Characterizing Microbial Community Composition and Function by Sequencing

Yssing Michaelsen, Thomas

DOI (link to publication from Publisher):
10.54337/aau456474762

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

CHARACTERIZING MICROBIAL COMMUNITY COMPOSITION AND FUNCTION BY SEQUENCING

BY
THOMAS YSSING MICHAELSEN

DISSERTATION SUBMITTED 2021

AALBORG UNIVERSITY
DENMARK

CHARACTERIZING MICROBIAL COMMUNITY COMPOSITION AND FUNCTION BY SEQUENCING

by

Thomas Yssing Michaelsen

AALBORG UNIVERSITY
DENMARK

Dissertation submitted October 2021

.

Dissertation submitted: October 2021

PhD supervisor: Prof. MSO Mads Albertsen,
Aalborg University

PhD committee: Professor Kåre Lehmann Nielsen (chair)
Aalborg University

Associate Professor Philip Pope
Norwegian University of Life Sciences

Assistant Professor Karoline Faust
KU Leuven

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Chemistry and Bioscience

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-996-7

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2021

ENGLISH SUMMARY

Microorganisms are everywhere, playing a pivotal role in all aspects of life on Earth – geochemical re-cycling of nutrients, sustaining ecosystems, cleaning wastewater, supporting human health as well as making us sick – to name a few. These abilities are by large enabled only when multiple microorganisms coexist, in microbial communities. Thousands of species can live together, forming entire miniscule ecosystems where members can share resources and collaborate to facilitate complex biological processes that no single organism is capable of. The focus of this PhD is studying microbial communities using high-throughput next generation sequencing, in particular their composition and putative function. Emphasis has been on applying data analysis approaches to improve understanding of the microbial communities at wastewater treatment plants (WWTPs) and in pouchitis, an inflammation of a surgically constructed J-shaped pouch made from the small intestine to treat ulcerative colitis, an inflammatory bowel disease.

A large part of this PhD were focused on WWTP microbes. We predicted community composition and abundance in activated sludge across several WWTPs, based on the matching influent wastewater community. Prediction was successful for many species, suggesting that assembly of the community at WWTPs is primarily shaped by immigration from the influent wastewater. This conclusion goes against the established dogma that the WWTP environment, governed by tight process control, is the main selective force. This makes our study a potential game-changer within WWTP management. We also performed metagenomics and metatranscriptomics studies of sludge from an anaerobic digester, a key component in the WWTP design, to recover and describe three novel syntrophic bacteria capable of degrading short-chain fatty acids. This is a key step in the anaerobic digestion process, which is known to cause operational problems. Therefore, better understanding of these bacteria may improve efficiency of digestion. Finally, we combined the metatranscriptomics data with similar data from other natural environments and performed a meta-analysis of antisense RNA (asRNA) as the first study ever, to our knowledge. We evaluated the putative function of asRNA and found no clear patterns in the data. Importantly, our study vitalizes topics for future research, which is important to gain a fundamental understanding of asRNA function in microbial communities.

Another part of this PhD project involved the study of pouchitis. We characterized its microbiome and found a reduced species richness and diversity compared to healthy pouches and controls. In a follow-up pilot-study, we attempted to perturb the pouchitis microbiome using fecal microbiota transplantation (FMT) from healthy donors. We found that half the study participants experienced remission and that the recipient microbiome approached that of the donor after FMT treatment. Interestingly, the degree of donor similarity seemed donor-depend. Collectively, our studies on pouchitis provided additional knowledge of its microbiome and promising results after FMT treatment. However, larger randomized controlled studies are needed to verify this.

On March 11th 2020, a nationwide lockdown in Denmark was issued due to COVID-19 and our laboratory was repurposed to whole-genome sequencing of SARS-CoV-2 virus. The first samples were processed within the following week and eventually scaled up to sequence >90% of all detected cases in Denmark. The focus of this PhD had to change accordingly. We developed the server infrastructure and bioinformatics pipeline to process the large amounts of data, as well as downstream interactive visualization tools and various custom reports, used by Danish authorities to track the pandemic and perform contact tracing. In late 2020 a new variant began to expand rapidly in the United Kingdom, later named Alpha, obtaining status as variant of concern in December 2020. Due to the extensive sequencing effort in Denmark, early modelling estimates enabled rapid adequate response by authorities to issue restrictions. However, spatiotemporal dynamics of spread across Denmark remained uninvestigated. We studied the introduction and transmission of Alpha in Denmark, identifying substantial early expansion of Alpha while it was still unmonitored. In addition, introductions from outside Denmark had a large impact in accelerating onwards spread. Our study shows the potential of genomic surveillance as a tool to monitor emerging variants and identify potential drivers of transmission, which can assist authorities in making balanced travel restrictions and self-isolation procedures while keeping society responsibly open.

DANSK RESUME

Mikroorganismer spiller en afgørende rolle for alle aspekter af livet på Jorden herunder genanvendelse af nærringsstoffer, opretholdelse af økosystemer og rensning af spildevand. De kan gøre os syge og andre kan gøre os raske. De gør det sjældent alene – tusinder af arter lever ofte sammen i mikrobielle samfund, som tilsammen udgør komplette økosystemer. Her kan individuelle medlemmer dele resourser og i samspil udføre komplekse biologiske processer som vil være umulige for enkelte organismer. Denne PhD afhandling studerer mikrobielle samfund ved hjælp af sekventering, særligt deres sammensætning og funktion. Fokus har været primært på anvendelse af data analyse metoder til at forbedre forståelsen af mikrobielle samfund i rensningsanlæg og pouchitis, en inflamation der kan opstå efter kirurgisk behandling af blødende tyktarmsinflammation.

Det mikrobiologiske samfund i rensningsanlæg blev undersøgt gennem flere studier. Vi prædikterede med lovende resultater både sammensætning og kvantitet af mikrober i aktiveret slam fra flere rensningsanlæg basseret på sammensætning og kvantitet af mikrober i indløbsspildevandet. Hermed viste vi at sammensætningen af mikrober inde på rensningsanlægget i høj grad er styret af hvad der introduceres fra indløbsspildevandet. Dette står i kontrast til den gængse opfattelse, at det er forholdende på anlægget og altså styrring af processen der er vigtig. Vores studie er derfor en potentiel game-changer for hvordan rensingsanlæg skal opereres fremover. Ved at udføre metagenomics og metatranscriptomics studier af slam fra en rådnetank, en vigtig komponent af et rensningsanlæg, fandt og beskrev vi tre nye syntrofiske bakterier som er i stand til at nedbryde kortkædede fedtsyrer. Dette er et vigtigt trin i metanisering, som kan volde operationelle problemer. Bedre forståelse af de bakterier som faciliterer dette kan potentielt være med til at optimere processen. Metatranscriptom data indgik ydermere sammen med lignende data fra andre naturlige miljøer i en meta-analyse af antisense RNA (asRNA), som efter vores orientering er det første af sin slags. Her undersøgte vi mulige funktioner af asRNA, uden at finde klare tendenser. Vores studie er dog stadig vigtigt, fordi det vitaliserer asRNA som et vigtigt emne for fremtidige studier, hvis vi skal gøre os forhåbninger om at få en fundamental forståelse af asRNAs rolle i mikrobielle samfund.

En anden væsentlig del af denne afhandling er studier i pouchitis. Vi fandt at pouchitis mikrobiomet har reduceret arts antal og diversitet sammenlignet med patienter uden pouchitis og kontroller. I et opfølgende pilot-studie forsøgte vi at ændre pouchitis mikrobiomet ved hjælp af fæcestransplantation fra raske donorer. Omkring halvdelen af patienterne oplevede bedring og at mikrobiomet hos patienterne ændrede sig imod at ligne donorenes mikrobiom efter behandling. Særligt interessant fandt vi at forskydningen var donor-afhængig, altså at nogle donorer var bedre end andre til at kolonisere patienten. I sammenfatning viste vores studier et distinkt mikrobiom ved pouchitis som tilsyneladende kan ændres ved fæcestransplantation. Sidstnævnte vil dog kræve større kliniske forsøg for at verificere.

Den 11 marts 2020 lukkede hele landet ned grundet spredningen af COVID-19. Vores laboratorie meldte sig friviligt til at udføre helgenom sekventering af SARS-CoV-2 viruset og sekventerede allerede indenfor en uge de første prøver, efterhånden >90% af alle detekterede tilfælde i Danmark. Afhandlingen skiftede fokus til at udvikle den it-infrastruktur og bioinformatikse pipeline der var nødvendig for at processere de store mængder data, foruden de interaktive visualiserings værktøjer og raporter som myndighederne havde brug for til at overvåge pandemien og udføre smittesporing. I efteråret 2020 begyndte en ny variant at sprede sig i England, senere navngivet Alpha, som fik status af variant of concern i december 2020. Takket være den omfattende sekventering i Danmark kunne væksten af Alpha modelleres tidligt i dens udbredelse, hvilket gjorde myndighederne i stand til hurtigt at indføre passende restriktioner. En detaljeret beskrivelse og forståelse af Alphas intog i Danmark manglede, så vi undersøgte introduktionen og spredningen af Alpha ved hjælp af det genomiske data. Resultaterne viste markant spredning i samfundet i perioden før Alpha blev overvåget og at løbende introduktioner fra udlandet har forøget spredning betydeligt. Vores studie viser potentialet for helgenom sekventering som et værktøj til at overvåge varianters udbredelse samt identificere potentielle årsager til øget spredning. Dette kan hjælpe myndighederne med at afbalancere rejserestriktioner og procedurer for selvisollering imod at holde samfundet åben i videst muligt omfang.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Mads Albertsen for giving me the opportunity to make this PhD thesis. I am grateful for our (mostly) concise and fruitful discussions, as well as allowing me to pursue my interest and ideas – including the bad ones. My time under your guidance have developed my academic skills tremendously and deeply influenced how I approach working life in general.

Thanks to all colleagues at the Albertsen Lab and at the Center for Microbial Communities for a warm and positive working environment, making it fun going to work (almost) every day. Especially my office mates, which counts many of you, for sharing and discussing new ideas as well as keeping the spirit high with good humor when needed.

I also want to show gratitude towards Nicola Segata and the people at Segata Lab for the opportunity to visit them. Thanks to Serena for taking me to the hospital when I crashed my bike, guess Italian roads were not ready for me yet. Special thanks to Federica and Nicolai for being outmost hospitable, treating me like a true friend and making me feel welcome instantly – you made me wish I had stayed a bit longer.

A sincere thanks to family and friends, for reminding me there is other things in life than work. Especially Rasmus, for inviting yourself to dinner rather spontaneously and pushing me out the house for a run on several occasions.

Thanks to my parents, without your loving support and encouragement to pursue my interests, I would not be where I am today. Finally, a deep heartfelt thanks to my wife Julie for giving me two wonderful kids while working on this PhD and making sure that our joint family-project stayed afloat during my sometimes long days and nights of working, especially when COVID-19 took all my time and effort. Standing on your shoulders allowed me to reach for the stars.

To my grandfathers, who newer got to "see me finish school" - I will never forget our long conversations about almost everything and the life-lessons you taught me. You both aspire me to be the best version of myself possible. Thank you!

