UNIVERSIDADE FEDERAL DO PAMPA
DOUTORADO EM CIÊNCIAS BIOLÓGICAS

PRISCILA CAROLINE THIAGO DOBBLER

O MICROBIOMA HUMANO COMO INDICADOR DE SAÚDE: CORRELAÇÕES E FERRAMENTAS ANALÍTICAS

São Gabriel
2021

PRISCILA CAROLINE THIAGO DOBBLER

O MICROBIOMA HUMANO COMO INDICADOR DE SAÚDE: CORRELAÇÕES E FERRAMENTAS ANALÍTICAS

Tese apresentada ao Programa de Pós-graduação *Stricto Sensu* em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutor em Ciências Biológicas.

Orientador: Luiz Fernando Wurdig Roesch

São Gabriel

2021

Ficha catalográfica elaborada automaticamente com os dados fornecidos
pelo(a) autor(a) através do Módulo de Biblioteca do
Sistema GURI (Gestão Unificada de Recursos Institucionais).

D959m Dobbler, Priscila Caroline Thiago
O MICROBIOMA HUMANO COMO INDICADOR DE SAÚDE: CORRELAÇÕES E FERRAMENTAS ANALÍTICAS / Priscila Caroline Thiago Dobbler.
228p.

Tese(Doutorado)-- Universidade Federal do Pampa, DOUTORADO EM CIÊNCIAS BIOLÓGICAS, 2021.
"Orientação: Luiz Fernando Wurdig Roesch ".

1. Microbiota vaginal no parto. 2. Microbiota intestinal recém-nascido. 3. Bioinformática. 4. Microbioma. I. Título.

PRISCILA CAROLINE THIAGO DOBBLER

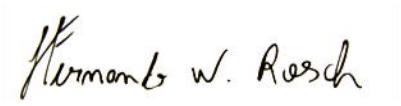
O MICROBIOMA HUMANO COMO INDICADOR DE SAÚDE: CORRELAÇÕES E FERRAMENTAS ANALÍTICAS

Tese apresentada ao Programa de Pós-graduação *Stricto Sensu* em Ciências Biológicas da Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Doutora em Ciências Biológicas.

Área de concentração: Qualidade ambiental

Tese defendida e aprovada em: 19 de março de 2021

Banca examinadora:



Prof. Dr. Luiz Fernando Wurdig Roesch
Orientador
UNIPAMPA

Prof. Dr. Filipe de Carvalho Victoria
UNIPAMPA

Dra. Karina Colonetti
HCPA

Prof. Dra. Luciana Campos Paulino
UFABC

Dedico a minha avó, Conceição Maria, que em vida foi exemplo de força e resiliência.

AGRADECIMENTOS

Agradeço aos meus pais, Angela e Milton, e irmã, Patricia, pelo apoio e incentivo inquestionáveis;

Ao meu orientador pelo apoio, orientação, exigência e parceria durante esses seis anos de pesquisa;

Ao Geferson e Anderson pela companhia, risadas e conversas durante o desenvolvimento deste trabalho;

E por fim, a esta universidade, ao corpo docente e administrativo, pois proporcionaram conhecimento e experiência, fundamentais para a minha jornada profissional;

RESUMO

Por um tempo, micro-organismos foram considerados danosos para a saúde e o funcionamento normal do corpo humano, e com isso, estudava-se principalmente as patologias associadas a esses organismos. Com a evolução da microbiologia, e mais recentemente, o desenvolvimento e expansão das tecnologias de sequenciamento, os micro-organismos passaram a ser vistos sob novas perspectivas. Com o uso do sequenciamento para catalogar comunidades microbianas, é possível identificar de centenas a milhares de micro-organismos em uma única amostra, e com isso se tornou acentuada a interdependência do ser humano com os diferentes tipos de comunidades microbianas. Sendo assim, pesquisas de microbiomas buscam identificar correlações não somente de micro-organismos isoladamente, mas também a nível comunitário, onde há alterações de correlações, ou coocorrência, que leva a desequilíbrios metabólicos dessa comunidade refletindo no hospedeiro. Além disso, gestantes possuem uma comunidade microbiana vaginal característica e distinta da mulheres não-gestantes. E durante a gestação essa comunidade também passa por transformações importantes numa gestação saudável. Durante o parto vaginal, o bebê passa pelo canal vaginal materno, portanto, também entra em contato com a microbiota do local. Com isso, buscamos entender com a microbiota vaginal materna, ao final da gestação, está associada com a microbiota do recém-nascido. Ao realizar esse estudo, percebemos que as ferramentas existes são insuficientes para lidar com a variabilidade intragrupo, a esparsidade dos dados. Por isso, criamos e validamos uma ferramenta chamada PIME (*Prevalence Interval for Microbiome Evaluation*, em português, Intervalo de Prevalência para Avaliação de Microbioma), que é capaz de diminuir a variabilidade intragrupo aplicando diferentes níveis de prevalência para filtragem. PIME também é capaz de, utilizando algoritmo *Random Forests*, classificar as unidades taxonômicas mais importantes para diferenciação entre grupos.

Palavras-Chave: 16S. Diversidade microbiana. Sequenciamento de nova geração. Gravidez. Core microbioma. Bioinformática.

ABSTRACT

For a while, microorganisms were considered harmful to the health and normal functioning of the human body, and the pathologies associated with these organisms were the chief objective of studies. With the evolution of the microbiology field, and more recently, the development and expansion of sequencing technologies, microorganisms came to be seen under a new perspective. The advent of sequencing for cataloging microbial communities has made possible to identify hundreds to thousands of microorganisms in a single sample. This highlighted the interdependency of humans with the different types of microbial communities. Therefore, microbiome research seeks out to identify correlations, not only with microorganisms in isolation, but also at community level, where there are altered correlations, or co-occurrences, that lead to community metabolic unbalancing that reflects on the host. Furthermore, pregnant women harbor characteristic vaginal microbial communities that are distinct from that of non-pregnant women. Also, during a healthy pregnancy, these communities also undergo important transformations. During a vaginal delivery, the baby goes through the birth canal and gets in contact with the local microbiota. Considering this, we sought to understand how the maternal vaginal microbiota, at the end of pregnancy, is associated with the newborn's microbiota. When carrying out this study, we realized that the tools available were not adequate to deal with the intragroup variability, or data sparsity. For this reason, we created and validated a tool called PIME (Prevalence Interval for Microbiome Evaluation) that is capable of reduce intragroup variability by using varying levels of prevalence for filtering. PIME is also capable of classifying the most important taxonomic units for differentiating between groups, by using Random Forests algorithm.

Keywords: 16S. Microbial diversity. Next Generation sequencing. Pregnancy. Core Microbiome.

SUMÁRIO

1. INTRODUÇÃO

No passado, o entendimento de comunidade de micro-organismos não patogênicos foi limitado, pois acreditava-se que os benefícios destes organismos para a saúde tinham pouco efeito comparado com seu potencial patogênico.

O avanço das metodologias de sequenciamento permitiu a descrição mais precisa de comunidades microbianas, e aprimorou nosso entendimento do estabelecimento e funcionamento dessas comunidades. O sequenciamento massivo de regiões especificas, como o gene da região 16S do RNA ribossomal, e o aperfeiçoamento de bancos de dados permitem a identificação e estimativa da abundância de micro-organismos presentes em amostras ambientais (WEINSTOCK, 2012).

O termo microbioma, pela definição mais recente por Berg et al. (2020), pode ser brevemente definido como uma comunidade de micro-organismos característica, presente em um ambiente bem definido e formando um ecossistema dinâmico e integrado, entre si e ao ambiente. O termo microbiota define todos os micro-organismos parte de um microbioma.

O microbioma tem se apresentado como um importante mecanismo para manutenção da saúde e bem-estar dos humanos, e até mesmo durante a gestação. No decorrer da gestação, a microbiota vaginal também passa por alterações e é um fator adicional na manutenção de um prognóstico positivo da gestação (ROMERO et al., 2014a; AVERSHINA et al., 2017). Ainda assim, não há estudos explorando como se apresenta a microbiota vaginal de gestantes brasileiras ao final da gestação, próximo do parto. Outra questão de grande relevância, é de como a microbiota vaginal materna está associada a microbiota intestinal do recém-nascido logo ao nascer e como contribui para a composição microbiana inicial. O primeiro contato do recém-nascido com o mundo externo se dá por meio do canal do parto, em partos vaginais, onde é exposto a microbiota vaginal materna. Esse contato inicial representa um aspecto relevante dado que a colonização microbiana inicial do bebê influencia expressivamente sua saúde ao longo da vida. Além disso, o estabelecimento de uma microbiota disfuncional também é atribuído a problemas no desenvolvimento do sistema imune como alergias e asma (MILANI et al., 2017).

Posto isso, é importante compreender como a microbiota da gestante brasileira se arranja no momento do parto, bem como entender como essa está associada a microbiota do seu recém-nascido. Esse entendimento pode ser relevante para futuros estudos onde se analisa intervenções perinatais, utilizando a microbiota como alvo, que visem o bem-estar materno e infantil.

Com isso, a primeira parte desta tese objetiva-se a analisar a microbiota vaginal de gestantes brasileiras saudáveis, ao final do terceiro trimestre de gestação, e como esta está associada a microbiota do recém-nascido no momento do parto.

Durante o desenvolvimento da pesquisa dessa tese, e de outros projetos desenvolvidos no decorrer do doutorado, percebeu-se que a grande variabilidade individual da microbiota impacta profundamente as análises que buscam encontrar diferenças na composição destas comunidades associadas a um tratamento, característica fenotípica ou ambiental.

Para além disso, amostras de origem humana são especialmente afetadas por essa grande variabilidade interindividual na composição da microbiota. Em muitos casos, essa variabilidade é tão acentuada que em um grupo de amostras, uma única espécie pode apresentar mais que 20% de todas as sequências, enquanto encontra-se ausente em outro (KRAAL et al., 2014). Atualmente utiliza-se pré-filtragens para retirada de sequências em baixa abundância, e também existem algumas ferramentas para comparação de fatores experimentais nos estudos de microbiomas. No entanto, essas abordagens ainda se apresentam insuficientes para solucionar a questão da variabilidade intragrupo. Se bem sucedidas, permitiria conhecer diferenças subjacentes, e correlações não detectados originalmente.

Posto isso, na tentativa de contornar essas dificuldades presentes nos estudos de microbiomas, formulamos a hipótese que a aplicação de níveis de prevalência intragrupo para remoção de micro-organismos contribui para diminuir a variabilidade interna. Com isso, buscamos construir e validar uma ferramenta para filtragem de sequências (unidades taxonômicas) por níveis de prevalência intragrupo.

Por fim, esta tese organiza-se da seguinte forma: Introdução, Referencial Teórico, Artigos, Conclusão, Referências, Apêndices e Anexo. Outros trabalhos e pesquisas desenvolvidos durante o doutorado, em colaboração com outros pesquisadores, são apresentados no formato de artigo e incluídos como Apêndice. Ressalta-se que as Referências ao final da tese, abrangem as citações na estrutura

da tese, Introdução e Referencial Teórico, uma vez que os Artigos têm suas referências apresentadas na sua respectiva composição.

2. REFERENCIAL TEÓRICO

2.1. Ecologia Microbiana

Ecologia microbiana é a ecologia dos micro-organismos, que observa suas interações, entre eles e com o habitat (fatores bióticos e abióticos). Micro-organismos, ou micróbios, são organismos microscópicos que abrangem todos os domínios da vida – *Eucarya*, *Archaea* e *Bacteria* – bem como os vírus, e se apresentam tanto em formas unicelulares ou pluricelulares (TORTORA; FUNKE; CASE, 2000).

Micro-organismos estão presentes em todos os diversos tipos de ambientes que impactam toda biosfera. desde a fixação de carbono, a digestão de alimentos no intestino de mamíferos (KONOPKA, 2009). Com grande variedade genética, metabólica e fisiológica, os processos coletivos chaves dos micro-organismos têm importante papel em regular sistemas biogeoquímicos (incluindo fixação de nitrogênio e metabolismo de metano), e reciclagem de nutrientes pela decomposição (FENCHEL; BLACKBURN; KING, 2012).

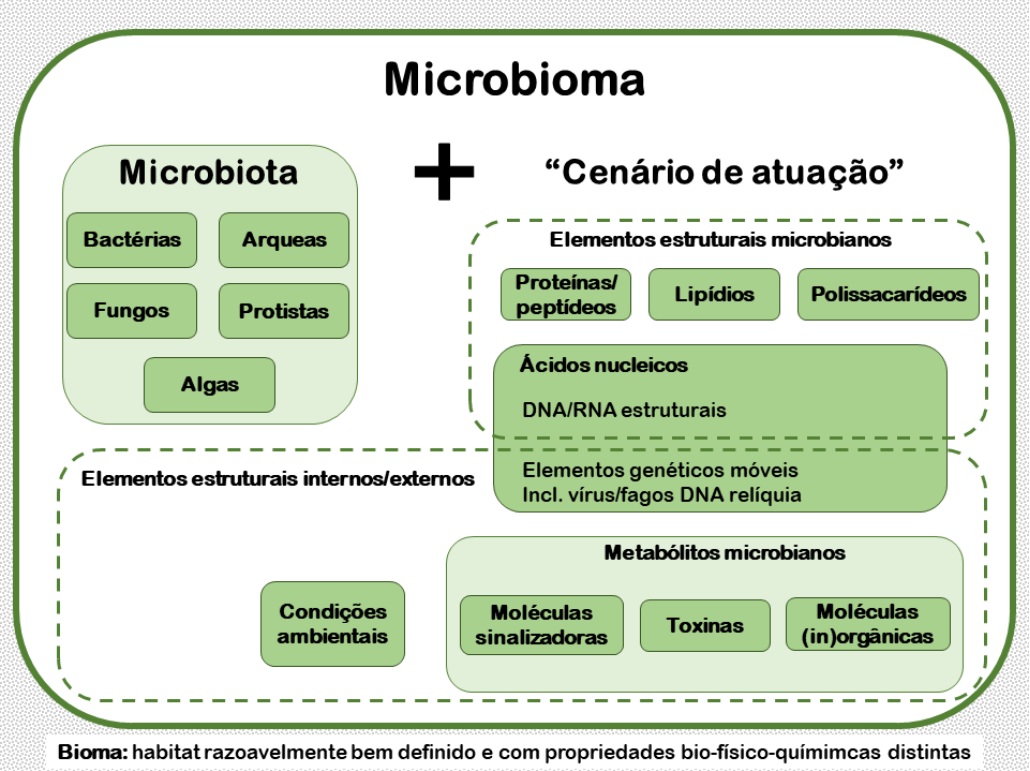
2.2. Microbioma e Microbiota

Historicamente, a área de pesquisa em microbioma emergiu da pesquisa em ecologia microbiana, e oferece uma plataforma interdisciplinar para diversas áreas como agricultura, biotecnologia, matemática (informática, estatística e modelagem), e especialmente, medicina humana.

Comumente, comunidades microbianas têm sido definidas como sendo uma coleção de micro-organismos, ou “montagem multi-espécies”, que interagem em um determinado ambiente. O termo “microbioma” foi primeiramente descrito por Whipps *et al*. (1988), como uma combinação dos termos “micro” e “bioma”, para nomear uma comunidade microbiana característica (micro) em um habitat razoavelmente bem definido e que possui propriedades físico-químicas distintas (bioma). Muitas outras definições para “microbioma” foram publicadas no decorrer dos anos, sendo a de Lederberg e Mccray (2001) a mais citada. Nessa definição, os microbiomas são descritos em um contexto ecológico, como uma comunidade com micro-organismos comensais, simbióticos e patogênicos, dentro de um corpo ou outro ambiente.

Berg *et al*. (2020) propõe um sistema conceitual que abarca a definição original de Whipps *et al.* (1988) e constrói uma definição que considera os novos modelos conceituais em ecologia, definindo o microbioma como uma comunidade microbiana característica ocupando um habitat razoavelmente bem definido com características físico-químicas bem definidas. O microbioma não se refere somente aos micro-organismos envolvidos, mas também o "cenário de atuação" destes, e que resulta na formação de nichos ecológicos específicos. O microbioma, que forma um micro ecossistema dinâmico e pró-ativo sujeito a mudanças em tempo e escala, é integrado num macro ecossistema incluindo hospedeiros eucarióticos, e cruciais para seu funcionamento e saúde (Figura 1).

Figura 1 - Representação da definição de microbioma e microbiota.



Fonte: Adaptado de Berg et al. (2020).

A microbiota consiste na reunião de micro-organismos de diferentes reinos como Procariotos (Bactéria e Arqueia), Eucariotos (e.g. Protozoários, Fungos e Algas).

2.3. Microbioma Humano

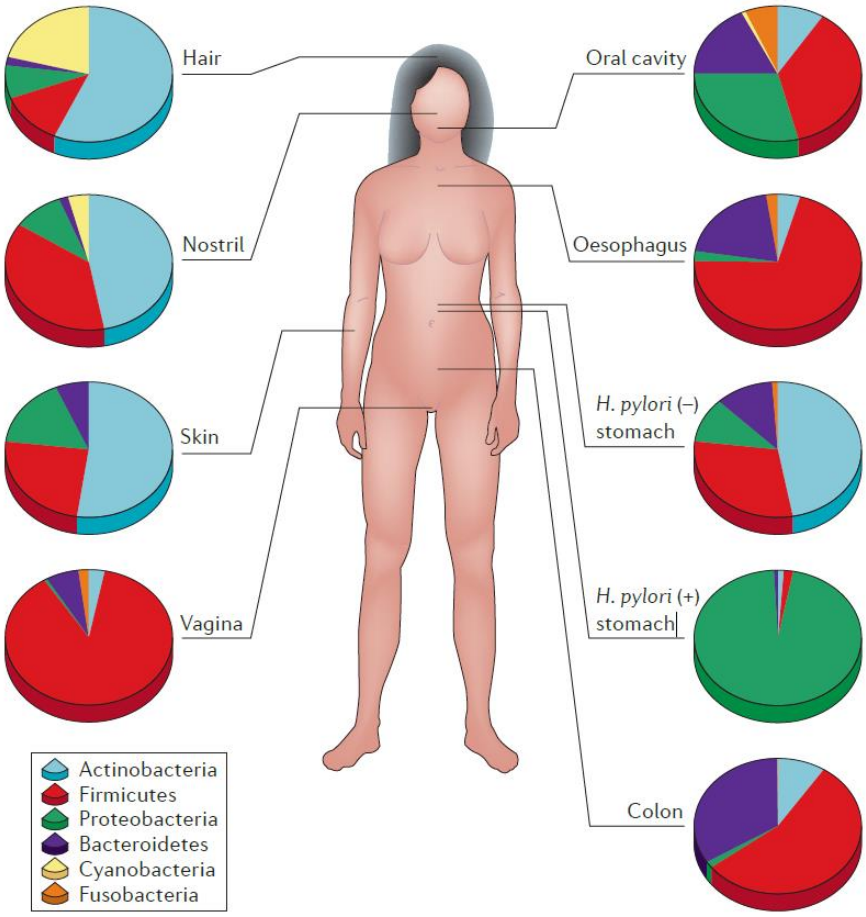
As propriedades da microbiota humana, já chamada de “flora” permaneceram desconhecidas por muito tempo. Novas técnicas moleculares, baseadas em DNA, permitiram pesquisadores desvendar essas informações, que usando somente métodos de cultura não eram acessíveis. Essas técnicas moleculares, sobretudo o de sequenciamento, permitem a identificação de milhões de moléculas de DNA microbiano presentes em uma única amostra, possibilitando a identificação de milhares de espécies de micro-organismos.

Com isso, permite-se uma visão menos enviesada destas comunidades e possibilita a identificação de micro-organismos que são difíceis, ou ainda impossíveis, de serem isolados em laboratório (FIERER et al., 2012). Estima-se que, coletivamente, a microbiota humana possui aproximadamente 10^{13} - 10^{14} células microbianas, com uma razão de 1:1 de células microbianas para células humanas (SENDER; FUCHS; MILO, 2016). E, talvez refletindo a essencialidade das diferentes comunidades, estima-se que a microbiota humana possui cerca de 50 – 100 vezes mais genes que o hospedeiro, servindo assim como uma expansão funcional do genoma hospedeiro. Esses genes extras adicionam uma diversidade de proteínas enzimáticas, não codificadas pelo hospedeiro, contribuindo para a regulação fisiológica e facilitando o metabolismo do hospedeiro (HOOPER; GORDON, 2001).

Estas comunidades associadas ao corpo humano, da pele, ao estômago e intestino, são compostas por diferentes filos e caracterizam microbiotas específicas, como demostrado na Figura 2 (CHO; BLASER, 2012). As diferentes partes do corpo humano podem ser vistas como diferentes pequenos habitats, com diferentes exposições ambientais e demandas energéticas e metabólicas. A diversidade de micro-organismos em um habitat pode ser definida como o número e abundância dos diferentes tipos de organismos. No corpo humano, comunidades microbianas orais e do intestino são especialmente diversas, enquanto que comunidades presentes na vagina são as menos diversas (HUTTENHOWER et al., 2012).

No decorrer do tempo, a composição da microbiota humana sofre menos variação em cada individuo do que entre indivíduos. Essa singularidade da comunidade microbiana de cada individuo aparenta ser estável ao longo do tempo, podendo ser uma característica associada a uma microbiota saudável (HUTTENHOWER et al., 2012).

Figura 2 - Diferenças na composição bacteriana do microbioma humano.



Fonte: CHO; BLASER (2012).

2.3.1. Microbioma Intestinal: homeostase e disbiose

O intestino humano abriga a maior parte das células microbianas do corpo humano e é colonizado por uma diversa comunidade de simbiontes e comensais. O desenvolvimento dessas relações pode ser uma consequência da coevolução adaptativa de humanos e micro-organismos. Relações simbióticas são aquelas onde se apresentam funções metabólicas únicas ou outros benefícios, enquanto que relações comensais são as que nenhuma parte é prejudicada (HOOPER; GORDON, 2001).

O relacionamento simbiótico entre microbiota intestinal e o hospedeiro é regulado e estabilizado por uma rede complexa de interações que englobam troca de sinais entre metabolismo, sistema imune e neuroendócrino. A fermentação de amido e fibras solúveis pela microbiota intestinal produz ácidos graxos de cadeia curta (como butirato, propionato, acetato e lactato). Esses ácidos graxos metabolizados pela microbiota contribuem para 70% da produção de ATP no cólon, contribuindo para a homeostase energética (DONOHOE et al., 2011; KHO; LAL, 2018).

Outra função importante da microbiota intestinal é a de resistência à colonização, onde a microbiota nativa: i) confere proteção ao hospedeiro contra a colonização de organismos patogênicos invasores e ii) controla o super crescimento de micro-organismos patogênicos, membros da microbiota. Membros dominantes da microbiota são essenciais na ocupação de nichos e assim, impede a colonização e crescimento de patógenos (UEMATSU et al., 2008; YIU; DORWEILER; WOO, 2017).

Alterações anormais, ou disbiose, da microbiota intestinal estão associadas a diversos tipos de doenças como infecção por *Clostridium difficile*, Doença Inflamatória Intestinal, e Doença Celíaca.

2.3.2. Microbioma Vaginal Materno

O trato genital inferior feminino abriga comunidades microbianas complexas e essenciais para manutenção da saúde e homeostase. Mesmo complexa em termos de composição microbiana, a microbiota vaginal, de uma mulher não gestante, é dominada por espécies de *Lactobacillus*. *Lactobacillus spp*. tem o ácido lático como principal produto da sua fermentação, o que contribui para manter o ambiente vaginal saudável devido ao efeito antimicrobiano associado a redução do pH (BOSKEY et al., 1999).

Outra função importante de uma microbiota vaginal saudável é o prognóstico de um recém-nascido. A microbiota intestinal dos recém-nascidos também é formada durante o parto, o que influencia a organização inicial da comunidade microbiana. (DOMINGUEZ-BELLO et al., 2010; MILANI et al., 2017).

Durante a gestação, o corpo passa por mudanças hormonais que contribuem para o ganho de peso e modulações das funções imunes que podem estar associadas a mudanças da composição da microbiota materna (NURIEL-OHAYON; NEUMAN;

KOREN, 2016). Diferentemente dos diversos estados de doença, onde alterações na microbiota são correlacionadas a efeitos adversos, as mudanças na microbiota durante a gestação podem contribuir para uma gestação saudável e completa.

No decorrer do primeiro trimestre, a abundância relativa de *Lactobacillus* spp. aumenta enquanto a de outras bactérias anaeróbicas, como *Sneathia*, *Gardnerella*, *Parvimonas*, *Gemella* e *Dialister* diminuem. No decorrer do terceiro trimestre a microbiota vaginal estabiliza, mas com menor diversidade quando comparada com não-gestantes (ROMERO et al., 2014a). No final da gestação, entre a 36ª semana e o parto, a diversidade tende a aumentar, com diminuição discreta de *Lactobacillus spp*. e aumento de outros micro-organismos das famílias *Staphylococcaceae*, *Sphingomonadaceae*, *Pseudomonadaceae*, *Chitinophagaceae*, entre outros (AVERSHINA et al., 2017).