TABLE OF CONTENTS

1 INTRODUCTION

Microorganisms and viruses are everywhere, making up thirty-five times the biomass of all animal life (*1*). Microbes are involved in almost every aspect of life on Earth, from biochemical nutrient cycling on global scale to the behavior and health of animals, including humans (*2*) – without them, life as we know it would simply not exist (*3*). Ever since the first microorganisms were discovered more than 300 years ago, humans have tried to understand, characterize, and utilize microorganisms for our own benefit. We now know that microorganisms can cause and cure human diseases (*4*) and are indispensable for human digestion, nutrient uptake, and the immune system (*5*). Various societal infrastructures as wastewater treatment, bioenergy production, and agriculture are completely dependent on an active microbial community to function (*3*).

On March 11th 2020, about halfway through this PhD study period, Denmark was thrown into a lockdown due to the novel SARS-CoV-2 virus causing COVID-19 disease was spreading rapidly across the world. Our laboratory was repurposed to whole-genome sequencing of SARS-CoV-2 positive cases from whole of Denmark (*6*), causing an abrupt shift of focus from microbial communities to the epidemiology of SARS-CoV-2. Therefore, this PhD-project is notably divided into two distinct parts. Both parts are, however, based on high-throughput sequencing data and shares many of the computational and statistical methods involved. The body of work on microbial communities was focused on wastewater treatment systems and the human microbiome. Thematically, these studies can be categorized based on the fundamental questions they in essence attempt to answer; “who are there?” or “what are they doing?” (*7*). First, studies of microbial community composition addressing “who are there?” are described. These involve a study of wastewater treatment plants, to predict community composition of sludge based on the community profile of influent wastewater. In two other studies we investigated the microbial community of pouchitis, a complication following surgical treatment of ulcerative colitis, and attempted to manipulate it using fecal microbiota transplantation. Next, the question “what are they doing?” is elaborated as a natural extension of the first, and will be explored in metagenomic and metatranscriptomic studies of microbial communities in wastewater treatment. Finally, for the second part of this PhD-project, the SARS-CoV-2 epidemic in Denmark is investigated from a genomics perspective, with focus on the highly transmissible Alpha variant.

1.1 STUDYING MICROBIAL COMMUNITIES

In almost every natural and engineered environment, microorganisms exists in communities of up to several thousand species living together from all major kingdoms of life (*7*). The ability of the community to carry out biological processes is often an emergent property of complex interactions between interdependent community members (*8*, *9*). Therefore, a detailed overview of constituent species and their abundance is necessary to deconstruct how biological functionality is obtained (*9*). This was previously done by culturing in the laboratory, using various media designed to select for specific organisms (*2*). The process is tedious and time-consuming, making the study of all relevant members of a community by culturing a monumental task. Furthermore, many members are interdependent and interacts in complex ways (*9*, *10*), which is currently not feasible to mimic nor investigate by culturing except for highly simplified systems (*11*). Next-generation sequencing (NGS) revolutionized the study of microbial communities, as a culture-free method that can generate massive amounts of high-quality sequence data for, in principle, the entire community simultaneously at low cost and turn-around time (*12*, *13*).

1.1.1 NEXT GENERATION SEQUENCING

Up until the mid-2000s almost all sequencing had been done exclusively using variations of Sanger sequencing (*14*), including sequencing of the first human genome (*15*). This method was largely replaced by NGS technologies, a term coined in the mid-2000s, as these has substantially higher throughput (*16*). NGS has practically replaced Sanger sequencing as the method of choice, in particular for studying microbial communities. Two different sequencing strategies are often applied in NGS; amplicon or shotgun sequencing (**Figure 1**). Amplicon sequencing uses polymerase chain reaction (PCR) to amplify specific genomic regions or genes of interest prior to sequencing, allowing sequencing of only the selected target(s) (*17*). In the study of microbial communities, the 16S ribosomal RNA (rRNA) gene is by far the most used. The 16S gene encodes a sub-unit of the prokaryotic ribosome involved in protein sequencing and is universally present in all prokaryotic cells (*2*). Furthermore, the gene has regions where the nucleotide sequence is highly conserved as well as variable regions, making it an ideal phylogenetic marker for taxonomic community profiling (*18*, *19*). However, due to the PCR step which uses primers to target specific nucleotide sequences, 16S amplicon sequencing is sensitive to bias during sample preparation and sequencing. Standard protocols only sequence a small region of the 16S gene and the choice of which region to sequence can have large impact on the resulting taxonomic profile (*20*). Primers may miss certain taxonomic groups and is therefore not able to capture the full community. Finally, PCR amplification can introduce biases in the quantitative signal due to variability in amplification efficiency (*21*). Shotgun sequencing, critically, does not involve such PCR amplification step. Instead, all input DNA or RNA are fragmented into approximately equal-sized fragments and sequenced, in principle capturing all the DNA or RNA within a sample.

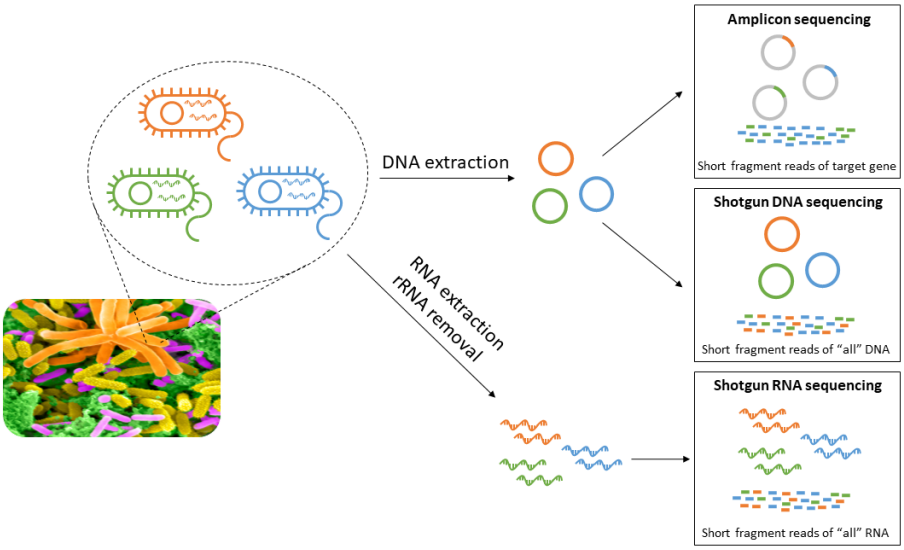

Figure 1 Principal overview of different approaches to sequence microbial communities. *Inspired by* (*22*)**.**

Specifically for RNA sequencing, rRNA is often removed during laboratory preparation as it constitutes more than 95% of total RNA and would otherwise dilute the signal from messenger RNA (mRNA) in the sample (*23*), **Figure 1**). To characterize the DNA content of samples, the choice to use amplicon or shotgun sequencing is context-dependent, and each method has its own advantages. Compared to shotgun DNA sequencing, amplicon sequencing is cheaper and give more robust high-level community characterization (*24*, *25*), at the expense of taxonomic resolution that can be achieved (*25*). Shotgun sequencing targets (in principle) all DNA of the sample, which also allows a complete profiling of all genes that can be used to assess functional potential (*23*). To sequence the whole SARS-CoV-2 genome, an alternative amplicon sequencing approach is used (*26*). Here, multiple amplicons are generated which overlaps across the full length of the genome. These are then assembled programmatically into one continuous genome sequence (*27*), somewhat similar to what is typically achieved using shotgun DNA sequencing.

1.1.2 OMICS

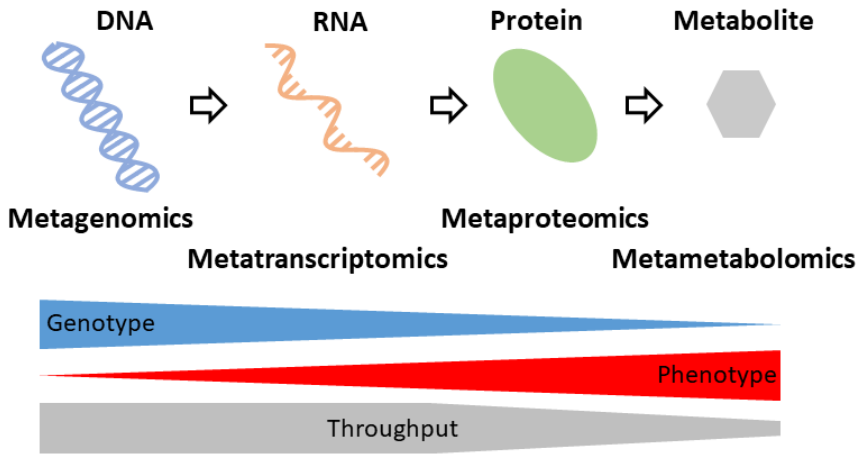

Figure 2 Illustration of the "omics cascade". *Inspired by* (*23*, *28*)

The "omics cascade" (**Figure 2**) refers to sequential levels of biomolecules that from DNA to metabolite holds information shifting from the genotype, information about metabolic potential (i.e. "what can happen") to the phenotype, the realized biological processes occurring (i.e. "what is happening") (*28*, *29*). Each biomolecule is associated with its own field of techniques and methods, as a whole referred as omics, with the semi-prefix "meta" referring to applications in microbial communities (**Figure 2**, (*25*). Metagenomics and metatranscriptomics can indicate potential function and activity, but not all genes may be transcribed and transcription does not necessarily lead to protein synthesis (*30*). Metaproteomics can be used to quantify protein levels and as proteins to some extent can be mapped back to the original RNA sequence template, it can in principle provide the link between genomic identity of community members and what functional role they have in the community (*25*). Metametabolomics is the analysis of all metabolites in a microbial community, which includes small molecules released by organisms into the immediate environment (*28*). Identifying and quantifying metabolites can give a snapshot view of all biological processes currently active, in principle providing the most direct indicator of the phenotype displayed (**Figure 2,** (*28*). An integrated approach, combining all different omics techniques is encouraged to balance out the strengths and weaknesses for various methods and develop more complete biological hypotheses. Studies are increasingly appearing (*23*), but several challenges remain. Metaproteomics has issues with complicated sample preparation, incomplete databases, and difficulties with quantitative analysis (*31*). Metametabolomics is still under development and as many metabolites from complex communities are completely unknown, its applications are limited. Metabolites span a wide array of different types of molecules further

complicating analysis (*23*). Collectively, approaches based on NGS are still highly popular, providing substantial insight into community composition and functionality with high-throughput.