Mesmo com estas variações, a microbiota vaginal tende a ser mais estável em gestantes do que em não gestantes (ROMERO et al., 2014b), podendo indicar que essa estabilidade confere maior resiliência e representa um papel protetivo a infecções ascendentes do trato genital.

2.3.3. Microbiota Intestinal Inicial do Recém-Nascido

Na primeira evacuação do recém-nascido (Mecônio), composto de células epiteliais e resíduos de líquido amniótico ingeridos a partir da 12ª semana gestação, já é possível encontrar uma microbiota um tanto complexa (GOSALBES et al., 2013; HANSEN et al., 2015; MOLES et al., 2013).

A microbiota vaginal materna contribui para a colonização inicial do intestino do recém-nascido e tem grande importância para saúde no curto e longo prazo. A transferência inicial anormal de micro-organismos pode afetar o desenvolvimento do sistema imune, como no desenvolvimento futuro de alergias e asma (JOHNSON; OWNBY, 2017; MILANI et al., 2017), e pode contribuir para complicações pós-natais como sepse precoce (DORNELLES et al., 2020; WORTHAM et al., 2016; MADAN et al., 2012). Durante o parto o bebê entra em contato com essa microbiota vaginal, durante a passagem pelo canal do parto, podendo então ter contato com diferentes micro-organismos, a depender do tipo de comunidade apresentada pela mãe logo antes do parto.

2.4. Sequenciamento como Ferramenta para Estudo da Microbiota

O sequenciamento de marcadores genético amplificados, e.g., o gene 16S do RNA ribossomal, é tradicionalmente usado para teste de hipóteses em composição de comunidades microbianas. O principal desafio em utilizar dados obtidos dessas pesquisas com amplificados é a sua interpretação para a descoberta de reguladores da diversidade microbiana. Aparte de microbiomas simples, como provenientes de ambientes extremos, pesquisas usando amplificados normalmente identificam um elevado número de *taxa* (também chamados de unidades de taxonomia operacional ou variantes de sequenciamento de amplificado que não são compartilhados entre todas amostras (também chamados de *taxa* de baixa prevalência) (SZE; SCHLOSS, 2016).

2.5. Processamento das Sequências e 'Core' Microbiano

Durante o processo de análise de dados provenientes de sequenciamento de amplificados, existem etapas de pré-filtragens, que normalmente removem muitas das *taxas* com baixa prevalência. Uma dessas etapas é a remoção de sequências encontradas somente uma vez em uma amostra. Essas sequências são chamadas de *singletons* (do inglês, "coisa única") (EDGAR, 2013), e de acordo com Tedersoo et al. (2010) esses *singletons* são artifícios e também são considerados como fonte da maior parte do viés no sequenciamento de nova geração. Outra consideração importante, é de que sequências com muito baixa abundância podem ser resultantes de contaminação em baixa escala provenientes de kits comerciais (EISENHOFER et al., 2019).

Outra abordagem durante a pré-filtragem, envolve a exclusão das unidades taxonômicas com baixa prevalência entre todas as amostras. A prevalência de micro-organismos no microbioma humano é caracterizado por padrões de distribuição bem variáveis (KRAAL et al., 2014) com abundância proeminente de algumas cepas em alguns indivíduos enquanto que virtualmente ausente em outros. Mesmo que a

presença de micro-organismos em baixa prevalência, e entre todas as amostras, pode representar um interesse de pesquisa para futuros estudos experimentais (KRAAL et al., 2014), a identificação destes micro-organismos presentes na maioria dos indivíduos, também chamado *core microbiome* (microbioma essencial), tem sido o principal objetivo do Projeto do Microbioma Humano (HUTTENHOWER et al., 2012). A identificação de um *core* microbiano pode ser importante para entender a estabilidade, plasticidade, e funcionamento por agregação complexa de micro-organismos em determinado ambiente. Podendo ainda, ser usado como um padrão para identificar variações significativas que podem estar associadas a estados de doença, outros tratamentos ou variáveis fenotípicas.

Com a exploração mais abrangente do microbioma, desenvolveram-se diversas ferramentas para contrastar fatores experimentais em estudos de microbiomas, como por exemplo Phyloseq (MCMURDIE; HOLMES, 2013), Qiime (CAPORASO et al., 2010), mg-rast (MEYER et al., 2008), Mothur (SCHLOSS et al., 2009), e MicrobiomeAnalyst (DHARIWAL et al., 2017). A escolha por um determinado pacote ou programa para análise é normalmente baseado nas questões de interesse do usuário, nível de experiência em bioinformática (já que alguns programas demandam familiaridade com uso de linhas de comando), e nos recursos disponíveis na instituição do usuário (POLLOCK et al., 2018). Ainda assim, a maior parte das abordagens inseridas nestes pacotes raramente consideram a prevalência microbiana intragrupo com uma opção a ser usada.

Por fim, dados provenientes de estudos de microbioma são desafiadores, com muitas unidades taxonômicas presentes de forma esparsa que leva a grande variação na distribuição. Novas abordagens aplicadas aos dados de microbiomas são imperativos para o desenvolvimento futuro da área.

3. ARTIGOS

Esta tese está organizada na forma de capítulos, onde cada capitulo é apresentado na forma deu um artigo. O Artigo 1 foi publicado na *World Journal of Microbiology and Biotechnology*, e formatado de acordo com as regras desta revista. O Artigo 2 foi publicado na *Molecular Ecology Resources*, e também foi formatado de acordo com as regras desta revista. Os materiais suplementares destes, estão presentes nos endereços localizados no início das respectivas apresentações.

3.1. Artigo 1

A microbiota vaginal de gestantes saudáveis brasileiras e sua correlação com a colonização do intestino do recém-nascido

Publicado na revista *World Journal of Microbiology and Biotechnology*.

Disponível em/DOI: https://doi.org/10.1007/s11274-019-2737-3

Submetido em 30 de julho de 2019, e aceito em 30 de setembro de 2019.

Material Suplementar disponível em:
https://static-content.springer.com/esm/art%3A10.1007%2Fs11274-019-2737-3/MediaObjects/11274_2019_2737_MOESM6_ESM.pdf

The vaginal microbial communities of healthy expectant Brazilian mothers and its correlation with the newborn's gut colonization

Priscila Dobbler[1], Volker Mai[2], Renato S Procianoy[3], Rita C Silveira[3], Andréa L Corso[3], Luiz Fernando Wurdig Roesch[1]

[1] Centro Interdisciplinar de Pesquisas em Biotecnologia – CIP-Biotec, Campus São Gabriel, Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil.
[2] Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA.
[3] Serviço de Neonatologia do Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

Corresponding author:
Luiz Fernando Wurdig Roesch1
Email: luizroesch@unipampa.edu.br

Funding:

This work was supported by the Bill & Melinda Gates Foundation Grant Number OPP1107597, CNPQ and DECIT/Ministério da Saúde do Brasil (Grant Number 401596/2013-3). PD was financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The female lower genital tract harbors a complex microbial community essential for homeostasis and health. During pregnancy, the female body undergoes unique hormonal changes that contribute to weight gain as well as modulations in immune function that can affect microbiota composition. Several studies have described the vaginal microbiota of pregnant women from the USA, Europe and Mexico. Here we expand our knowledge about the vaginal microbial communities during the third trimester to healthy expectant Brazilian mothers. Vaginal samples were collected from patients delivering at the Hospital de Clínicas de Porto Alegre, Brazil. Microbial DNA was isolated from samples and the V4 region of the 16S rRNA gene was amplified and sequenced using the PGM Ion Torrent. Brazilian pregnant women presented three distinct types of microbial community at the time of labor. Two microbial communities, Cluster 1 and Cluster 3, presented an overall dominance of *Lactobacillus* while Cluster 2 tended to present higher diversity and richness, with the presence of *Pseudomonas*, *Prevotella* and other vaginosis related bacteria. About half of the Brazilian mothers sampled here had dominance of *L. iners*. The proportion of mothers without dominance of any *Lactobacillus* was higher in Brazil (22%) compared to UK (2.4%) and USA, where this community type was not detected. The vaginal microbiota showed significant correlation with the composition of the babies' gut microbiota (p-value=0.002 with a R2 of 15.8%). Mothers presenting different vaginal microbiota shared different microorganisms with their newborns, which would reflect on initial colonizers of the developing newborns' gut.

Keywords: 16S rRNA; Microbial diversity; Next Generation Sequencing; Pregnancy; Vaginal Microbiome

1 Introduction

The female lower genital tract harbors a complex microbial community essential for homeostasis and health. Although complex in terms of microbial composition, the healthy vaginal microbiota, in non-pregnant woman, is dominated by *Lactobacillus* species. *Lactobacillus spp*. produce lactic acid as their main fermentation product which contributes to maintaining a healthy vaginal environment by antimicrobial effect associated with a reduced pH (Boskey et al. 1999; Tachedjian et al. 2017). Newborn's health outcome is another important role of a healthy vaginal microbiota. Infant's gut microbial community is shaped also during birth, trough the birth canal, influencing the initial gut microbial community assembly (Dominguez-Bello et al. 2010; Milani et al. 2017).

During pregnancy, the female body undergoes hormonal changes contributing to weight gain as well as modulations in immune function that could be associated with changes in mothers' microbiota composition (Nuriel-Ohayon et al. 2016). In contrast to various disease states, where microbiota alterations correlate with adverse outcomes, microbiota changes during pregnancy might contribute to a healthy full term pregnancy. Throughout the first trimester, the relative abundance of Lactobacillus spp. increases while the abundance of other anaerobic bacteria such as *Sneathia*, *Gardnerella*, *Parvimonas*, *Gemella* and *Dialister* decreases. Towards the last trimester the vaginal microbiota stabilizes but with lower diversity compared to non-pregnant woman (Romero et al. 2014a, b).

The presence of *Lactobacillus spp*. as a member of the healthy vaginal microbiota seems to occur irrespective of geography or racial background, though varying in overall abundance and prevalence. One of the questions this study seeks to answer is if this would also be found on a Brazilian cohort. Several studies have described the vaginal microbiota of pregnant women from the USA, Europe or Mexico (Hernández-Rodríguez et al., 2011; Hyman et al., 2014; MacIntyre et al., 2015; Romero et al., 2014a, 2014b). Romero et al. (2014b) compared the vaginal microbiota of non-pregnant (N=32), 50% African American (AA) with pregnant women (N=22, 86% AA) and monitored microbiota changes throughout a term pregnancy. In another study Romero et al. (2014a) investigated microbiota differences between term (N=72, 86% AA) and preterm (N=18, 94% AA) delivery. Hernández-Rodríguez et al. (2011) described the vaginal microbiota during the third trimester of gestation in 23 pregnant Mexican women. MacIntyre et al., (2015) found a higher proportion of women with

dominance of *L*. *jensenii* in the UK compared to women from USA. Bisanz et al. (2015) also described the vaginal microbiota of 56 pregnant women (53 with term gestation) in a rural region of Tanzania, though there were no description of types of vaginal communities. They found that the majority of the women sampled had dominance of *Lactobacillus spp*. (no species resolution). *Prevotella*, *Gardnerella*, *Sneathia* were also found in lower proportions.

The maternal vaginal microbiota contributes to the colonization of the newborn's gut. Initial infant's gut colonization is very important for early and long-term health. Initial abnormal microbial transfer can affect the immune system development, allergy and asthma future incidence (Johnson and Ownby 2017; Milani et al. 2017), and can contribute to postnatal complications including early onset sepsis (Madan et al. 2012; Wortham et al. 2016). Women with distinct vaginal microbial communities during labor onset might transfer different microbial seeds to their newborn's gut. Thus, understanding how these different vaginal microbial communities are presented during labor, could provide an avenue for developing microbiota-targeting interventions that can improve maternal and newborn's health.

The purpose of this study was to characterize the vaginal microbial community of healthy pregnant Brazilian women at the end of their third trimester, and understand how it correlates with their respective infant's gut microbiota colonization at time of birth. To our knowledge, there are no reports on how the vaginal microbiota of healthy pregnant women from Brazil.

2 Materials and Methods

We performed an observational, cross-sectional study based on a convenience sampling strategy. Participants were recruited at the Neonatology Section of Hospital de Clínicas de Porto Alegre (HCPA), Brazil, between the years of 2014 and 2015. Expectant mothers were enrolled at hospital admission for delivery and provided written informed consent. The study protocol was approved by the Ethics Committee of Hospital de Clínicas de Porto Alegre (HCPA), approval number 39164114.0.0000.532. Exclusion criteria: 1) HIV carrier, 2) recreational drug user or alcohol dependent (self-reported), 3) urinary tract infections, 4) any antibiotic usage during third trimester, 5) gestational diabetes and 6) congenital infections in newborn.