Metagenomics allows the study of all genomic content of a sample, however, interpretation of data requires good reference databases that sufficiently represent the microbes of the community being studied. Arguably these only exists for the extensively studied human microbiome (*32*), but substantial efforts are put into generating databases for other environments (*33*). Such databases of microbial genomes are to a wide extent assembled directly from metagenomics data using a genome-centric approach (*34*). This involves multiple bioinformatics steps to generate metagenome-assembled genomes (MAGs) (*35*). Simplified, the reads are first assembled in to longer contigs which are then grouped into bins, based on various parameters such as sequence composition and coverage similarity (*34*). The set of bins are typically subjected to refinement, in particular to remove duplicated genomes that are from the same organisms (*36*). The finalized set of MAGs are annotated using gene-prediction software (*37*), allowing to reconstruct their putative metabolism and function in the community. However, obtaining complete genomes using this approach has limitations and may not be possible for all organisms, primarily due to high strain heterogeneity causing chimeric and fragmented genomes as well as low-prevalence microbes with insufficient data to reconstruct their genome (*38*). Metatranscriptomics is often used in combination with metagenomics to quantify gene expression of the retrieved MAGs from the same environment, as a natural extension to further substantiate their putative function and metabolism. It is also commonly used to study changes of functional activity in response to perturbation, by identifying genes with changed expression patterns using differential expression analysis (*39*). Alternatively, 16S amplicon sequencing can be used for genomic profiling of the community composition and abundance (*23*). Databases of 16S sequences has existed for longer time than full-genome databases and has good coverage of high-level microbial diversity. This is by large due long-running initiatives as the Human Microbiome Project (*40*) and the Earth Microbiome Project (*41*), as well as environment-specific initiatives as the Microbial Database for Activated Sludge (MiDAS) for wastewater systems (*42*). All these projects have subsequently included shotgun DNA sequencing and whole genomes into their databases. Prior to analysis, the amplicon reads are traditionally clustered into operational taxonomic units (OTUs) based on sequence identity, typically 97%, to reduce impact of sequencing errors (*43*). However, clustering may miss subtle real biological variation, such as single-nucleotide variants (SNPs) that may distinguish between different species or strains (*44*). Instead of OTUs exact sequences are increasingly used, referred as amplicon sequencing variants (ASVs), with sequencing errors corrected directly using denoising algorithms (*45*). Importantly, 16S amplicon sequencing typically targets only a sub-region of the 16S gene and cannot provide the same taxonomic resolution as using the full-length 16S gene (*19*), for which new methods are being developed (*45*, *46*).

1.2 COMPOSITION OF MICROBIAL COMMUNITIES

"Who is there?" is a fundamental question often addressed in studies of microbial communities (*7*). As NGS continues to increase output quantity and sequence quality, the resulting data complexity also increase to reflect more and more of the real-world complexity of microbial communities (*47*). In this PhD project three papers (**Paper 1, 2, and 3**) were devoted to studying the composition of microbial communities, to gain insight into the constituent members, as well as identify which members were able to differentiate between different environmental conditions. The background and results of these studies are outlined in this section.

1.2.1 COMMUNITY ASSEMBLY IN WASTEWATER TREATMENT PLANTS

Wastewater treatment plants (WWTPs) is an integral part of modern society, reducing environmental impact of human activity. This is facilitated by the microbial community, primarily bacteria, by degrading organic matter, removing pollutants, and recover nutrients such as nitrogen and phosphorus (*48*). Aforementioned processes require different microorganisms to interact with each other in complex and partially unknown ways, making the study of community assembly and what drives it crucial. The processes driving community assembly can be grouped into stochastic processes (e.g. dispersal, immigration) and deterministic processes (e.g. environmental conditions, biotic interactions) (*49–55*). Clarifying which processes dominate in WWTP systems has been debated, but studies investigating this are scarce (*52*). In **Paper 1** we sought to determine the impact of immigration from influent wastewater (IWW) on the microbial community in activated sludge (AS). This was done by simultaneously taking paired samples of IWW and AS (**Figure 3**) across several months, for 11 different WWTPs. We used a data-driven multivariate analysis approach (partial least squares regression [PLSR]) to predict community profile in AS from the community profile in IWW. Of the taxa in AS that were above certain quantity thresholds (**Paper 1**), we were able to successfully predict the relative abundance of 44% of them. We further developed a heuristic to type taxa that could be predicted in AS by either their own abundance in IWW (prediction by univariate regression performed well, termed "univocal association") or a combination of several taxa in AS (univariate prediction performed poorly and PLSR performed well, termed "interspecies association"). We found interspecies associations to drive predictability in AS for most taxa, however the exact interpretation and implication of this is not clear. Overall, we suggested that stochastic processes, manifested here as immigration, likely are more important than previously thought in shaping the community of WWTPs.

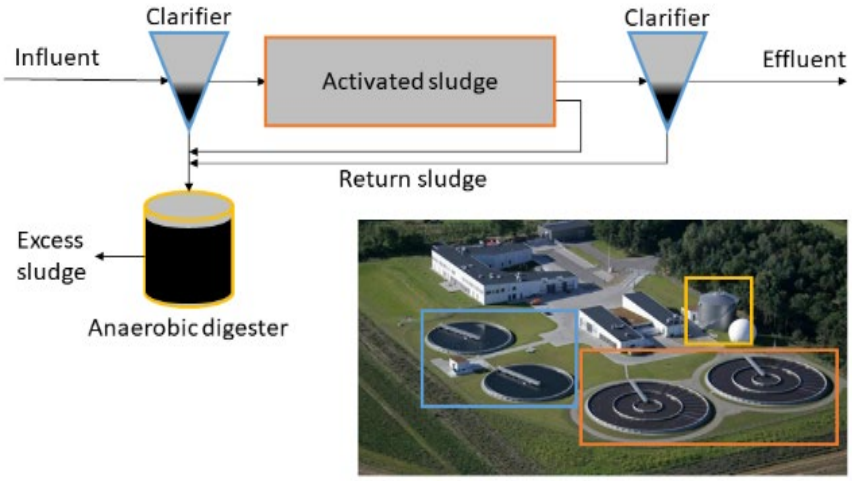

Figure 3 Overview of a typical wastewater treatment plant. *A schematic overview of the main components is provided, with arrows indicating the direction of flow. Borders of each component are colored according to rectangles overlaid on the picture in bottom-right, showing their placement on a real wastewater treatment plant, here Mariagerfjord Renseanlæg in North Denmark. Inspired by* (*56*)*. Image credit: mariagerfjordvand.dk*

The key idea outlined in **Paper 1** were transferred to another study (*57*), where we evaluated the effect of transplanting AS from a donor to recipient WWTP on the microbial community. This was done by emptying out 75% of all AS in the recipient plant, before filling it up again with activated sludge from the donor plant, transported by several trucks within one day (*57*). The transplantation only affected the recipient AS microbial community temporally before returning to pre-transplant state after approximately 40 days. Importantly, the IWW community of the recipient WWTP were also sampled in the study period and showed some similarity to the recipient AS community. This is in agreement with what we observed in **Paper 1**, suggesting that immigration from IWW induce resilience to perturbations of the AS community (*57*).

1.2.2 THE MICROBIOME OF CHRONIC POUCHITIS AND POSSIBLE TREATMENT

The human microbiome is the collection of all microorganisms living inside and on the human body. The number of genes associated with these are likely 100 times that of human genes and shows substantial diversity, emphasizing the enormous functional potential of the microbiome to modulate many aspects of human biology (*4*). The microbiome has been associated with a wide array of diseases, from inflammatory bowel disease (IBD) (*58*) to behavioral disorders (*59*). Ulcerative colitis (UC) is an IBD in the large intestine (colon) and rectum causing pain, fatigue, rectal bleeding, and weight loss amongst other symptoms (*60*). Surgical treatment of UC, called ileal pouch-anal anastomosis (IPAA; **Figure 4**), first involves removal of the colon and rectum. Then a pouch is constructed from the end of the small intestine to hold waste which is attached directly to the anus, allowing relatively normal bowel movement (*60*). Pouchitis is a common long-term complication following IPAA, occurring in 23% to 60% of cases (*61*). The exact cause of pouchitis remains uncertain, but recent studies have suggested that the gut microbiome may play a major role in development of pouchitis (*62*).

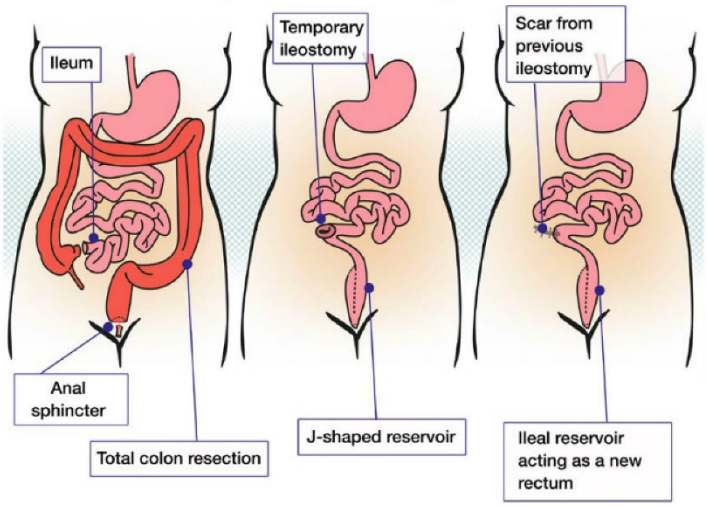

Figure 4 Overview of the surgical procedure for ileal pouch-anal anastomosis. *First the colon and rectum are removed. Then a J-shaped pouch is constructed from the end of the small intestine (ileum). A temporary ileostomy is created to allow waste disposal during the healing process, which is later removed. Adapted from endoinflamatoria.com*

To improve our understanding of the pouchitis microbiome, we investigated the difference in microbiome composition of patients with a normal pouch and pouchitis, and compared these to healthy controls (**Paper 2**). Results showed that the richness and diversity was highest in healthy controls and lowest in patients with pouchitis, while patients with normal pouch was intermediate of these. We also identified species associated specifically with pouchitis (**Paper 2**). In a related pilot-study (**Paper 3**), the potential of fecal microbiota transplantation (FMT) to change the microbiome of pouchitis was investigated. Across a period of 14 days, each patient was treated daily with FMT from five different donors. We sampled the microbiome prior to FMT and after the treatment period, as well as the five donors. Overall, patient microbiome shifted towards being more similar to the donors after FMT. To gain more details, an engraftment analysis (*63*) was performed, which is used to infer the origin of species found in the patient after FMT treatment. Interestingly, we found weak support for a donor-specific ability to engraft, suggesting that some donor communities are better than others to establish themselves in recipient patients (**Paper 3**). The same methodology of engraftment analysis was applied in another study investigating the effect of FMT on behavior of rats (*64*). Rats were treated with FMT from a human healthy donor and a patient suffering from major depressive disorder (MDD), to assess potential changes in behavior. FMT from MDD patients induced depressive-like behavior in the form of increased immobility and less struggling compared to FMT from healthy individuals during a forced swim test. This could be linked to differential abundance of certain microbes (*64*).

1.3 FUNCTIONAL CHARACTERIZATION OF MICROBIAL COMMUNITIES

Studies of the community composition (i.e. **Papers 1, 2, and 3**), can provide valuable insights about the identity and abundance of species in specific environments, as well as discriminate between them (*47*) – i.e. "who are there?". Knowing which species are present from 16S amplicon sequencing can give some functional insight by matching to databases with functional information (*42*), or metagenomics can be used to generate a database *de novo*, from which the functional potential can be predicted (*33*). However, neither approaches provide much information in answering "what are they doing?" This can be done by measurement of gene expression activity, which can provide indications of the metabolic processes actively occurring (*23*). Two papers (**Paper 4 and 5**) were devoted to functional studies of microbial communities using metagenomics and metatranscriptomics. The background and results of these studies are outlined in this section.

1.3.1 SYNTROPHIC BACTERIA IN ANAEROBIC DIGESTERS

Anaerobic digestion is a multi-step biological process, converting organic matter into methane which can be used for biogas production (**Figure 5**). It is a key component of modern WWTPs, by removing excess sludge from the activated sludge process (**Figure 3**). Furthermore, excess material from the anaerobic digester (AD) can be refined to higher-value products as bioplastic (*65*) or used as fertilizer.

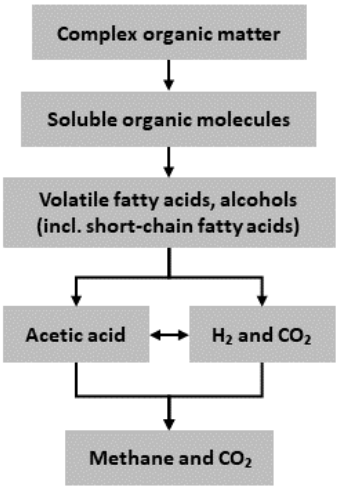

Figure 5 The anaerobic digestion process. *Inspired by* (*66*).

The AD process consists of four overall steps, orchestrated by different organisms, with each step dependent on the previous one (*67*). Although the overall process is well characterized from laboratory culturing, the organisms performing these reactions in a full-scale AD are still relatively unknown and their interactions with other organisms even more so (*67*). We studied the response of microorganisms in AD sludge to stimuli of multiple short-chain fatty acids (SCFA) in **Paper 4**. Degradation of SCFAs is a key step in anaerobic digestion (**Figure 5**), which otherwise induce SCFA accumulation and performance instability, which often occur during operation of full-scale ADs (*68*). Syntrophic bacteria plays a major role in degradation of SCFAs (*69*), however, their exact identity and characteristics remain largely unknown (*70*). In **Paper 4** we analyzed gene expression response of the AD community to SCFA stimuli. We used a genome-centric approach, where metatranscriptomes were mapped to MAGs derived from metagenomes obtained from the same environment. Doing so, database biases are mitigated and gene expression patterns can be linked directly to individual community members (*71*). We annotated the MAGs, allowing us to identify those with upregulated expression of pathways and genes related to SCFA degradation. Three novel syntrophic bacteria were identified and their genomic content described, adding to the understanding of syntrophic bacteria in AD systems (**Paper 4**). Since the completion of this study, we combined shotgun DNA sequencing with long-read Oxford Nanopore sequencing technology to recover more than 1000 high-quality genomes from activated sludge sampled from 23 different WWTPs (*33*). These all fulfilled the stringent minimum information about a MAG (MIMAG) criteria (*35*) and collectively accounted for 30% of the community, thus representing a high-quality environment-specific database of reference genomes for activated sludge (*33*).

1.3.2 ANTISENSE RNA – SIGNAL OR NOISE?

The majority of the bacterial transcriptome (>95%) is non-coding rRNA and transfer RNA (tRNA) (*72*). A substantial part of the remainder is messenger RNA (mRNA), which is typically the target of most analysis (*23*). In addition, a plethora of small regulatory noncoding RNAs (sRNAs) exists, which regulate transcription through various mechanism whose exact nature is an area of active research (*30*). Antisense RNA (asRNA) is a type of sRNA matching the template DNA strand (the antisense strand) inside a gene (**Figure 6,** (*73*) and is ubiquitous in both bacteria (*74*) and archaea (*75*). Some studies suggest a functional purpose of asRNA by binding to their mRNA counterpart, hereby regulating translation (*74*). In contrast, other studies suggest it has no functional purpose and is just a byproduct of transcription, due to spurious promotor sites appearing randomly across the genome which can initiate transcription (*76*). The promotor site in prokaryotes typically consists of the nucleotides TATAAT (*2*). Therefore, genomes with high AT-content should produce more promotor sites at random, in turn producing more asRNA and induce a positive relation, which was tested and verified across multiple different bacteria (*77*).

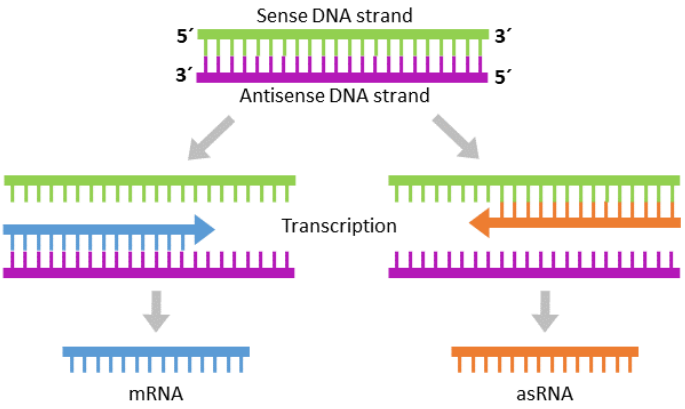

Figure 6 Relation between messenger RNA (mRNA) and antisense RNA (asRNA). *Adapted from* (*78*).

Despite the growing number of studies on asRNA and its putative function, all have been performed exclusively on pure cultures of bacteria and archea – except one study on the human microbiome (*73*). Many shotgun RNA sequencing methods provides stranded reads as output, meaning that their orientation relative to the reference genome can be determined. The study of asRNA can therefore be performed directly on metatranscriptomics data, without the need for additional laboratory procedures and preparations. In **Paper 5** we performed a combined study and mini-review on asRNA across multiple environments with different microbial communities, to quantify asRNA and explore its potential function. We found that levels of asRNA varied considerably between environments, in some cases dominating over regular gene expression (**Paper 5**). We did not find support for asRNA being the result of transcriptional noise. We performed an enrichment analysis of functional categories comparing the fraction of genes with highest relative asRNA expression against the rest. The asRNA-enriched genes were consistently enriched for unknown functions, emphasizing the need for further studies on asRNA in complex communities to elucidate their potential role in complex microbial communities (**Paper 5**).

1.4 SARS-COV-2 SEQUENCING WHOLE OF DENMARK

Since it was first observed in Wuhan China, the novel SARS-CoV-2 virus causing COVID-19 disease spread rapidly across the globe with the first Danish case registered on 27 February 2020 (*79*). On 11 March the prime minister issued a nationwide lockdown, effectively closing all Danish universities from 12 March onwards. A few days later on 14 March, my supervisor prof. Mads Albertsen mobilized relevant stakeholders and co-established the Danish Covid19 Genome Consortium (DCGC) as a collaborative effort between our lab at Aalborg University (AAU), Statens Serum Institute (SSI), Hvidovre Hospital, and Aalborg University Hospital (*6*). Our lab started sequencing of the first positive SARS-CoV-2 samples already a week later. During summer 2020 sequencing at AAU was optimized and sequencing nodes were established at most Danish hospitals to enable rapid local sequencing. In parallel, I co-developed a bioinformatics pipeline to generate full length SARS-CoV-2 genomes centered on the artic protocol (*27*). We also enabled interactive visualizations of the data, on GDPR-secure servers which were setup for the purpose, using the Nextstrain platform (*80*) and various custom reports. This allowed SSI to explore the genomic data put into context by rich individual-level metadata, including summary reports stratified across administrative regions and sociodemographic divides, as well as reports targeting specific outbreaks. As case numbers started rising during autumn 2020 (Figure 5) DCGC upscaled considerably, reaching 5000 samples per week from December 2020 onwards, just as case numbers peaked. As cases dropped the percentage of sequenced cases has been stable at >90% since January 2021 (**Figure 7**), making Denmark world leading in SARS-CoV-2 sequencing and public genome sharing via the GISAID initiative (*81*).

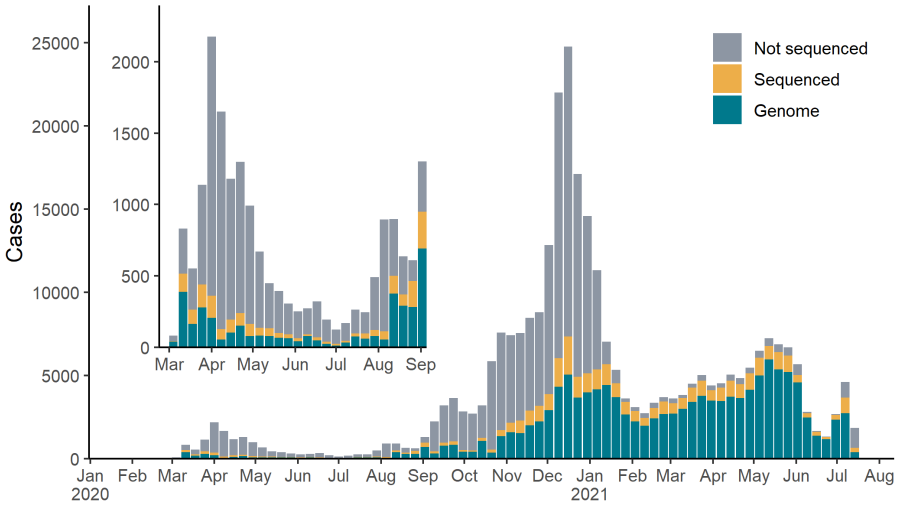

Figure 7 Sequencing rate by the Danish Covid-19 Genome Consortium across time. *Total height corresponds to the total number of cases for each week, while colour indicate if a genome was available, if sequenced but yielded no genome and if not sequenced. The inserted plot shows data for a sub-period with rescaled y-axis.*

1.4.1 THE ALPHA VARIANT

To enable surveillance of SARS-CoV-2 variants for governmental decision making, the COVID-19 Genomics UK Consortium (COG-UK) developed a SARS-CoV-2 classification and definition system which has become the standard of classifying variants based on their potential impact on SARS-CoV-2 countermeasures such as vaccines, therapeutics, and diagnostics (*82*). The system is composed of three levels of escalating impact on health; variant of interest (VOI), variant of concern (VOC), variant of high consequence (VOHC). The first level, VOI, has features that might be associated with increased health impact, such as genetic markers for enhanced virus fitness and/or disease severity, or increased prevalence in a limited geographical area. However, further investigation is needed to assess this. If a VOI proves to display aforementioned features across multiple peer-reviewed studies and geographical regions it will be escalated to VOC status. In this case, additional countermeasures might be needed such as increased testing and restrictions, as well as investigations to determine the effectiveness of vaccines and treatments against the variant. Finally, a variant might reach VOHC status. Currently, no variant exist or have existed that rise to this level, which would have profound impact on the global society as current countermeasures would be ineffective to sustain epidemic control (*82*).