We obtained samples from a total of 45 pregnant women delivering at 37-40 weeks of gestational age and 45 first fecal samples (meconium) from their babies. Samples from 18 women were excluded from the analysis based on: collection after delivery (n=1), lack of records for collection time (n=8), urinary tract infection in the third trimester (n=2), intrapartum antibiotic treatment (n=2), gestational diabetes (n=2), and low sequence coverage, with less than 1000 sequences (n=3). Thus, vaginal samples from 27 expecting mothers were retained for this analysis, and 26 samples of first pass meconium from their respective newborns. All babies, except one, were vaginally delivered. Vaginal samples were collected after hospital admission and shortly before delivery by rotating a sterilized swab five times along the vaginal lumen with a circular motion. Speculum was not used. There were no occurrences of PROM (Premature Rupture of Membranes) or administration of intravenous antibiotics during delivery. Meconium samples were collected within 24h of birth from a single diaper directly into a sterile collection tube. All samples were immediately stored at -80°C for later analysis.

2.1 Microbial DNA extraction, 16S rRNA amplification and library preparation

Microbial DNA isolation from vaginal and meconium samples, amplification of the 16S rRNA, and sequencing protocol were performed following Dobbler et al., (2017, 2018). Raw sequences were deposited in the Sequence Read Archive (SRA), accession SRP093885. Records are accessible at https://www.ncbi.nlm.nih.gov/sra/SRP093885. Run numbers SRR7657414 to SRR7657440.

2.2 Sequence processing and statistical analysis

The 16S rRNA raw sequences were analyzed following the recommendations of the Brazilian Microbiome Project (Pylro et al. 2014) and as previously described (Dobbler et al. 2017). For downstream analysis, the data set was filtered by removing Chloroplast/Cyanobacteria sequences and only OTUs with more than 5 sequence reads were kept before rarefying to the same number of sequences (Lemos et al. 2011). Observed OTU richness and Shannon diversity index estimators

were calculated using the “phyloseq” package (McMurdie and Holmes 2013), and plotted using the “ggpubr” package, both in the R environment. Alpha diversity measurements were tested for normality with Shapiro-Wilk test, and clusters differences were evaluated with the Kruskall-Wallis test. Clinical data was also evaluated, including testing of quantitative variables for normality with Shapiro-Wilk Normality Test. Quantitative variables with normal distribution were compared by the ANOVA test while the non-normal distributed variables were compared by Kruskal-Wallis rank sum test.

We applied an unsupervised clustering approach on the different vaginal microbial communities occurring in Brazilian expectant mothers. First, a Bray-Curtis dissimilarity matrix was built with the OTUs identified in each sample. A Hopkins statistic test was used to verify cluster tendency, followed by the Gap statistical analysis (Tibshirani et al. 2001) to discover the number of clusters in the dataset. Gap statistic was performed with 500 Monte Carlo simulations. The members of each cluster were then identified using k-means with the number of clusters derived from the previous analysis, with 25 different random starting assignments. Analysis was carried out using the “cluster” and “phyloseq” packages (Maechler 2013; McMurdie and Holmes 2013) implemented in R environment.

To test the hypothesis that different vaginal microbial communities occur in healthy Brazilian mothers, Bray-Curtis dissimilarity matrix was ordinated by Multidimensional Scaling (MDS) and differences among community states were tested by Permutational Multivariate Analysis of Variance (PERMANOVA) (Anderson 2001) implemented in the vegan package (Oksanen J et al. 2015), and a pairwise PERMANOVA. Also, in order to identify the main taxa responsible for the differences among each community type, the 30 most abundant OTUs were biploted with the Bray-Curtis dissimilarity matrix in the MDS space and the mean relative abundance were computed for each community type.

High-level phenotype of these microbial communities was investigated through BugBase platform (Ward et al. 2017). For that, the raw 16S rRNA dataset was prepared following the instructions of Langille et al., (2013). After quality filtering and trimming, OTUs were picked against the Greengenes (McDonald et al. 2012) database. Hypothesis testing was performed with Pairwise Mann-Whitney-Wilcoxon Tests.

After exploration of maternal vaginal microbial communities, we also sought to understand whether these different vaginal microbial communities were associated with differences in the newborn's gut microbial community assembly at birth. To accomplish this, OTUs with more than 5 reads were retained and Bray-Curtis dissimilarity and Binary matrices were constructed and ordinated by MDS, where babies' samples were grouped according with their respective mother's cluster. Hypothesis testing was performed with PERMANOVA. Also, in order to visualize how maternal and newborn's sample are clustered, a heatmap was constructed with taxa present in at least 10% in one sample using the 'pheatmap' R package (Kolde 2019).

MetaCoMET (Metagenomics Core Exploration Tool) (Wang et al. 2016) was used to find shared OTUs between mother's clusters and their newborns. An OTU was considered member of a group when the cumulative relative abundance was above 0.1%.

3 Results

3.1 Overall 16S sequencing report and diversity description

After initial quality filtering that retained all OTU's except singletons, Good's coverage at 97% similarity cutoff ranged from 89 to 99% of sequencing coverage (Supplementary Table S1). Further analyses were performed after removing OTUs with less than six sequences across all samples. In all, after quality assessment and pruning of low representative OTUs, 745,688 sequences were retained with a median of 7,236 sequences per sample.

Alpha diversity of the vaginal microbial communities at 37-40 weeks of gestational age varied greatly among mothers. On average, the number of observed OTUs among the subjects was 26, with a minimum of 9 and a maximum of 345 OTUs. The Shannon diversity index ranged from 0.14 to 5.18 with an average of 1.27. The number of phyla and genera also presented great variation among the mothers ranging from one to 15 and 1 to 81 respectively.

Similarly, the newborn's gut microbiota also presented great variation in alpha diversity. The number of OTUs ranged from 12 to 199 with a mean of 74.7 OTUs per sample. Shannon Diversity Index ranged from 0.4 to 4.1 with a mean of 2.1 per sample.

3.2 Determining whether vaginal microbial communities differ among healthy pregnant mothers

To determine whether the vaginal microbial community of Brazilian pregnant woman represented distinct clusters, we applied an unsupervised machine learning approach (Supplementary Table S2). The first step consisted in verifying the cluster tendency using the Hopkins statistic. The Hopkins statistical analysis of the Bray-Curtis dissimilarity distance matrix at OTU level was 0.19, which indicated the presence of clusters. Gap statistic, using the same dissimilarity matrix ordinated by a Multidimensional Scaling space, identified that Brazilian mothers had three distinct vaginal microbiota clusters. With a Gap statistic value of 0.407 and a standard error of 0.037 (Supplementary Table S2).

The k-means clustering function was applied to determine cluster membership. The analysis was carried out using the number of clusters specified by the Gap statistic and with 25 different random starting cluster assignments. K-means clustering method selects the best assignment of cluster members that produces the lowest within cluster variation. Seven pregnant mothers were assigned to the Cluster 1, six pregnant mothers were assigned to Cluster 2 and fourteen pregnant mothers were assigned to Cluster 3.

3.3 Microbial community analysis among clusters

The Shannon Diversity Index and the number of OTUs were used in order to evaluate how alpha diversity compared between the clusters. We found that there was a marginal difference in alpha diversity between the vaginal microbial clusters, global p-value for Shannon Index was 0.069, but no overall difference in number of observed OTUs (p-value = 0.19 (Fig. 1a, 1b). There was a tendency of higher Shannon diversity in Cluster 2 compared to Cluster 3 (p-value = 0.076). There was no difference in Shannon diversity (Fig. 1C) and number of observed OTUs (Fig. 1d) of the microbial community of the babies' gut when considering their respective mothers cluster.

In order to understand how the structure of the microbial communities of each vaginal sample compared between groups. We applied the MDS to the Bray-

Curtis dissimilarity matrix, as a measure of Beta-diversity, and plotted it with taxa abundance. The first two axis alone explained 54.7% of the variation between the microbial communities, and the mothers' vaginal microbial communities formed three clearly distinct groups (Fig. 2). In addition, a PERMANOVA analysis revealed that the microbial composition of the three community types were different (p-value = 0.001, R^2 = 0.50), and indicated that 50% of the distance variation was explained by the different microbial communities' membership (Table 1). All clusters were statistically different (p-adjusted = 0.003) and the largest R^2 was observed between Cluster 3 and Cluster 1, which explained more than 56% of the difference between these communities (Table 1). The large difference observed among clusters (measured by the R^2) indicated this contrast present biological relevance. In addition, analysis of multivariate homogeneity of group dispersions suggests that differences between Cluster 1 and Cluster 3 are not caused by differences in homogeneity of variance (Supplementary Table S3).

3.4 Clinical features of each vaginal community type and microbial composition

After community state assignment, the clinical characteristics of each group were evaluated. We investigated whether gestational age, maternal age, number of pregnancies or number of prenatal visits differed between community states (Table 2). After testing for data normality, Kruskal-Wallis was applied for non-normal data, and Analysis of Variance were applied for data with normal distribution. We found that there was no significant difference in gestational age (p-value = 0.5), mother's age (p-value = 0.54), in the number of pregnancies (p-value = 0.25) or number of prenatal visits (p-value = 0.3801) between the community states. Though, our few samples might not have enough power to detect differences.

Vaginal microbial composition differed greatly between groups. On average mothers assigned to Cluster 1 had dominance of an unidentified species of *Lactobacillus* making up an average of 68.1% of the vaginal microbial community. Also, *L. iners*, *L. antri, L. jensenii*, *Prevotella timonensis*, *P. bivia* and *P. copri* were also detected (Fig. 3). This vaginal community could not be matched to a specific CST (Community State Type) described by Romero et al. (2014a, b). Women grouped into Cluster 2 was marked mainly by very low abundance, or absence, of Lactobacillus spp.

and presence of *P. bivia*, P. copri, P. disiens, *Gardnerella vaginalis* and *Bacteroides* (Fig. 3). This vaginal community matches the CSTIV described by Romero et al. (2014a, b), which has low abundance of Lactobacillus spp. and high frequency and abundance of taxa related to bacterial vaginosis, such as *Gardnerella*, *Prevotella*, and *Atopobium*.

Moreover, more than half the women sampled were grouped into Cluster 3, which had the highest abundance of *L. iners* among the three vaginal community types, with mean abundance of 74.73%. And in addition to dominance of *L. iners*, *G. vaginalis*, *Atopobium vaginae*, *Sneathia sanguinegens* and *Veillonellaceae* were also found (Fig. 3). This description matches the CSTIII described by Romero et al. (2014a, b). Also, *L. iners* was present in all vaginal communities, though, with different overall mean abundance.

Following the characterization of vaginal communities' microbial profiles, we sought to find whether the different microbial composition also reflected on overall differences on hither phenotype characteristics. The BugBase (Ward et al. 2017) platform was applied for phenotype prediction. We found that differences on vaginal microbial composition also reflected on overall community phenotype characteristics. Cluster 1 had higher mean abundance of aerobic bacteria than Cluster 3, 50.19% and 4.99%, respectively (Table 3). Mean abundance of facultative anaerobic was different between all clusters, ranging from 5.78%, for Cluster 2, to 70%, for Cluster 3. There was no significant difference in abundance of anaerobic bacteria between vaginal communities. Also, Cluster 1 and 3 were mainly composed of gram-positive bacteria, 80.56 and 85.38% respectively, while Cluster 2 was composed of mainly gram-negative, 69.18% (Table 3).

3.5 Newborn's gut similarity with mother's vaginal microbial community cluster

After finding and characterizing three different vaginal microbial community clusters in Brazilian mothers, we explored how much of those differences were associated with the composition of babies' gut microbiota at birth.

Beta diversity of infants' samples was measured with Bray-Curtis dissimilarity and Binary distance, ordinated with MDS (Fig. 4), and tested with PERMANOVA. Overall, the three vaginal maternal clusters were sufficient to cluster infants' samples regarding the presence and/or absence (p-value=0.002), however explained little of the variation, with R^2 of 15.8% (Fig. 4a). Bray-Curtis dissimilarity showed no difference, p-value 0.509, between infants from mothers from different clusters (Fig. 4b).

In order to better examine how maternal vaginal microbial composition compared to the newborn's gut composition, we constructed a heatmap with the most abundant taxa across mothers and infants. We found that there was not a clear clustering of these samples, with some common low frequent taxa in low abundance shared between them (Fig. 5), such as Bacteroides, *Clsotridiales* and *Faecalibacterium prausnitzii*. *Pseudomonas lini* and *Prevotella copri* were the shared taxa with the highest abundance in the newborns' gut, and that were also present in the maternal vaginal bacterial community.