The Alpha variant (B.1.1.7 lineage) was first reported in the United Kingdom in September 2020 (*83*) and subsequently spread across the world. It was among the first variants to receive VOC status on 18 December 2020 due to an estimated increased transmissibility of 71% (95% CI: [67%, 75%]) compared to other variants (*84*). Since then, several studies have affirmed an increased transmissibility of Alpha ranging from 29% to 130% (*85–89*). In Denmark, early estimates from January 2021 were between 36%-55% increased transmissibility (*90*, *91*). In **Paper 6** we updated those estimates to 58% (95% CI: [56%, 60%]), substantiating the accuracy of early results and confirming the value of comprehensive and real-time genomic surveillance of SARS-CoV-2 for early warning to sustain epidemic control. We outlined the introduction and transmission of Alpha in Denmark using ancestral state reconstruction from a phylogenetic analysis of genomic data (*92*) with detailed metadata on every positive case from Danish authorities. We found that introductions contributed substantially to case numbers throughout the study period, despite restrictive measures being in effect mid-study and onwards. In addition, spread of Alpha within the country was to great extent associated with the first reported case of it in Denmark, revealing that sustained transmission in society caused substantial expansion before restrictive measures were taken. The conclusions of **Paper 6** have wide implications to the nation-level handling of VOCs to follow such as the Delta variant, first observed in India, which as of July 2021 constitute 90% of all new cases in Denmark (*93*).

The massive sequencing effort by DCGC coupled with detailed and comprehensive metadata collected by the Danish authorities made it possible to study the epidemiological characteristics of Alpha in great detail. In (*94*) we assessed the risk of COVID-19 hospitalization in individuals with Alpha compared to those with other SARS-CoV-2 variants. We found an increased risk of hospitalization, but only after adjusting for covariates. In (*95*) we estimated the household transmissibility of Alpha compared to other variants, across groups of age and viral load. We found an increased transmissibility of 50% to 70%, which varied across age groups with highest transmission rate and risk amongst children and elderly.

2 DISCUSSION AND PERSPECTIVE

In **Paper 1, 2, and 3** we conducted studies on the microbial community composition in different environments. In **Paper 1** we showed that community composition and abundance could be predicted in AS from the IWW across comparable Danish WWTP. Albeit data-driven approaches have been used to investigate the relationship between IWW and AS before (*96–100*), they did not perform prediction and focused only at community-level associations or by simple presence/absence of taxa. Alternatively, approaches to evaluate impact of immigration based on ecological theoretical concepts and models exists (*52*, *101–103*), but do not quantify interactions between taxa directly and often model the community as a whole. The results presented in **Paper 1** points toward a revised understanding of WWTPs, where the design and operation should not only be focused at optimizing the AS community directly at the plant, but also by evaluating the source community from the sewer systems. In **Papers 2 and 3** we identified a distinct pouchitis microbiome and showed promising results of FMT from healthy donors to change the microbiome of pouchitis. Although studies of pouchitis microbiome has been performed previously (*104*), our studies are unique by using consistent sampling and sequencing protocols across groups. Engraftment analysis further revealed that the success of FMT might be donor-specific. This could have implications for possible FMT treatment strategy, as it might be necessary to perform donor matching with the recipient patient prior to FMT. Using engraftment analysis to study changes in microbial composition after FMT is a powerful approach which can be used to investigate which factors are important to ensure engraftment of donor microbes in the recipient (*63*). It is important to recognize that **Paper 2 and 3** are both based on relatively few patient samples and can only be considered suggestive. Further studies with larger sample sizes and potentially using metagenomics instead of 16S amplicon data are needed, to obtained the statistical power necessary and also verify that changes in microbial community composition is related to changes in the functional potential of the community, which might be linked to the pathology of pouchitis.

In **Paper 4 and 5** we studied the metagenome and metatranscriptome of microbial communities. We found three novel syntrophic bacteria in sludge from anaerobic digesters that could metabolize SCFAs (**Paper 4**) and demonstrated the power of a combined metagenomics and metatranscriptomics approach as a first-line screening methodology to link community potential to activity under certain conditions (**Paper 4**). The MAGs recovered in this study were assembled exclusively from conventional shotgun DNA sequencing data, with short read lengths of maximum 300 base pairs (**Paper 5**). This makes it difficult to handle repetitive regions during assembly and increase the computational effort needed (*71*). By combining short reads with long-read technologies as Oxford Nanopore in a hybrid approach, such issues can be alleviated substantially, as we showed (*33*). Using high-quality MAGs derived by this hybrid-approach as reference databases for metatranscriptomic studies may provide

even better resolution of gene expression analysis, in particular to investigate small noncoding RNAs as for example asRNA. In **Paper 5** we explored antisense RNA expression across different microbial environments. The levels of asRNA varied considerably across environments and were driven primarily by a few genes with unknown function. This study is to our knowledge (still) the second study, after Bao and colleagues (*73*), to address asRNA in microbial community and the first performing a comparative study across multiple environments. We did not find support for any particular function of asRNA, highlighting the need for further studies to elucidate this. The suggested function or asRNA are many (*30*, *74*, *75*, *105*) and deciphering the extent to which each of them is occurring from metatranscriptomics data will arguably be very difficult. Primarily due to the compounded signal, but also to setup the correct data analysis method to test hypotheses (**Paper 5**). In a transcriptomics study of a single cultured organisms' response to cyanate treatment, I incorporated asRNA into the differential expression analysis between the two conditions. We were able to identify several interesting genes which had different ratios of sense/antisense RNA in the controls compared to cyanate treatment (*106*). Of course, such results are preliminary and may be artefacts of other mechanisms such as changes in mRNA expression only, but may also suggest a regulatory role of asRNA. These results are preliminary, and further studies are needed to formalize and verify this.

In **Paper 6** we explored the introduction and transmission of the SARS-CoV-2 Alpha variant. We found that the first introduction in November 2021 caused substantial onwards transmission and that continuous introductions from outside Denmark further fueled the Alpha expansion. Our study highlights the powerful potential of genomic surveillance to reconstruct spread of variants and from that gain insight into the underlying causal drivers. The epidemiological characteristics of the Alpha variant was further investigated in two additional studies (*94*, *95*). At the time of publishing, no studies had investigated Alpha hospitalization risk nor transmissibility on individual-level population data as that provided by the Danish surveillance effort, enabling proper inclusion of relevant covariates. In conclusion, all these studies enable policy makers to make more informed decisions on how to balance opening the society while maintaining control of the epidemic. This is by large possible due to the scale and extent of the Danish genomic surveillance effort of SARS-CoV-2, which is amongst the most comprehensive in the world (*81*). However, the usage of modern phylogenetic analysis incorporating epidemiological data to characterize the state of the epidemic (contact numbers, superspreading events, etc.) or investigate outbreaks are lagging behind, largely due to the massive amounts of data being generated and lack of training in how to analyse it (*107*). This is known and initiatives are taken to alleviate it, for example in the United Kingdom (*108*). The SARS-CoV-2 virus will continue to exist indefinitely in some form (*109*) and with it, surveillance programs will play an important part in the long-term solutions to handle it.

3 CONTRIBUTIONS TO OTHER PAPERS

- J. Wiesinger, C. W. Herbold, **T. Y. Michaelsen**, M. Palatinszky, M. T. Wolfinger, S. M. Karst, M. Albertsen, P. Pjevac, M. Wagner. Transcriptional response of the ammonia-oxidizing Archaeon ca. nitrososphaera gargensis to growth on cyanate as substrate. *In preparation*
- G. Dottorini, D. S. Wágner, M. Stokholm-Bjerregaard, S. Kucheryavskiy, **T. Y. Michaelsen**, M. Nierychlo, R. William, P. H. Nielsen, P. H. Nielsen. Full-scale activated sludge transplantation reveals highly resilient community structure. *In preparation*
- T. B. Rasmussen, J. Fonager, C.S. Jørgensen, R. Lassauniére, A. S. Hammer, M. L. Quaade, A. Boklund, L. Lohse, B. Strandbygaard, M. Rasmussen, **T. Y. Michaelsen**, S. Mortensen, A. Fomsgaard, G. J. Belsham, A. Bøtner. Infection, recovery and re-infection of farmed mink with SARS-CoV-2. *Submitted*
- J. K. Knudsen, **T. Y. Michaelsen**, C. Bundgaard-Nielsen, R. E. Nielsen, S. Hjerrid, P. Leutscher, G. Wegener, S. Sørensen. Faecal microbiota transplantation from patients with depression into rats modulates mood-related behavior. *Preprint at Scientific Reports*
- F. P. Lyngse, K. Mølbak, R. L. Skov, L. E. Christiansen, L. H. Mortensen, M. Albertsen, C. H. Møller, T. G. Krause, M. Rasmussen, **T. Y. Michaelsen**, M. Voldstedlund, J. Fonager, N. Steenhard, C. Kirkeby, The Danish Covid-19 Genome Consortium, Increased Transmissibility of SARS-CoV-2 Lineage B.1.1.7 by Age and Viral Load: Evidence from Danish Households. *Preprint on medRxiv*
- T. S. Jørgensen, K. Blin, F. Kuntke, H. K. Salling, **T. Y. Michaelsen**, M. Albertsen, H. Larsen. A rapid, cost efficient and simple method to identify current SARS-CoV-2 variants of concern by Sanger sequencing part of the spike protein gene. *Preprint on medRxiv*
- P. Bager, J. Wohlfahrt, J. Fonager, M. Rasmussen, M. Albertsen, **T. Y. Michaelsen**, C. H. Møller, S. Ethelberg, R. Legarth, M. S. F. Button, S. Gubbels, M. Voldstedlund, K. Mølbak, R. L. Skov, A. Fomsgaard, T. G. Krause, Danish Covid-19 Genome Consortium, Risk of hospitalisation associated with infection with SARS-CoV-2 lineage B.1.1.7 in Denmark: an observational cohort study. *Lancet Infect. Dis.* (2021)
- C. M. Singleton, F. Petriglieri, J. M. Kristensen, R. H. Kirkegaard, **T. Y. Michaelsen**, M. H. Andersen, Z. Kondrotaite, S. M. Karst, M. S. Dueholm, P. H. Nielsen, M. Albertsen, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat. Commun.* **12**, 2009 (2021)
- R. F. Brøndum, **T. Y. Michaelsen**, M. Bøgsted, Regression on imperfect class labels derived by unsupervised clustering. *Brief. Bioinform.* **22**, 2012–2019 (2021)
- N. de Jonge, **T. Y. Michaelsen**, R. Ejbye-Ernst, A. Jensen, M. E. Nielsen, S. Bahrndorff, J. L. Nielsen, Housefly (Musca domestica L.) associated microbiota across different life stages. *Sci. Rep.* **10**, 7842 (2020)
- **T. Y. Michaelsen**, J. Richter, R. F. Brøndum, W. Klapper, H. E. Johnsen, M. Albertsen, K. Dybkær, M. Bøgsted, A B-cell-associated gene signature classification of diffuse large B-cell lymphoma by NanoString technology. *Blood Adv.* **2**, 1542–1546 (2018)

In addition, as member of the Danish Covid19 Genome Consortium:

- Â. O'Toole, V. Hill, O. G. Pybus, A. Watts, I. I. Bogoch, K. Khan, J. P. Messina, COVID-19 Genomics UK (COG-UK) consortium, Network for Genomic Surveillance in South Africa (NGS-SA), Brazil-UK CADDE Genomic Network, H. Tegally, R. R. Lessells, J. Giandhari, S. Pillay, K. A. Tumedi, G. Nyepetsi, M. Kebabonye, M. Matsheka, M. Mine, S. Tokajian, H. Hassan, T. Salloum, G. Merhi, J. Koweyes, J. L. Geoghegan, J. de Ligt, X. Ren, M. Storey, N. E. Freed, C. Pattabiraman, P. Prasad, A. S. Desai, R. Vasanthapuram, T. F. Schulz, L. Steinbrück, T. Stadler, Swiss Viollier Sequencing Consortium, A. Parisi, A. Bianco, D. García de Viedma, S. Buenestado-Serrano, V. Borges, J. Isidro, S. Duarte, J. P. Gomes, N. S. Zuckerman, M. Mandelboim, O. Mor, T. Seemann, A. Arnott, J. Draper, M. Gall, W. Rawlinson, I. Deveson, S. Schlebusch, J. McMahon, L. Leong, C. K. Lim, M. Chironna, D. Loconsole, A. Bal, L. Josset, E. Holmes, K. St George, E. Lasek-Nesselquist, R. S. Sikkema, B. Oude Munnink, M. Koopmans, M. Brytting, V. Sudha Rani, S. Pavani, T. Smura, A. Heim, S. Kurkela, M. Umair, M. Salman, B. Bartolini, M. Rueca, C. Drosten, T. Wolff, O. Silander, D. Eggink, C. Reusken, H. Vennema, A. Park, C. Carrington, N. Sahadeo, M. Carr, G. Gonzalez, SEARCH Alliance San Diego, National Virus Reference Laboratory, SeqCOVID-Spain**, Danish Covid-19 Genome Consortium (DCGC)**, Communicable Diseases Genomic Network (CDGN), Dutch National SARS-CoV-2 surveillance program, Division of Emerging Infectious Diseases (KDCA), T. de Oliveira, N. Faria, A. Rambaut, M. U. G. Kraemer. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. *Wellcome Open Res.* **6**, 121 (2021)
- E. Alm, E. K. Broberg, T. Connor, E. B. Hodcroft, A. B. Komissarov, S. Maurer-Stroh, A. Melidou, R. A. Neher, Á. O'Toole, D. Pereyaslov, **WHO European Region sequencing laboratories and GISAID EpiCoV group***. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. *Euro Surveill.* **25** (2020)

PAPER 1. MASS-IMMIGRATION DETERMINES THE ASSEMBLY OF ACTIVATED SLUDGE MICROBIAL COMMUNITIES

G. Dottorini*, **T. Y. Michaelsen***, S. Kucheryavskiy, K. S. Andersen, J. M. Kristensen, M. Peces, D. S. Wagner, M. Nierychlo, P. H. Nielsen

* Contributed equally

Published in Proceedings of the National Academy of Sciences of the United States of America (PNAS)

PAPER 2. THE MICROBIOTA PROFILE IN INFLAMED AND NON-INFLAMED ILEAL POUCH-ANAL ANASTOMOSIS

S. J. Kousgaard, **T. Y. Michaelsen**, H. L. Nielsen, K. F. Kirk, M. Albertsen, O. Thorlacius-Ussing, The Microbiota profile in inflamed and non-inflamed ileal pouch-anal anastomosis.

Published in Microorganisms

PAPER 3. CLINICAL RESULTS AND MICROBIOTA CHANGES AFTER FAECAL MICROBIOTA TRANSPLANTATION FOR CHRONIC POUCHITIS: A PILOT STUDY

S. J. Kousgaard, **T. Y. Michaelsen**, H. L. Nielsen, K. F. Kirk, J. Brandt, M. Albertsen, O. Thorlacius-Ussing

Published in Scandinavian Journal of Gastroenterology

PAPER 4. NOVEL SYNTROPHIC BACTERIA IN FULL-SCALE ANAEROBIC DIGESTERS REVEALED BY GENOME-CENTRIC METATRANSCRIPTOMICS

L. Hao, **T. Y. Michaelsen**, C. Singleton, G. Dottorini, R. Kirkegaard, M. Albertsen, P. H. Nielsen, M. Dueholm.

Published in The ISME Journal

PAPER 5. THE SIGNAL AND THE NOISE – CHARACTERISTICS OF ANTISENSE RNA IN COMPLEX MICROBIAL COMMUNITIES

T. Y. Michaelsen, J. Brandt, C. M. Singleton, R. H. Kirkegaard, J. Wiesinger, N. Segata, M. Albertsen

Published in mSystems

PAPER 6. INTRODUCTION AND TRANSMISSION OF SARS-COV-2 LINEAGE B.1.1.7 (ALPHA) IN DENMARK

T. Y. Michaelsen, M. Bennedbæk, L. E. Christiansen, M. S. F. Jørgensen, C. H. Møller, E. A. Sørensen, S. Knutsson, J. Brandt, T. B. N. Jensen, C. Chiche-Lapierre, E. F. Collados, T. Sørensen, C. Petersen, V. Le-Quy, M. Sereika, F. T. Hansen, M. Rasmussen, J. Fonager, S. M. Karst, R. L. Marvig, M. Stegger, R. N. Sieber, R. Skov, R. Legarth, T. G. Krause, A. Fomsgaard, M. Albertsen, The Danish Covid-19 Genome Consortium (DCGC)

Submitted and in revision, Preprint on medRxiv

LITERATURE LIST

1. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6506–6511 (2018).

2. T. D. Brock, etc., *Biology of microorganisms* (Prentice-Hall, London, England, ed. 6, 1991).

3. J. A. Gilbert, J. D. Neufeld, Life in a world without microbes. *PLoS Biol.* **12**, e1002020 (2014).

4. J. A. Gilbert, M. J. Blaser, J. G. Caporaso, J. K. Jansson, S. V. Lynch, R. Knight, Current understanding of the human microbiome. *Nat. Med.* **24**, 392–400 (2018).

5. M. Blaser, P. Bork, C. Fraser, R. Knight, J. Wang, The microbiome explored: recent insights and future challenges. *Nat. Rev. Microbiol.* **11**, 213–217 (2013).

6. Danish Covid-19 Genome Consortium, covid19genomics.dk. *covid19genomics.dk* (2021), (available at https://www.covid19genomics.dk/home).

7. A. Konopka, What is microbial community ecology? *ISME J.* **3**, 1223–1230 (2009).

8. H.-C. Flemming, J. Wingender, U. Szewzyk, P. Steinberg, S. A. Rice, S. Kjelleberg, Biofilms: an emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).

9. K. Faust, J. Raes, Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).

10. E. E. L. Muller, K. Faust, S. Widder, M. Herold, S. Martínez Arbas, P. Wilmes, Using metabolic networks to resolve ecological properties of microbiomes. *Current Opinion in Systems Biology*. **8**, 73–80 (2018).

11. E. J. Stewart, Growing unculturable bacteria. *J. Bacteriol.* **194**, 4151–4160 (2012).

12. S. C. Schuster, Next-generation sequencing transforms today's biology. *Nat. Methods*. **5**, 16–18 (2008).

13. E. E. Allen, J. F. Banfield, Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* **3**, 489–498 (2005).

14. F. Sanger, G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, C. A. Fiddes, C. A. Hutchison, P. M. Slocombe, M. Smith, Nucleotide sequence of bacteriophage phi X174 DNA. *Nature*. **265**, 687–695 (1977).

15. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blöcker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A.

Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki, International Human Genome Sequencing Consortium, Initial sequencing and analysis of the human genome. *Nature*. **409**, 860–921 (2001).

16. J. Shendure, H. Ji, Next-generation DNA sequencing. *Nat. Biotechnol.* **26**, 1135–1145 (2008).

17. K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51 Pt 1**, 263–273 (1986).

18. D. McDonald, M. N. Price, J. Goodrich, E. P. Nawrocki, T. Z. DeSantis, A. Probst, G. L. Andersen, R. Knight, P. Hugenholtz, An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).

19. J. S. Johnson, D. J. Spakowicz, B.-Y. Hong, L. M. Petersen, P. Demkowicz, L. Chen, S. R. Leopold, B. M. Hanson, H. O. Agresta, M. Gerstein, E. Sodergren, G. M. Weinstock, Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* **10**, 5029 (2019).

20. P. Rausch, M. Rühlemann, B. M. Hermes, S. Doms, T. Dagan, K. Dierking, H. Domin, S. Fraune, J. von Frieling, U. Hentschel, F.-A. Heinsen, M. Höppner, M. T. Jahn, C. Jaspers, K. A. B. Kissoyan, D. Langfeldt, A. Rehman, T. B. H. Reusch, T. Roeder, R. A. Schmitz, H. Schulenburg, R. Soluch, F. Sommer, E. Stukenbrock, N. Weiland-Bräuer, P. Rosenstiel, A. Franke, T. Bosch, J. F. Baines, Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*. **7**, 133 (2019).

21. M. F. Polz, C. M. Cavanaugh, Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**, 3724–3730 (1998).

22. M. Lee, Happy Belly Bioinformatics: an open-source resource dedicated to helping biologists utilize bioinformatics. *J. Open Source Educ.* **2**, 53 (2019).

23. E. A. Franzosa, T. Hsu, A. Sirota-Madi, A. Shafquat, G. Abu-Ali, X. C. Morgan, C. Huttenhower, Sequencing and beyond: integrating molecular "omics" for microbial community profiling. *Nat. Rev. Microbiol.* **13**, 360–372 (2015).

24. M. Tessler, J. S. Neumann, E. Afshinnekoo, M. Pineda, R. Hersch, L. F. M. Velho, B. T. Segovia, F. A. Lansac-Toha, M. Lemke, R. DeSalle, C. E. Mason, M. R. Brugler, Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* **7**, 6589 (2017).

25. N. Segata, D. Boernigen, T. L. Tickle, X. C. Morgan, W. S. Garrett, C. Huttenhower, Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* **9**, 666 (2013).

26. J. Quick, nCov-2019 sequencing protocol V.1. *protocols.io* (2020), (available at https://www.protocols.io/view/ncov-2019-sequencing-protocol-bbmuik6w?version_warning=no).

27. N. Loman, W. Rowe, A. Rambaut, nCoV-2019 novel coronavirus bioinformatics protocol. *ARTIC network* (2020), (available at https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html).

28. G. J. Patti, O. Yanes, G. Siuzdak, Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

29. K. Dettmer, P. A. Aronov, B. D. Hammock, Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **26**, 51–78 (2007).

30. R. Sorek, P. Cossart, Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* **11**, 9–16 (2010).

31. H. Schiebenhoefer, T. Van Den Bossche, S. Fuchs, B. Y. Renard, T. Muth, L. Martens, Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis. *Expert Rev. Proteomics*. **16**, 375–390 (2019).

32. Human Microbiome Jumpstart Reference Strains Consortium, K. E. Nelson, G. M. Weinstock, S. K. Highlander, K. C. Worley, H. H. Creasy, J. R. Wortman, D. B. Rusch, M. Mitreva, E. Sodergren, A. T. Chinwalla, M. Feldgarden, D. Gevers, B. J. Haas, R. Madupu, D. V. Ward, B. W. Birren, R. A. Gibbs, B. Methe, J. F. Petrosino, R. L. Strausberg, G. G. Sutton, O. R. White, R. K. Wilson, S. Durkin, M. G. Giglio, S. Gujja, C. Howarth, C. D. Kodira, N. Kyrpides, T. Mehta, D. M. Muzny, M. Pearson, K. Pepin, A. Pati, X. Qin, C. Yandava, Q. Zeng, L. Zhang, A. M. Berlin, L. Chen, T. A. Hepburn, J. Johnson, J. McCorrison, J. Miller, P. Minx, C. Nusbaum, C. Russ, S. M. Sykes, C. M. Tomlinson, S. Young, W. C. Warren, J. Badger, J. Crabtree, V. M. Markowitz, J. Orvis, A. Cree, S. Ferriera, L. L. Fulton, R. S. Fulton, M. Gillis, L. D. Hemphill, V. Joshi, C. Kovar, M. Torralba, K. A. Wetterstrand, A. Abouellleil, A. M. Wollam, C. J. Buhay, Y. Ding, S. Dugan, M. G. FitzGerald, M. Holder, J. Hostetler, S. W. Clifton, E. Allen-Vercoe, A. M. Earl, C. N. Farmer, K. Liolios, M. G. Surette, Q. Xu, C. Pohl, K. Wilczek-Boney, D. Zhu, A catalog of reference genomes from the human microbiome. *Science*. **328**, 994–999 (2010).

33. C. M. Singleton, F. Petriglieri, J. M. Kristensen, R. H. Kirkegaard, T. Y. Michaelsen, M. H. Andersen, Z. Kondrotaite, S. M. Karst, M. S. Dueholm, P.

H. Nielsen, M. Albertsen, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. *Nat. Commun.* **12**, 2009 (2021).

34. M. Albertsen, P. Hugenholtz, A. Skarshewski, K. L. Nielsen, G. W. Tyson, P. H. Nielsen, Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* **31**, 533–538 (2013).

35. R. M. Bowers, The Genome Standards Consortium, N. C. Kyrpides, R. Stepanauskas, M. Harmon-Smith, D. Doud, T. B. K. Reddy, F. Schulz, J. Jarett, A. R. Rivers, E. A. Eloe-Fadrosh, S. G. Tringe, N. N. Ivanova, A. Copeland, A. Clum, E. D. Becraft, R. R. Malmstrom, B. Birren, M. Podar, P. Bork, G. M. Weinstock, G. M. Garrity, J. A. Dodsworth, S. Yooseph, G. Sutton, F. O. Glöckner, J. A. Gilbert, W. C. Nelson, S. J. Hallam, S. P. Jungbluth, T. J. G. Ettema, S. Tighe, K. T. Konstantinidis, W.-T. Liu, B. J. Baker, T. Rattei, J. A. Eisen, B. Hedlund, K. D. McMahon, N. Fierer, R. Knight, R. Finn, G. Cochrane, I. Karsch-Mizrachi, G. W. Tyson, C. Rinke, A. Lapidus, F. Meyer, P. Yilmaz, D. H. Parks, A. Murat Eren, L. Schriml, J. F. Banfield, P. Hugenholtz, T. Woyke, Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).

36. M. R. Olm, C. T. Brown, B. Brooks, J. F. Banfield, dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).

37. D. Hyatt, G.-L. Chen, P. F. Locascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. **11**, 119 (2010).

38. S. M. Karst, thesis, Aalborg University (2016).

39. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

40. P. J. Turnbaugh, R. E. Ley, M. Hamady, C. M. Fraser-Liggett, R. Knight, J. I. Gordon, The human microbiome project. *Nature*. **449**, 804–810 (2007).

41. J. A. Gilbert, J. K. Jansson, R. Knight, The Earth Microbiome project: successes and aspirations. *BMC Biol.* **12**, 69 (2014).

42. M. Nierychlo, K. S. Andersen, Y. Xu, N. Green, C. Jiang, M. Albertsen, M. S. Dueholm, P. H. Nielsen, MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters

reveals species-level microbiome composition of activated sludge. *Water Res.* **182**, 115955 (2020).

43. R. Knight, A. Vrbanac, B. C. Taylor, A. Aksenov, C. Callewaert, J. Debelius, A. Gonzalez, T. Kosciolek, L.-I. McCall, D. McDonald, A. V. Melnik, J. T. Morton, J. Navas, R. A. Quinn, J. G. Sanders, A. D. Swafford, L. R. Thompson, A. Tripathi, Z. Z. Xu, J. R. Zaneveld, Q. Zhu, J. G. Caporaso, P. C. Dorrestein, Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **16**, 410–422 (2018).

44. B. J. Callahan, P. J. McMurdie, S. P. Holmes, Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639–2643 (2017).

45. M. S. Dueholm, K. S. Andersen, S. J. McIlroy, J. M. Kristensen, E. Yashiro, S. M. Karst, M. Albertsen, P. H. Nielsen, Generation of comprehensive ecosystem-specific reference databases with species-level resolution by high-throughput full-length 16S rRNA gene sequencing and automated taxonomy assignment (AutoTax). *MBio*. **11** (2020), doi:10.1128/mBio.01557-20.

46. S. M. Karst, R. M. Ziels, R. H. Kirkegaard, E. A. Sørensen, D. McDonald, Q. Zhu, R. Knight, M. Albertsen, High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods*. **18**, 165–169 (2021).

47. A. Ramette, Multivariate analyses in microbial ecology. *FEMS Microbiol. Ecol.* **62**, 142–160 (2007).

48. P. H. Nielsen, Microbial biotechnology and circular economy in wastewater treatment. *Microb. Biotechnol.* **10**, 1102–1105 (2017).

49. J. S. Clark, M. Dietze, S. Chakraborty, P. K. Agarwal, I. Ibanez, S. LaDeau, M. Wolosin, Resolving the biodiversity paradox. *Ecol. Lett.* **10**, 647–59; discussion 659-62 (2007).

50. M. Vellend, Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206 (2010).

51. T. Curtis, M. Polchan, J. Baptista, R. Davenport, W. Sloan, Microbial community assembly, theory and rare functions. *Front. Microbiol.* **4**, 68 (2013).

52. I. D. Ofiţeru, M. Lunn, T. P. Curtis, G. F. Wells, C. S. Criddle, C. A. Francis, W. T. Sloan, Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15345–15350 (2010).

53. E. B. Graham, J. C. Stegen, Dispersal-Based Microbial Community Assembly Decreases Biogeochemical Function. *Processes*. **5**, 65 (2017).

54. J. Zhou, D. Ning, Stochastic community assembly: Does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* **81** (2017), doi:10.1128/mmbr.00002-17.

55. D. Ning, Y. Deng, J. M. Tiedje, J. Zhou, A general framework for quantitatively assessing ecological stochasticity. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16892–16898 (2019).

56. M. H. Andersen, thesis, Aalborg University (2021).

57. G. Dottorini, D. S. Wágner, M. Stokholm-Bjerregaard, S. Kucheryavskiy, T. Y. Michaelsen, M. Nierychlo, R. William, P. H. Nielsen, P. H. Nielsen, Full-scale activated sludge transplantation reveals highly resilient community structure. *In preparation* (2021).

58. D. N. Frank, A. L. St Amand, R. A. Feldman, E. C. Boedeker, N. Harpaz, N. R. Pace, Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 13780–13785 (2007).

59. H. Jiang, Z. Ling, Y. Zhang, H. Mao, Z. Ma, Y. Yin, W. Wang, W. Tang, Z. Tan, J. Shi, L. Li, B. Ruan, Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behav. Immun.* **48**, 186–194 (2015).

60. Mayo Clinic, Ulcerative colitis, (available at https://www.mayoclinic.org/diseases-conditions/ulcerative-colitis/symptoms-causes/syc-20353326).

61. W. Szeto, F. A. Farraye, Incidence, prevalence, and risk factors for pouchitis. *Semin. Colon Rectal Surg.* **28**, 116–120 (2017).

62. J. P. Segal, N. S. Ding, G. Worley, S. Mclaughlin, S. Preston, O. D. Faiz, S. K. Clark, A. L. Hart, Systematic review with meta-analysis: the management of chronic refractory pouchitis with an evidence-based treatment algorithm. *Aliment. Pharmacol. Ther.* **45**, 581–592 (2017).

63. C. S. Smillie, J. Sauk, D. Gevers, J. Friedman, J. Sung, I. Youngster, E. L. Hohmann, C. Staley, A. Khoruts, M. J. Sadowsky, J. R. Allegretti, M. B. Smith, R. J. Xavier, E. J. Alm, Strain tracking reveals the determinants of bacterial engraftment in the human gut following fecal Microbiota transplantation. *Cell Host Microbe*. **23**, 229-240.e5 (2018).

64. S. Sørensen, J. Knudsen, T. Michaelsen, C. Bundgaard-Nielsen, R. E. Nielsen, S. Hjerrild, P. Leutscher, G. Wegener, Faecal Microbiota transplantation from

patients with depression into rats modulates mood-related behaviour. *Research Square* (2021), , doi:10.21203/rs.3.rs-671170/v1.

65. L. Hagman, A. Blumenthal, M. Eklund, N. Svensson, The role of biogas solutions in sustainable biorefineries. *J. Clean. Prod.* **172**, 3982–3989 (2018).

66. N. M. C. Saady, Homoacetogenesis during hydrogen production by mixed cultures dark fermentation: Unresolved challenge. *Int. J. Hydrogen Energy*. **38**, 13172–13191 (2013).

67. D. Wu, L. Li, X. Zhao, Y. Peng, P. Yang, X. Peng, Anaerobic digestion: A review on process monitoring. *Renew. Sustain. Energy Rev.* **103**, 1–12 (2019).

68. K. Boe, D. J. Batstone, J.-P. Steyer, I. Angelidaki, State indicators for monitoring the anaerobic digestion process. *Water Res.* **44**, 5973–5980 (2010).

69. N. Müller, P. Worm, B. Schink, A. J. M. Stams, C. M. Plugge, Syntrophic butyrate and propionate oxidation processes: from genomes to reaction mechanisms. *Environ. Microbiol. Rep.* **2**, 489–499 (2010).

70. F. Mosbæk, H. Kjeldal, D. G. Mulat, M. Albertsen, A. J. Ward, A. Feilberg, J. L. Nielsen, Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters by combined protein-based stable isotope probing and metagenomics. *ISME J.* **10**, 2405–2418 (2016).

71. H. Mallick, S. Ma, E. A. Franzosa, T. Vatanen, X. C. Morgan, C. Huttenhower, Experimental design and quantitative analysis of microbial community multiomics. *Genome Biol.* **18**, 228 (2017).

72. F. C. Neidhardt, R. Curtiss, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, Eds., *Escherichia coli and salmonella* (American Society for Microbiology, Washington, D.C., DC, ed. 2, 1996).

73. G. Bao, M. Wang, T. G. Doak, Y. Ye, Strand-specific community RNA-seq reveals prevalent and dynamic antisense transcription in human gut microbiota. *Front. Microbiol.* **6**, 896 (2015).