Considering the correlation of mothers' microbes with the composition of the babies' gut, we also found that common OTUs between mothers and infants differed between the clusters. Overall, pairs of babies and mothers form Cluster 1 shared 15 different OTUs, while Cluster 2 and 3 shared 25 and 7, respectively. Babies from Cluster 2 had 60.31% of their gut bacterial composition similar with the mother's vaginal microbiota, while babies from the *Lactobacillus* dominated clusters, Cluster 1 and 3, had 32.54 and 50.32%, respectively (Table 4). Babies from Cluster 2 had the highest proportion of *Lactobacillus*, 5.67%, while babies from mothers on Cluster 1 and 3 had only 2.03 and 0.75%. In addition, on average babies from all three clusters had the same two OTUs, the same 200nt sequence, as the most abundant of the shared OTUs and identified as *Pseudomonas lini* and *Prevotella copri*. The 30 most abundant OTUs comprising the gut microbiota of each newborn is presented in Supplementary Fig. S1.

Caution is warranted regarding detection of *Pseudomnoas lini*, as it's commonly found on soil samples, and should be considered part of the "kitome" (Salter et al. 2014).

4 Discussion

The Brazilian population is composed mainly by European, African and Amerindian ancestry. However, race in Brazil does not refer directly to ancestry rather it refers mostly to phenotype, such as skin color. This ambiguity nature of race in Brazil allows for individuals drift from one racial category to another, for example siblings and parents can often identify themselves as member of different racial groups (Telles 2004). This ambiguity is due to the highly miscegenation of the Brazilian population and therefore skin color becomes irrespective of ancestry. Recent research has shown that European ancestry in the Brazilian population is larger than expected, ranging between 60 to 77% depending on the region (Pena et al. 2011; Rodrigues de Moura et al. 2015). However, when Parra et al. (2003) compared white Brazilians with Portuguese (Europeans) and black Brazilians with Africans, they found that these populations were statistically different regarding to the alleles surveyed. They also found extensive overlaps in the African ancestry index among white, intermediate (pardos/brown) and blacks. Therefore, it is imprudent to use the standard stratifications of Caucasian/white and black/African Americans with the Brazilian population. It also strengthens the need to independently investigate populations with high miscegenation rates.

The vaginal microbiota changes rapidly over time, and fluctuations may occur in a matter of days (Gajer et al. 2012). Here we described the vaginal microbiota of pregnant healthy Brazilian mothers, right before delivery. To our knowledge, there is few descriptions of the vaginal microbiota of laboring mothers, immediately before delivery (Martín et al. 2007; Avershina et al. 2017).

A literature review of studies addressing the vaginal microbial communities at third trimester of pregnancy is presented in Table 5. *Lactobacillus* spp. were detected in higher frequency and in higher abundance among all women irrespective of the women background. However, a low proportion of pregnant women presented a vaginal microbial community that was not dominated by *Lactobacillus*. Those women did not present clinical symptoms of vaginosis, their vaginal microbial community was more diverse (greater number of taxa) and presented greater abundance of *Atopobium*, *Pseudomonas*, *Gardnerella* and *Prevotella*.

In this study, we found three different vaginal microbial community assemblies in Brazilian mothers at their third trimester of a healthy gestation. The

Cluster 3 found here, matches the descriptions of the CSTIII (Community State Type) described in other populations, which has dominance of *Lactobacillus iners*. Cluster 2 matches descriptions of the CSTIV, which has low abundance of *Lactobacillus spp*. and high frequency and abundance of taxa related to bacterial vaginosis, such as *Gardnerella*, *Prevotella*, and *Atopobium* (Romero et al. 2014a, b). However, our Cluster 1, dominated by unidentified species of *Lactobacillus spp*., can be matched to any other CSTs dominated by *Lactobacillus spp*. (*L. crispatus*, *L. gaseri*, *L. jensenii*) described in the literature, even though it presented low abundance of *L. antri*, *L. iners*, *L. jensenii*. Nevertheless, despite the similarities of the community clusters dominated by *Lactobacillus* with the others already described, the clusters dominated by *Lactobacillus* found in Brazilian mothers had prevalence, albeit low abundance, of bacterial vaginosis associated bacteria. The majority of the sampled Brazilian mothers presented a cluster dominated by *Lactobacillus spp*. About half (51.8%) of the mothers had dominance of *L. iners*, which are in consonance with both reports (59.4% and 59.1%) from USA (Romero et al. 2014a, b), though reports from UK had lower rates (31%) of dominance of *L. iners*. It is important to highlight the high proportions of black women in both reports from USA (90% and 86%). The prevalence of the diverse cluster in Brazilian mothers was the same as one of the reports form USA, 22%, however it was much higher compared to UK and another report from USA, 2.4% and 0%, respectively (Romero et al. 2014b; MacIntyre et al. 2015). Although our primers are able to amplify *L. crispatus* and *L. gasseri*, these two microbial species were not detected in our samples. They might not be present in our dataset or their abundance was below the detection level of our technique.

During pregnancy, increasing levels of oestrogen lead to the maturation of the vaginal epithelium and accumulation of glycogen, which is broken down into maltose, maltotriose, and maltotetraose supporting *Lactobacillus spp*. colonization (Spear et al. 2014). This increase in oestrogen levels is thought to drive the increase in proportion of Lactobacillus spp. in the vagina throughout pregnancy. On the other hand, Avershina et al. (2017) investigated the vaginal microbiota of women at labor and found that by the time of labor onset the number of observed species are increased. In particular, the phylotypes that are characteristic of CST IV (*Peptoniphilus*, *Anaerococcus*, *Corynebacterium*, *Finegoldia*, *Prevotella*) were overrepresented at labor. This supports our findings, that even the vaginal microbial communities

dominated by *Lactobacillus spp*. had considerable abundance of BV related bacteria at labor onset.

The uterine environment has been considered sterile, in which babies were thought to be born sterile, acquiring their gut microbial community after birth. Recent several studies have described the microbial composition of first pass meconium (Jiménez et al. 2008; Mshvildadze et al. 2010; Madan et al. 2012; Dobbler et al. 2017), placenta (Aagaard et al. 2014) and amniotic fluid (Collado et al. 2016) suggesting that microbial seeding of the fetus gut might occur before birth. Overall, the mothers' vaginal microbial community cluster at time of labor was associated with the microbial presence and absence in the gut microbiota of their newborn at birth, though not strong enough to affect the community structure. Reflecting the different composition of each vaginal community type, common OTUs between mothers and babies were also different.

More than half of the composition of the babies' gut microbiota of mothers from Cluster 2 was found in their mothers, which could be a result of the higher diversity. Even though *Lactobacillus spp*. were most frequently the most abundant in the vagina, it was in very low abundance in the meconium samples, while OTUs identified as *Pseudomonas lini* and *Prevotella* copri were the most frequently shared and abundant in the babies' gut at time of birth. Low resemblance of the newborn gut with the maternal vaginal microbiota have been recently reported in vaginally delivered babies. It was suggested that babies receive microbes from several maternal body sites, though the microbes from maternal gut were more persistent (Ferretti et al. 2018). The OTUs shared between babies and mothers of different clusters, might reflect on initial colonizers of the developing newborns' gut.

The reads obtained by high throughput 16S rRNA gene sequencing surveys represent a random sample of the relative abundance of DNA molecules. According to Gloor et al. (2017), due to the nature of the data it cannot be related to the absolute number of microbes in a sample. The data presenting such random component are referred to as compositional (Aitchison et al. 2000; Gloor and Reid 2016) and the multivariate approaches, usually applied in microbial ecology studies, such as ordination and clustering are considered inappropriate (Pawlowsky-Glahn et al. 2015). While the arguments in favor of compositional analyses are plausible, most tools available for microbiome analysis do not take into account the compositionality of the

data. This opens a discussion on whether or not any other work based on non-compositional models should be rejected. Here we reanalyzed our results using a compositional approach described by Gloor and Reid (2016) by converting 16S rRNA counts using the centered log-ratio (clr) transformation. The results are presented in the supplementary material (Supplementary Data S1) of this manuscript. For our particular dataset the same biological conclusion was reached irrespective of the approach chosen for data analysis.

5 Conclusion

Here, we characterize three different vaginal microbial community types found in Brazilian mothers at time of labor. Two of these community types were dominated by *Lactobacillus spp*. and one was marked by lower abundance of *Lactobacillus spp*. and higher abundance of BV related bacteria. Irrespective of cluster membership, vaginosis related bacteria were frequently found in Brazilian mothers. Other community types were not detected in this cohort and might be due to our small number of women sampled here. In addition, the vaginal microbiota showed significant association with presence of microbes in the babies' gut at the time of birth. On the other hand, high abundance of those vaginal microbes did not correlate with high abundance in the infant's gut microbiota. Overall, maternal vaginal microbiota had low resemblance with initial baby's gut colonization, and maternal vaginal clusters dominated with *Lactobacillus* were not associated with *Lactobacillus* in the babies' meconium at time of birth.

6 References

Aagaard K, Ma J, Antony KM, et al (2014) The placenta harbors a unique microbiome. Sci Transl Med 6:237ra65–237ra65

Aitchison J, Barcelo-Vidal C, Martın-Fernandez JA, Pawlowsky-Glahn V (2000) Logratio Analysis and Compositional Distance. Math Geol 5

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance: NON-PARAMETRIC MANOVA FOR ECOLOGY. Austral Ecol 26:32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Avershina E, Slangsvold S, Simpson MR, et al (2017) Diversity of vaginal microbiota increases by the time of labor onset. Sci Rep 7:17558. doi: 10.1038/s41598-017-17972-0

Bisanz JE, Enos MK, PrayGod G, et al (2015) Microbiota at Multiple Body Sites during Pregnancy in a Rural Tanzanian Population and Effects of Moringa-Supplemented Probiotic Yogurt. Appl Environ Microbiol 81:4965–4975. doi: 10.1128/AEM.00780-15

Boskey ER, Telsch KM, Whaley KJ, et al (1999) Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. Infect Immun 67:5170–5175

Chu DM, Ma J, Prince AL, et al (2017) Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. Nat Med 23:314–326. doi: 10.1038/nm.4272

Collado MC, Rautava S, Aakko J, et al (2016) Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. Sci Rep 6:. doi: 10.1038/srep23129

Dobbler PCT, Laureano ÁM, Sarzi DS, et al (2018) Differences in bacterial composition between men's and women's restrooms and other common areas within a public building. Antonie Van Leeuwenhoek 111:551–561. doi: 10.1007/s10482-017-0976-6

Dobbler PT, Procianoy RS, Mai V, et al (2017) Low Microbial Diversity and Abnormal Microbial Succession Is Associated with Necrotizing Enterocolitis in Preterm Infants. Front Microbiol 8:. doi: 10.3389/fmicb.2017.02243

Dominguez-Bello MG, Costello EK, Contreras M, et al (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci 107:11971–11975. doi: 10.1073/pnas.1002601107

Ferretti P, Pasolli E, Tett A, et al (2018) Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. Cell Host Microbe 24:133-145.e5. doi: 10.1016/j.chom.2018.06.005

Gajer P, Brotman RM, Bai G, et al (2012) Temporal Dynamics of the Human Vaginal Microbiota. Sci Transl Med 4:132ra52. doi: 10.1126/scitranslmed.3003605

Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol 8:. doi: 10.3389/fmicb.2017.02224

Gloor GB, Reid G (2016) Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can J Microbiol 62:692–703. doi: 10.1139/cjm-2015-0821

Hernández-Rodríguez C, et al (2011) Vaginal Microbiota of Healthy Pregnant Mexican Women is Constituted by Four Lactobacillus Species and Several Vaginosis-Associated Bacteria. In: Infect. Dis. Obstet. Gynecol. https://www.hindawi.com/journals/idog/2011/851485/. Accessed 28 Aug 2017

Hyman RW, Fukushima M, Jiang H, et al (2014) Diversity of the Vaginal Microbiome Correlates With Preterm Birth. Reprod Sci 21:32–40. doi: 10.1177/1933719113488838

Jiménez E, Marín ML, Martín R, et al (2008) Is meconium from healthy newborns actually sterile? Res Microbiol 159:187–193. doi: 10.1016/j.resmic.2007.12.007

Johnson CC, Ownby DR (2017) The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. Transl Res 179:60–70. doi: 10.1016/j.trsl.2016.06.010

Kolde R (2019) pheatmap: Pretty Heatmaps

Langille MGI, Zaneveld J, Caporaso JG, et al (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821. doi: 10.1038/nbt.2676

Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW (2011) Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods 86:42–51. doi: 10.1016/j.mimet.2011.03.014

MacIntyre DA, Chandiramani M, Lee YS, et al (2015) The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep 5:. doi: 10.1038/srep08988

Madan JC, Salari RC, Saxena D, et al (2012) Gut microbial colonisation in premature neonates predicts neonatal sepsis. Arch Dis Child-Fetal Neonatal Ed 97:F456–F462

Maechler M (2013) Cluster analysis extended Rousseeuw et al. R CRAN

Martín R, Heilig GHJ, Zoetendal EG, et al (2007) Diversity of the Lactobacillus group in breast milk and vagina of healthy women and potential role in the colonization of the infant gut. J Appl Microbiol 103:2638–2644. doi: 10.1111/j.1365-2672.2007.03497.x

McDonald D, Price MN, Goodrich J, et al (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618. doi: 10.1038/ismej.2011.139

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8:e61217. doi: 10.1371/journal.pone.0061217

Milani C, Duranti S, Bottacini F, et al (2017) The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. Microbiol Mol Biol Rev 81:e00036-17. doi: 10.1128/MMBR.00036-17

Mshvildadze M, Neu J, Shuster J, et al (2010) Intestinal Microbial Ecology in Premature Infants Assessed with Non–Culture-Based Techniques. J Pediatr 156:20–25. doi: 10.1016/j.jpeds.2009.06.063

Nuriel-Ohayon M, Neuman H, Koren O (2016) Microbial Changes during Pregnancy, Birth, and Infancy. Front Microbiol 7:. doi: 10.3389/fmicb.2016.01031

Oksanen J, Blanchet F G, Kindt R, et al (2015) Vegan: community ecology package. R package vegan, vers. 2.2-1

Parra FC, Amado RC, Lambertucci JR, et al (2003) Color and genomic ancestry in Brazilians. Proc Natl Acad Sci U S A 100:177–182. doi: 10.1073/pnas.0126614100

Pawlowsky-Glahn V, Egozcue JJ, Tolosana-Delgado R (2015) Modeling and Analysis of Compositional Data, 1 edition. Wiley, Chichester, West Sussex, UK

Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, et al (2011) The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. PLoS ONE 6:. doi: 10.1371/journal.pone.0017063

Pylro VS, Roesch LFW, Morais DK, et al (2014) Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods 107:30–37. doi: 10.1016/j.mimet.2014.08.018

Rodrigues de Moura R, Coelho AVC, de Queiroz Balbino V, et al (2015) Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. Am J Hum Biol 27:674–680. doi: 10.1002/ajhb.22714

Romero R, Hassan SS, Gajer P, et al (2014a) The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome 2:18. doi: 10.1186/2049-2618-2-18

Romero R, Hassan SS, Gajer P, et al (2014b) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome 2:4. doi: 10.1186/2049-2618-2-4

Salter SJ, Cox MJ, Turek EM, et al (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. doi: 10.1186/s12915-014-0087-z

Spear GT, French AL, Gilbert D, et al (2014) Human α-amylase Present in Lower-Genital-Tract Mucosal Fluid Processes Glycogen to Support Vaginal Colonization by Lactobacillus. J Infect Dis 210:1019–1028. doi: 10.1093/infdis/jiu231

Tachedjian G, Aldunate M, Bradshaw CS, Cone RA (2017) The role of lactic acid production by probiotic Lactobacillus species in vaginal health. Res Microbiol. doi: 10.1016/j.resmic.2017.04.001

Telles EE (2004) Race in Another America: The Significance of Skin Color in Brazil. Princeton University Press

Tibshirani R, Walther G, Hastie T (2001) Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Ser B Stat Methodol 63:411–423. doi: 10.1111/1467-9868.00293

Wang Y, Xu L, Gu YQ, Coleman-Derr D (2016) MetaCoMET: a web platform for discovery and visualization of the core microbiome. Bioinformatics 32:3469–3470. doi: 10.1093/bioinformatics/btw507

Ward T, Larson J, Meulemans J, et al (2017) BugBase Predicts Organism Level Microbiome Phenotypes. bioRxiv 133462. doi: 10.1101/133462

Wortham JM, Hansen NI, Schrag SJ, et al (2016) Chorioamnionitis and Culture-Confirmed, Early-Onset Neonatal Infections. Pediatrics 137:. doi: 10.1542/peds.2015-2323

FIGURES

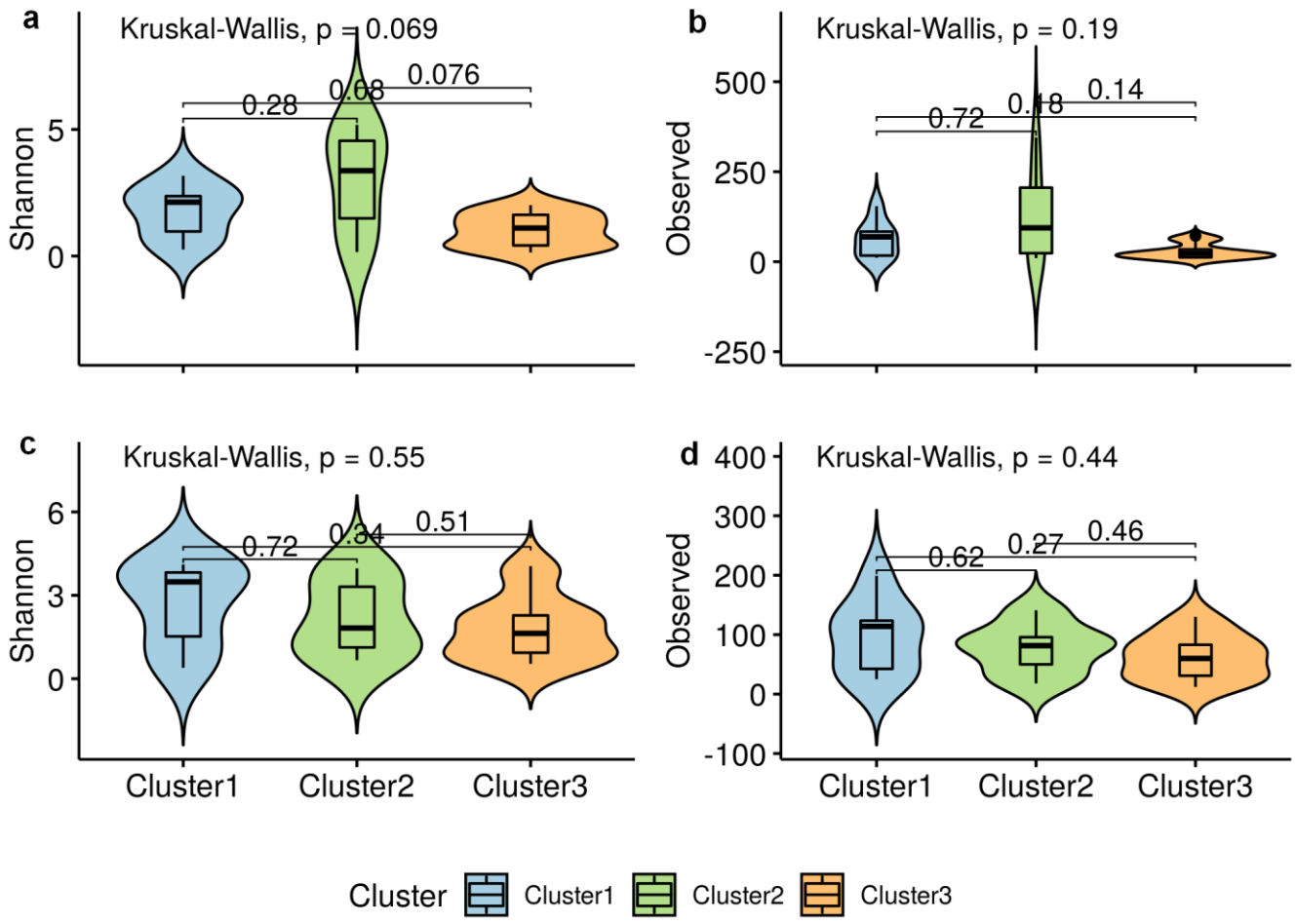


Fig. 1. Violin plot representing alpha diversity measurements. (A) Shannon Diversity Index and (B) Observed OTUs of maternal vaginal samples. (C) Shannon Diversity Index and (D) Observed OTUs of newborns' meconium samples. Colors indicate cluster classification, babies samples are colored according to mother's cluster. Observed stands for the number of OTUs found in each cluster and Shannon stands for Shannon Diversity Index. Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median black dots represents all samples in each group and red dots represent outliers.

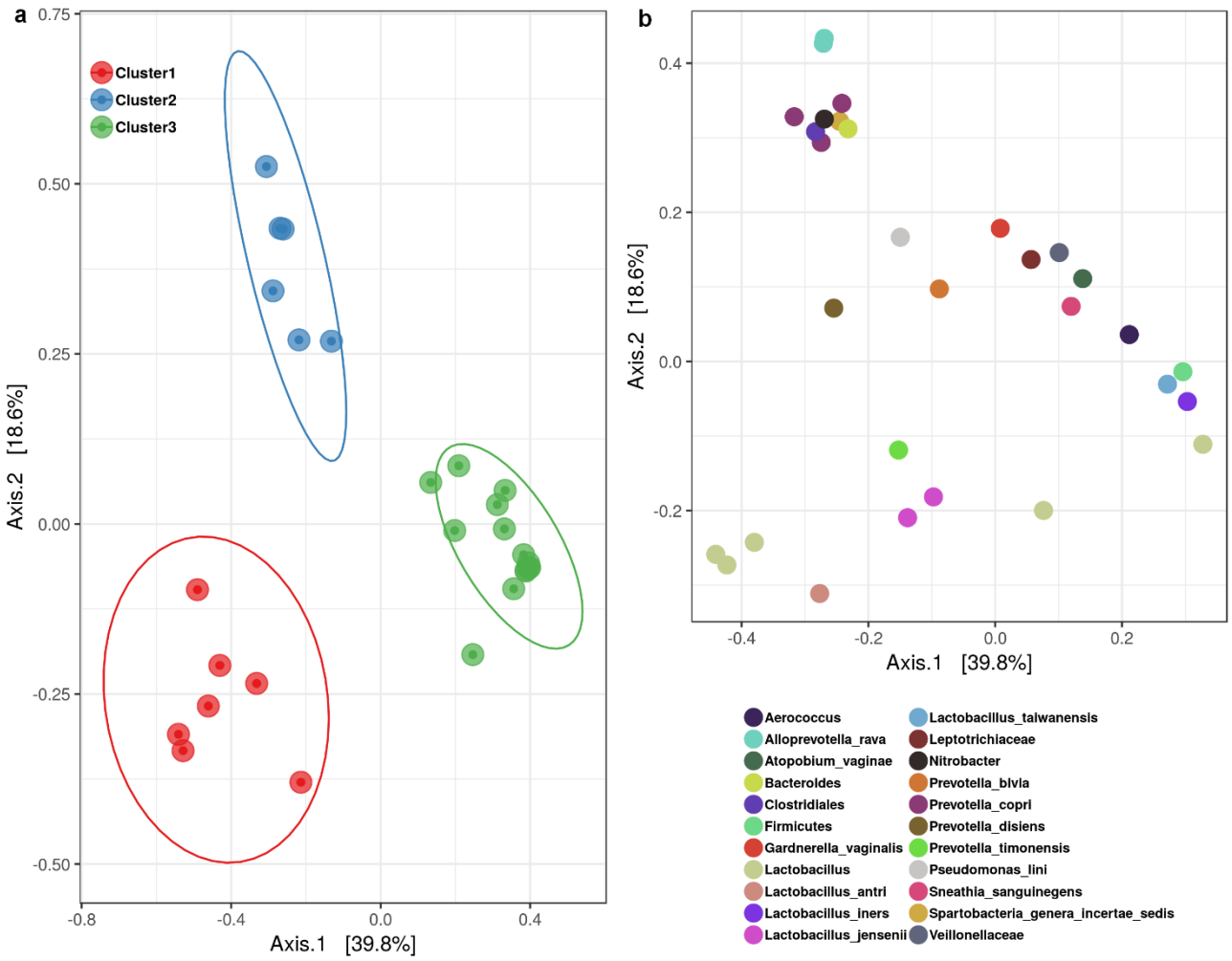


Fig. 2. **Multidimensional Scaling of the Bray-Curtis distance of vaginal samples.** Each symbol represents a microbial community of an individual sample. (A) Presents the different clusters of vaginal microbial communities. Each color represents a cluster, large circles around samples represents a confidence ellipse of 95%. (B) Represents the 30 most abundant OTUs across all vaginal samples, summarized at the highest taxonomy level with at least 80% confidence, into 22 different taxa. Each circle represents a different OTU, while different colors represent different taxonomy assignments and indicate which OTU/taxa are driving sample clustering.