74. M. K. Thomason, G. Storz, Bacterial Antisense RNAs: How Many Are There, and What Are They Doing?* (2010), doi:10.1146/annurev-genet-102209-163523.

75. D. R. Gelsinger, J. DiRuggiero, The Non-Coding Regulatory RNA Revolution in Archaea. *Genes* . **9** (2018), doi:10.3390/genes9030141.

76. R. Raghavan, D. B. Sloan, H. Ochman, Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio*. **3** (2012), doi:10.1128/mBio.00156-12.

77. V. Lloréns-Rico, J. Cano, T. Kamminga, R. Gil, A. Latorre, W.-H. Chen, P. Bork, J. I. Glass, L. Serrano, M. Lluch-Senar, Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv*. **2**, e1501363 (2016).

78. D. P. Clark, N. J. Pazdernik, *Biotechnology* (Academic Cell, 2016).

79. Statens Serum Institut, Aktuelt og presse. *SSI.dk* (2020), (available at https://www.ssi.dk/aktuelt/nyheder/2020/02_27_foerste-tilfaelde-af-ny-coronavirus-i-dk).

80. J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. **34**, 4121–4123 (2018).

81. GISAID initiative, Submission Tracker Global. *gisaid.org* (2021), (available at https://www.gisaid.org/index.php?id=208).

82. SARS-CoV-2 Interagency Group, SARS-CoV-2 Variant Classifications and Definitions. *CDC* (2021), (available at https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html).

83. European Centre for Disease Prevention and Control, "Risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA – first update" (ECDC, 2021).

84. New and Emerging Respiratory Virus Threats Advisory Group, "NERVTAG meeting on SARS-CoV-2 variant under investigation VUI-202012/01" (2020).

85. N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, W. J. Edmunds, Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*. **372** (2021), doi:10.1126/science.abg3055.

86. H. S. Vohringer, T. Sanderson, M. Sinnott, N. De Maio, T. Nguyen, R. Goater, F. Schwach, I. Harrison, J. Hellewell, C. Ariani, S. Goncalves, D. Jackson, I. Johnston, A. W. Jung, C. Saint, J. Sillitoe, M. Suciu, N. Goldman, E. Birney, S. Funk, E. Volz, D. Kwiatkowski, M. Chand, I. Martincorena, J. C. Barrett, M. Gerstung, The Wellcome Sanger Institute Covid-19 Surveillance Team, The

COVID-19 Genomics UK (COG-UK) Consortium, Genomic reconstruction of the SARS-CoV-2 epidemic across England from September 2020 to May 2021, , doi:10.1101/2021.05.22.21257633.

87. T. Alpert, A. F. Brito, E. Lasek-Nesselquist, J. Rothman, A. L. Valesano, M. J. MacKay, M. E. Petrone, M. I. Breban, A. E. Watkins, C. B. F. Vogels, C. C. Kalinich, S. Dellicour, A. Russell, J. P. Kelly, M. Shudt, J. Plitnick, E. Schneider, W. J. Fitzsimmons, G. Khullar, J. Metti, J. T. Dudley, M. Nash, N. Beaubier, J. Wang, C. Liu, P. Hui, A. Muyombwe, R. Downing, J. Razeq, S. M. Bart, A. Grills, S. M. Morrison, S. Murphy, C. Neal, E. Laszlo, H. Rennert, M. Cushing, L. Westblade, P. Velu, A. Craney, L. Cong, D. R. Peaper, M. L. Landry, P. W. Cook, J. R. Fauver, C. E. Mason, A. S. Lauring, K. St George, D. R. MacCannell, N. D. Grubaugh, Early introductions and transmission of SARS-CoV-2 variant B.1.1.7 in the United States. *Cell* (2021), doi:10.1016/j.cell.2021.03.061.

88. N. L. Washington, K. Gangavarapu, M. Zeller, A. Bolze, E. T. Cirulli, K. M. Schiabor Barrett, B. B. Larsen, C. Anderson, S. White, T. Cassens, S. Jacobs, G. Levan, J. Nguyen, J. M. Ramirez 3rd, C. Rivera-Garcia, E. Sandoval, X. Wang, D. Wong, E. Spencer, R. Robles-Sikisaka, E. Kurzban, L. D. Hughes, X. Deng, C. Wang, V. Servellita, H. Valentine, P. De Hoff, P. Seaver, S. Sathe, K. Gietzen, B. Sickler, J. Antico, K. Hoon, J. Liu, A. Harding, O. Bakhtar, T. Basler, B. Austin, D. MacCannell, M. Isaksson, P. G. Febbo, D. Becker, M. Laurent, E. McDonald, G. W. Yeo, R. Knight, L. C. Laurent, E. de Feo, M. Worobey, C. Y. Chiu, M. A. Suchard, J. T. Lu, W. Lee, K. G. Andersen, Emergence and rapid transmission of SARS-CoV-2 B.1.1.7 in the United States. *Cell* (2021), doi:10.1016/j.cell.2021.03.052.

89. E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, COVID-19 Genomics UK (COG-UK) consortium, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson, Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. **593**, 266–269 (2021).

90. Ekspertgruppen for matematisk modellering, "Notat om prognoser for smittetal og indlæggelser ved scenarier for genåbning af 0.-4. klasse i grundskolen" (2021).

91. Ekspertgruppen for matematisk modellering, "Prognoser for smittetal med fokus på udviklingen i cluster B.1.1.7" (Statens Serum Institut, 2021).

92. S. A. Ishikawa, A. Zhukova, W. Iwasaki, O. Gascuel, A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evol.* **36**, 2069–2085 (2019).

93. Statens Serum Institut, SARS-CoV-2-virusvarianter. *SSI's portal med information om covid-19* (2021), (available at https://covid19.ssi.dk/virusvarianter).

94. P. Bager, J. Wohlfahrt, J. Fonager, M. Rasmussen, M. Albertsen, T. Y. Michaelsen, C. H. Møller, S. Ethelberg, R. Legarth, M. S. F. Button, S. Gubbels, M. Voldstedlund, K. Mølbak, R. L. Skov, A. Fomsgaard, T. G. Krause, Danish Covid-19 Genome Consortium, Risk of hospitalisation associated with infection with SARS-CoV-2 lineage B.1.1.7 in Denmark: an observational cohort study. *Lancet Infect. Dis.* (2021), doi:10.1016/S1473-3099(21)00290-5.

95. F. P. Lyngse, K. Mølbak, R. L. Skov, L. E. Christiansen, L. H. Mortensen, M. Albertsen, C. H. Møller, T. G. Krause, M. Rasmussen, T. Y. Michaelsen, M. Voldstedlund, J. Fonager, N. Steenhard, C. Kirkeby, The Danish Covid-19 Genome Consortium, Increased Transmissibility of SARS-CoV-2 Lineage B.1.1.7 by Age and Viral Load: Evidence from Danish Households, , doi:10.1101/2021.04.16.21255459.

96. S.-H. Lee, H.-J. Kang, H.-D. Park, Influence of influent wastewater communities on temporal variation of activated sludge communities. *Water Res.* **73**, 132–144 (2015).

97. D. Frigon, G. Wells, Microbial immigration in wastewater treatment systems: analytical considerations and process implications. *Curr. Opin. Biotechnol.* **57**, 151–159 (2019).

98. S. Jauffur, S. Isazadeh, D. Frigon, Should activated sludge models consider influent seeding of nitrifiers? Field characterization of nitrifying bacteria. *Water Sci. Technol.* **70**, 1526–1532 (2014).

99. A. Gonzalez-Martinez, A. Rodriguez-Sanchez, T. Lotti, M.-J. Garcia-Ruiz, F. Osorio, J. Gonzalez-Lopez, M. C. M. van Loosdrecht, Comparison of bacterial communities of conventional and A-stage activated sludge systems. *Sci. Rep.* **6**, 18786 (2016).

100. M. Ali, Z. Wang, K. W. Salam, A. R. Hari, M. Pronk, M. C. M. van Loosdrecht, P. E. Saikaly, Importance of species sorting and immigration on the bacterial assembly of different-sized aggregates in a full-scale aerobic granular sludge plant. *Environ. Sci. Technol.* **53**, 8291–8301 (2019).

101. W. T. Sloan, M. Lunn, S. Woodcock, I. M. Head, S. Nee, T. P. Curtis, Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ. Microbiol.* **8**, 732–740 (2006).

102. J. C. Stegen, X. Lin, J. K. Fredrickson, X. Chen, D. W. Kennedy, C. J. Murray, M. L. Rockhold, A. Konopka, Quantifying community assembly processes and identifying features that impose them. *ISME J.* **7**, 2069–2079 (2013).

103. L. Wu, D. Ning, B. Zhang, Y. Li, P. Zhang, X. Shan, Q. Zhang, M. R. Brown, Z. Li, J. D. Van Nostrand, F. Ling, N. Xiao, Y. Zhang, J. Vierheilig, G. F. Wells, Y. Yang, Y. Deng, Q. Tu, A. Wang, Global Water Microbiome Consortium, T. Zhang, Z. He, J. Keller, P. H. Nielsen, P. J. J. Alvarez, C. S. Criddle, M. Wagner, J. M. Tiedje, Q. He, T. P. Curtis, D. A. Stahl, L. Alvarez-Cohen, B. E. Rittmann, X. Wen, J. Zhou, Global diversity and biogeography of bacterial communities in wastewater treatment plants. *Nat. Microbiol.* **4**, 1183–1195 (2019).

104. I. Angriman, M. Scarpa, I. Castagliuolo, Relationship between pouch microbiota and pouchitis following restorative proctocolectomy for ulcerative colitis. *World J. Gastroenterol.* **20**, 9665–9674 (2014).

105. I. Lasa, A. Toledo-Arana, A. Dobin, M. Villanueva, I. R. de los Mozos, M. Vergara-Irigaray, V. Segura, D. Fagegaltier, J. R. Penadés, J. Valle, C. Solano, T. R. Gingeras, Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20172–20177 (2011).

106. J. Wiesinger, C. W. Herbold, T. Y. Michaelsen, M. Palatinszky, M. T. Wolfinger, S. M. Karst, M. Albertsen, P. Pjevac, M. Wagner, Transcriptional response of the ammonia-oxidizing Archaeon ca. nitrososphaera gargensis to growth on cyanate as substrate. *In preparation* (2021).

107. E. B. Hodcroft, N. De Maio, R. Lanfear, D. R. MacCannell, B. Q. Minh, H. A. Schmidt, A. Stamatakis, N. Goldman, C. Dessimoz, Want to track pandemic variants faster? Fix the bioinformatics bottleneck. *Nature*. **591**, 30–33 (2021).

108. COVID-19 GENOMICS UK CONSORTIUM, COVID-19 Genomics (COG) Train. *COG-UK* (2021), (available at https://www.cogconsortium.uk/cog-train/about-cog-train/).

109. McKinsey & Company, When will the COVID-19 pandemic end? *McKinsey & Company* (2021), (available at https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/when-will-the-covid-19-pandemic-end#).

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-996-7

AALBORG UNIVERSITY PRESS