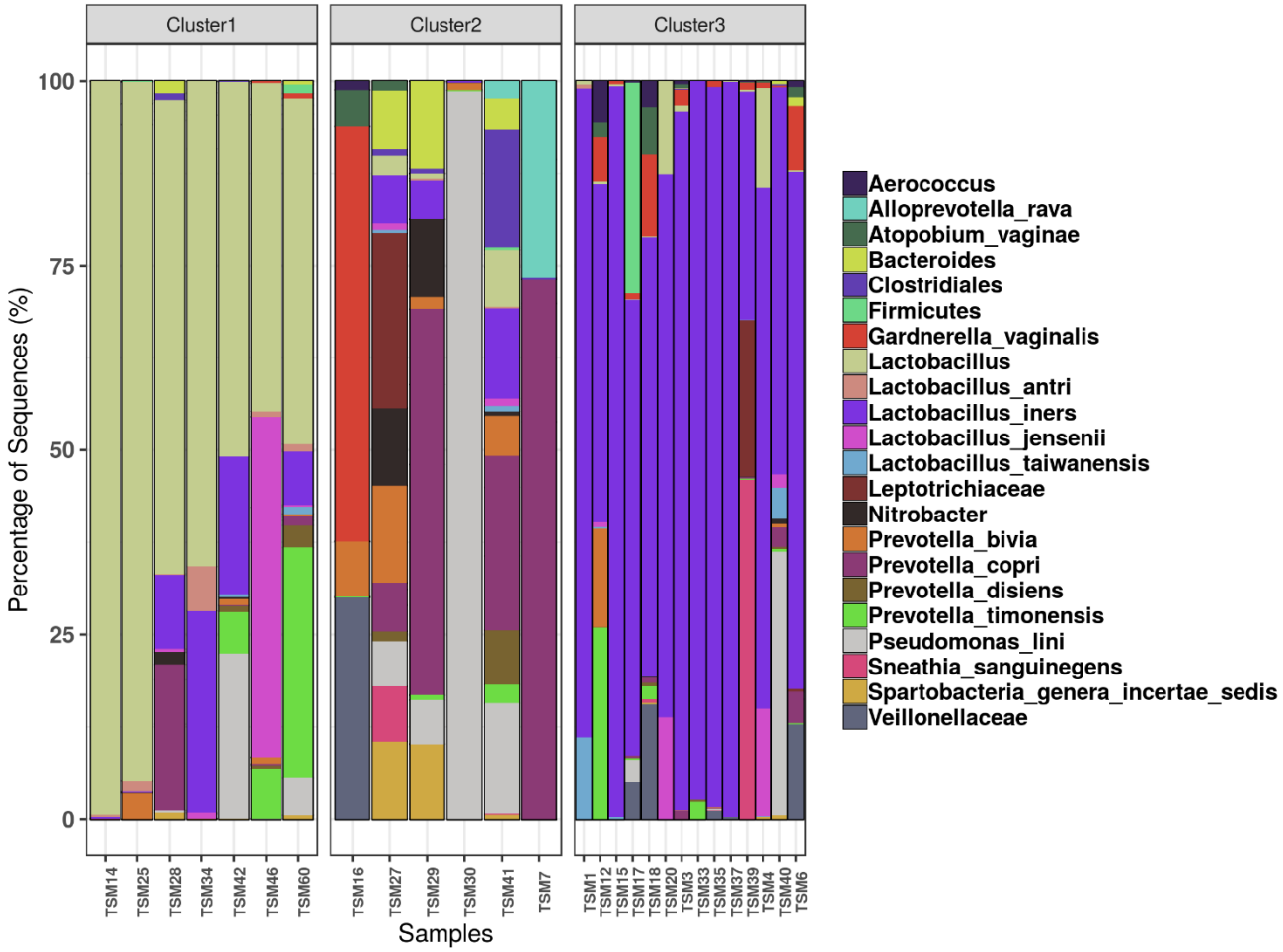


Fig. 3. Bar plot presenting the relative abundance of the 30 most abundant OTUs of vaginal samples. OTUs were summarized at the highest taxonomy level with at least 80% confidence, into 22 different taxa. Each stacked bar represents the relative abundance of each vaginal maternal sample, grouped according to its respective community cluster.

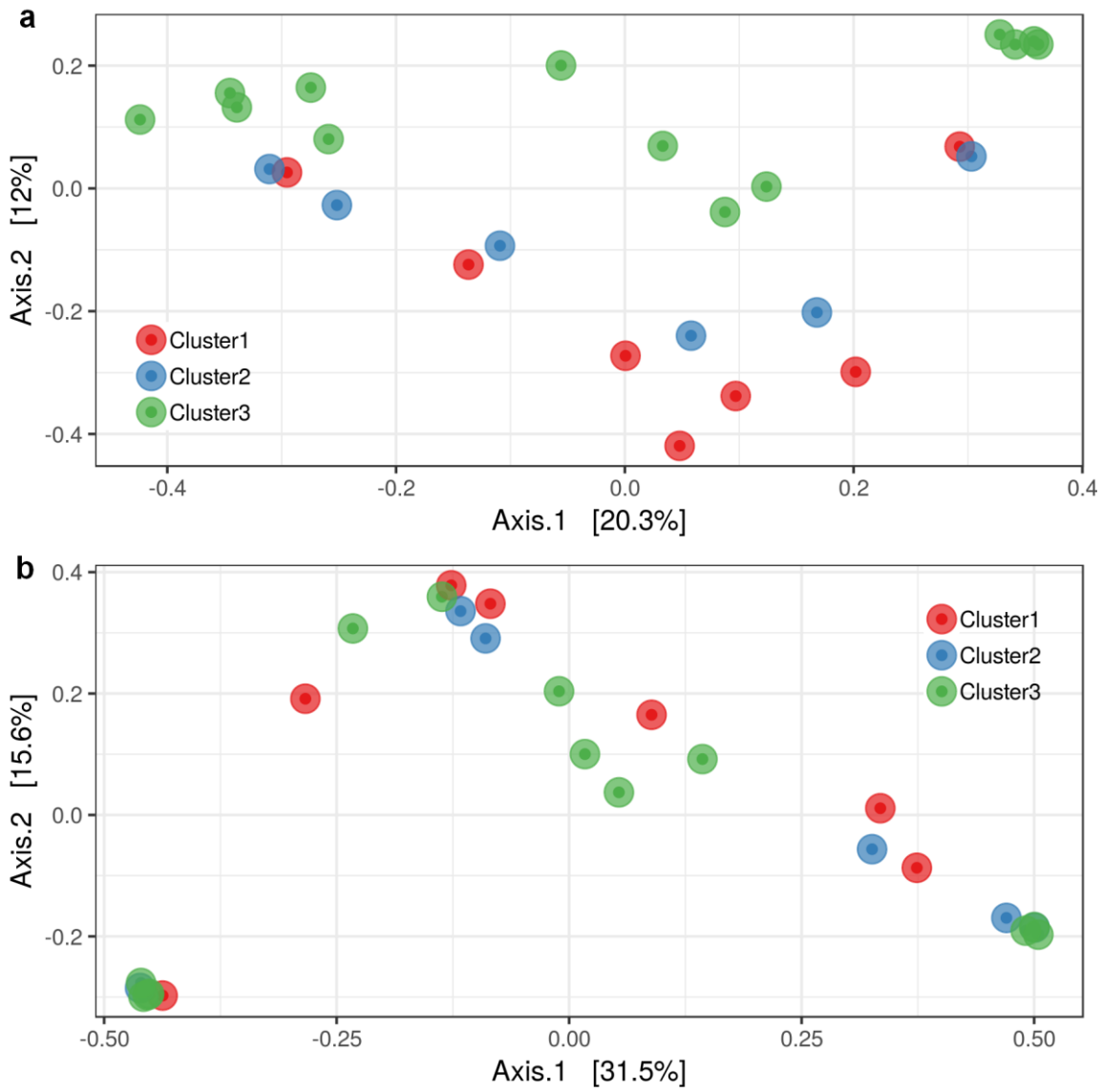


Fig. 4. Beta diversity ordinated with MDS of microbial communities present in meconium samples. Each symbol represents a microbial community of an individual sample and each color represents a cluster assigned to newborns' respective mother. (A) Binary distance of microbial communities, based on presence and absence. (B) Bray-Curtis dissimilarity.

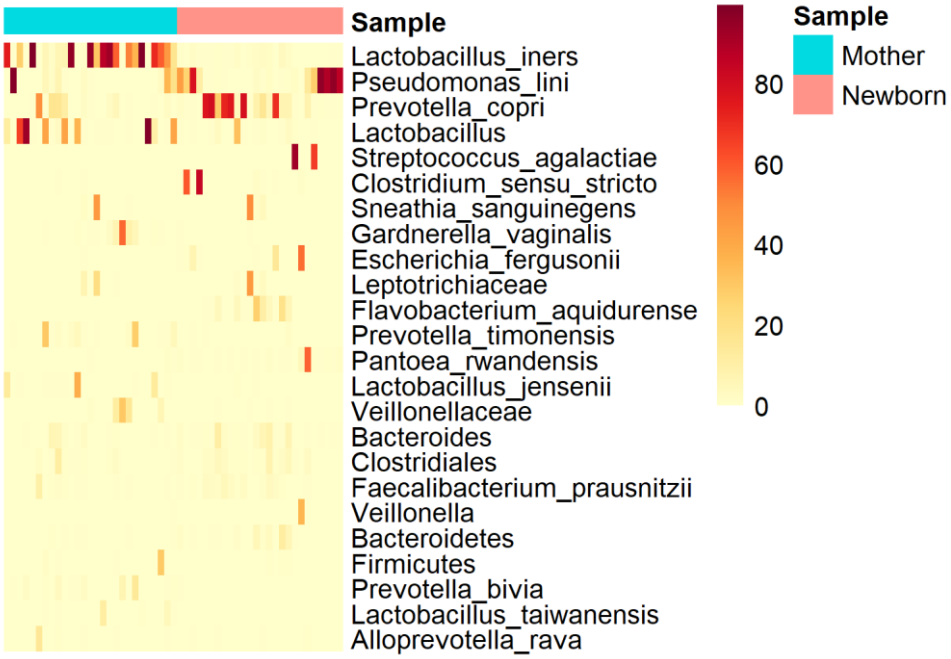


Fig. 5. Heatmap with the most abundant taxa across maternal and infant microbiota.

Each line represents a taxon, summarized at the highest taxonomy level, and each column represents an individual sample. These are the taxa with more than 10% relative abundance in at least one sample.

TABLES

Table 1. Permutational Multivariate Analysis of Variance among the vaginal microbial communities of different community clusters found in this study.

	DF	SS	F. Model	R^2	P-*value*	P-adjusted
Clusters	2	4.283	12.245	0.50505	0.001	-
Residuals	24	4.197	-	0.49495	-	-
Total	26	8.481	-	1.0	-	-
Pairwise Clusters Comparisons						
Cluster 3 vs Cluster 2			10.873101	0.3765824	0.001	0.003*
Cluster 3 vs Cluster 1			24.232420	0.5605150	0.001	0.003*
Cluster 2 vs Cluster 1			5.125949	0.3178696	0.001	0.003*

*P-value adjusted for multiple comparisons with Bonferroni

Table 2. Sample group characteristics summarized according to the different vaginal microbial communities found in this study.

	Cluster 1 (n=7)	Cluster 2 (n=6)	Cluster 3 (n=14)	p-values
Mothers' Characteristics				
Gestational Age (weeks)***	39.77 ± 1.0	39.14 ± 1.1	39.73 ± 1.1	0.5*
Mother's Age (years)	27.14 ± 7.4	24.33 ± 3.6	23.71 ± 6.0	0.3931**
Number of Pregnancies	2.3 ± 1.1	2.3 ± 0.5	1.7 ± 0.9	0.2308**
Prenatal Visits	6.71 ± 2.14	8.17 ± 1.72	7.14 ± 1.92	0.3801**
Newborns' Characteristics				
Weight at Birth (grams)	3362.86 ± 414.82	3370.83 ± 279.15	3924.43 ± 270.78	0.838*
Length (cm)	48.92 ± 1.5	48.5 ± 1.61	48.54 ± 0.82	0.782*
APGAR 1	8.29 ± 1.25	8.67 ± 0.52	8.21 ± 2.33	0.7581**
APGAR 5	9.14 ± 0.69	9.67 ± 0.52	9.29 ± 0.83	0.3412**
Head Circumference (cm)	33.17 ± 1.51	33.5 ± 1	34.14 ± 1.51	0.337*
Thoracic Circumference (cm)	33.67 ± 1.37	33.58 ± 1.24	33.54 ± 1.25	0.978*

Values expressed in means and standard deviations of the mean.

*ANOVA

**Kruskal-Wallis rank sum test

*** Gestational age for delivery time and sample collection time

Table 3. Predicted phenotypes of different vaginal microbial cluster found in this study.

	Mean abundance (%)			p-value		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1 vs Cluster 2	Cluster 1 vs Cluster 3	Cluster 2 vs Cluster 3
Aerobic	50.19	34.20	4.99	0.2948	0.00003*	0.2391
Anaerobic	13.35	53.09	24.21	0.1014	0.8557	0.0757
Facultative Anaerobic	28.69	5.78	70.13	0.0011*	0.0022*	0.00005*
Gram-Negative	19.44	69.18	14.62	0.0081*	0.5352	0.0006*
Gram-Positive	80.56	30.82	85.38	0.0081*	0.5352	0.0006*
Mobile Elements	11.28	19.15	2.87	0.945	0.0007*	0.1093

*Pairwise Mann-Whitney-Wilcoxon Test

Table 4. Relative mean abundance of shared OTUs between vaginal and meconium samples within each cluster.

OTU ID	Taxonomy	Meconium	Vaginal
Cluster 1	15 Shared Phylotypes		
OTU1	Pseudomonas_lini	10.97%	3.46%
OTU2	Prevotella_copri	8.10%	1.01%
OTU3	Prevotella_copri	3.53%	0.46%
OTU4	Bacteroides	3.08%	0.11%
OTU5	Prevotella_copri	2.23%	0.21%
OTU6	Lactobacillus	1.09%	55.78%
OTU7	Nitrobacter	1.02%	0.25%
OTU8	Lactobacillus_iners	0.77%	8.07%
OTU9	Alistipes	0.45%	0.06%
OTU10	Lachnospiraceae	0.32%	0.22%
OTU11	Prevotella_timonensis	0.28%	5.52%
OTU12	Spartobacteria_genera_incertae_sedis	0.21%	0.08%
OTU13	Firmicutes	0.21%	0.21%
OTU14	Lactobacillus	0.17%	1.65%
OTU15	Flavobacterium	0.11%	0.18%
Cluster 2	25 Shared Phylotypes		
OTU1	Pseudomonas_lini	32.53%	18.02%
OTU2	Prevotella_copri	7.73%	4.22%
OTU6	Lactobacillus	4.66%	0.44%

OTU5	Prevotella_copri	3.06%	1.24%
OTU3	Prevotella_copri	2.43%	6.73%
OTU4	Bacteroides	1.68%	1.37%
OTU16	Bacteroides_uniformis	1.31%	0.44%
OTU8	Lactobacillus_iners	1.01%	1.36%
OTU17	Faecalibacterium_prausnitzii	0.90%	0.88%
OTU18	Clostridiales	0.88%	1.42%
OTU19	Parabacteroides_merdae	0.69%	0.16%
OTU20	Bacillus_bataviensis	0.41%	0.25%
OTU21	Prevotella_copri	0.34%	0.23%
OTU9	Alistipes	0.34%	0.33%
OTU22	Bacillus	0.31%	0.23%
OTU23	Spartobacteria_genera_incertae_sedis	0.29%	0.98%
OTU32	Ruminococcaceae	0.28%	0.10%
OTU24	Faecalibacterium_prausnitzii	0.26%	0.44%
OTU25	Prevotella_copri	0.26%	0.16%
OTU26	Leptotrichiaceae	0.23%	0.93%
OTU27	Gp1	0.20%	0.36%
OTU11	Prevotella_timonensis	0.18%	0.41%
OTU28	Pseudomonas	0.16%	0.57%
OTU29	Faecalibacterium_prausnitzii	0.11%	0.72%
OTU30	Bacteroides_coprocola	0.05%	0.33%
Cluster 3	7 Shared Phylotypes		
OTU1	Pseudomonas_lini	22.50%	2.68%
OTU2	Prevotella_copri	14.93%	0.19%
OTU3	Prevotella_copri	4.43%	0.46%
OTU31	Sneathia_sanguinegens	4.00%	3.24%
OTU26	Leptotrichiaceae	3.70%	1.54%
OTU8	Lactobacillus_iners	0.57%	69.82%
OTU6	Lactobacillus	0.18%	0.67%

Table 5. Review of studies addressing the vaginal microbial communities at the third trimester of pregnancy.

Authors	Country	Type of study and methodology	Study Characteristics	Most abundant Taxa
Avershina et al. (2017)	Norway	Randomized double blind clinical trial V3-V4 hypervariable region of the 16 S rRNA	256 pregnant women with term gestation	*L. iners, L. crispatus, Enterobacteriaceae and Prevotella*
Chu et al. (2017)	USA	Prospective cohort study V5-V3 hypervariable region of the 16S rRNA gene	81 pregnant women with term gestation	*Lactobacillus* spp*., Prevotella* spp*., Streptococcus* spp*., Corynebacterium* spp*.*
MacIntyre et al. (2015)	UK	Longitudinal study V1-V2 hypervariable regions of 16S rRNA gene	42 pregnant women with term gestation; 23 (54.8%) White, 5 (11.9%) Black; 13 (31%) Asian	*Lactobacillus crispatus, L. iners, L. jensenii, L. gasseri, Prevotella* spp.
Bisanz et al. (2015)	Tanzania	Longitudinal open-label study V4 hypervariable region of 16S rRNA gene	56 pregnant women, 53 with term gestation	*Lactobacillus* spp., *Prevotella* spp., *Gardnerella* spp., *Sneathia* spp.
Romero et al. (2014a)	USA	Retrospective case–control longitudinal study V1-V2 hypervariable regions of 16S rRNA gene	22 pregnant women with term gestation; 19 (86%) African American, 2 (9%) White, 1 (5%) Hispanic	*L. iners, L. crispatus, Atopobium vaginae, Lactobacillus, L. Jensenii*
Romero et al. (2014b)	USA	Nested case–control study V1-V3 hypervariable regions of 16S rRNA gene	72 pregnant women with term gestation; 62 (86.1%) African American, 4 (5.6%) White, 6 (8.3%) Others	*L. iners, L. crispatus, Gardnerella vaginalis, L. Jensenii,* BVAB1

Hernández-Rodríguez et al. (2011)	Mexico	Transversal study V3 hypervariable region of 16S rRNA gene	23 pregnant women with term gestation	*L. acidophilus*, *L. iners*, *Ureaplasma urealyticum*, *L. gasseri*, BVAB1
This work	Brazil	Cohort study V4 hypervariable region of 16S rRNA gene	27 pregnant women with term gestation	*L. iners*, *Lactobacillus*, *Pseudomonas lini*, *G. vaginalis*, *Prevotella copri*

3.2. Artigo 2

PIME: um pacote para descoberta de novas diferenças entre comunidades microbianas.

Publicado na revista *Molecular Ecology Resources*

Submetido em 10 de junho de 2019, e aceito em 4 de novembro de 2019.

Disponível em/DOI: https://doi.org/10.1111/1755-0998.13116

Material suplementar em:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2F1755-0998.13116&file=men13116-sup-0001-Supinfo.pdf

PIME: a package for discovery of novel differences among microbial communities

Luiz Fernando W. Roesch[1], Priscila Thiago Dobbler[1], Victor Satler Pylro[2], Bryan Kolaczkowski[3], Jennifer C. Drew[3] and Eric W. Triplett[3]

[1]Interdisciplinary Research Center on Biotechnology-CIP-Biotec, Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil.
[2]Microbial Ecology and Bioinformatics Lab., Department of Biology, Universidade Federal de Lavras, MG, Brazil.
[3]Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA.

Abstract

The data used for profiling microbial communities is usually sparse with some microbes having high abundance in a few samples and being nearly absent in others. However, current bioinformatics tools able to deal with this sparsity are missing. PIME (Prevalence Interval for Microbiome Evaluation) was designed for remove those taxa that may be high in relative abundance in just a few samples but have a low prevalence overall. The reliability and robustness of PIME were compared against existing methods and tested using 16S rRNA independent datasets. PIME filters microbial taxa not shared in a per treatment prevalence interval starting at 5% prevalence with increasing increments of 5% at each filtering step. For each prevalence interval, hundreds of decision trees are calculated to predict the likelihood of detecting differences in treatments. The best prevalence-filtered dataset is user-selected by choosing the prevalence interval that keeps a large portion of the 16S rRNA sequences in the dataset while also showing the lowest error rate. To obtain the likelihood of introducing type I error while building prevalence-filtered datasets, an error detection step based is also included. A PIME reanalysis of published datasets uncovered other expected microbial associations then previously reported, which may be masked when only relative abundance was considered.

Keywords: microbial prevalence, Next Generation Sequencing, 16S-rRNA, taxa filtering, Core microbial taxa, microbial biomarkers.

1 INTRODUCTION

Sequencing of amplified genetic markers (amplicon survey), *e.g.* the 16S rRNA gene, is traditionally used for testing hypotheses on microbial community composition. The major challenge for using the data obtained by these surveys is their interpretation for the discovery of the drivers of microbial diversity. Excluding microbiomes from simple ecosystems (e.g. habitats with extreme temperature or pH), amplicon surveys usually identify a large number of taxa (also called Operational Taxonomy Units – OTUs or Amplicon Sequence Variants – ASVs) not shared among all samples (also called low prevalent taxa) (Sze & Schloss, 2016). Often pre-filtering steps in the data analysis eliminate many of these taxa with low prevalence. Those steps include, but are not limited to, the exclusion of sequences found only once in a sample. These are the so-called singletons (Edgar, 2013; Edgar & Flyvbjerg, 2015). According to Tedersoo et al., (2010), singletons are artifactual and account for the greatest source of bias in Next Generation Sequencing. Also, very low abundant reads might be the result of a low level of contaminants from commercial kits (Eisenhofer et al., 2019; Salter et al., 2014).

Another pre-filtering approach involves the exclusion of microbial taxa of low prevalence across all samples. The prevalence of microbes in the human microbiome is characterized by variable distribution patterns (Kraal, Abubucker, Kota, Fischbach, & Mitreva, 2014) with prominent abundance of some strains in some subjects while nearly absent in others. While the presence of microbes with low prevalence across all samples can be the focus of research for future experimental study (Kraal et al., 2014), the identification of microbial taxa present in the majority of the subjects, also known as the core microbiome, has been one of the primary goals of the Human Microbiome Project (Consortium, 2012; Huse, Ye, Zhou, & Fodor, 2012). The microbial core can be used as standard to identify significant variations that might be associated with disease states or other treatments.

Many tools such as Phyloseq (McMurdie & Holmes, 2013), Qiime (Caporaso et al., 2010), UPARSE (Edgar, 2013), MG-RAST (Meyer et al., 2008), Mothur (Schloss et al., 2009), and MicrobiomeAnalyst (Dhariwal et al., 2017) have been developed to contrast experimental factors in microbiome studies. The choice of a given analysis package is usually based on the user's questions of interest, level of experience in bioinformatics, and on the available resources at the user's host institution (Pollock,

Glendinning, Wisedchanwet, & Watson, 2018). Nevertheless, most approaches embedded in these packages rarely consider microbial prevalence within treatments.

Here we propose a new workflow based on the core microbiome concept that is designed to identify and remove the within group variation found in amplicon surveys (16S rRNA datasets) by capturing only biological differences at high sample prevalence levels. That means in an experiment comparing two treatments (e.g. healthy versus diseased subjects) one core for each treatment will be calculated and relevant microbial taxa responsible for differences within microbial cores will be detected. That is, we are asking the question of the extent to which core microbiomes differ. To implement this concept, we developed an R package called PIME (Prevalence Interval for Microbiome Evaluation). PIME is a tool specifically designed to work with datasets with high variation among samples. PIME removes low abundance taxa in each treatment or group keeping only those taxa that are shared at some level of prevalence. It calculates prevalence levels in 5% intervals from 5% to 95%. For each prevalence level a list of the most relevant taxa responsible for differences between or among treatments is provided. We also implemented an error detection step based on randomizations. It consists in calculating the likelihood of false predictions (i.e. existence of distinct groups when there are not) throughout the dataset filtration process.

2 MATERIALS AND METHODS

2.1 Program Description

2.1.1 Bioinformatics Workflow

The bioinformatics workflow described here is embedded in an R package called PIME (Prevalence Intervals for Microbiome Evaluation) available at: https://github.com/microEcology/PIME. PIME identifies statistically significant bacterial community differences considering the proportion of samples hosting a specific microbial community in a given time period. For the purpose of this work, prevalence was defined as the proportion of samples in a specific group (e.g. treatment or any

other factor the user want to compare) that share taxa, irrespective of the relative abundance, at the time of sampling. For example, a prevalence cutoff of 50% means that the taxa selected at this prevalence interval are found in 50% of the samples. PIME's strategy is based on four fundamental steps depicted in Figure 1 that we describe below.

I) Prediction of differences in full dataset:

PIME takes a phyloseq object (McMurdie & Holmes, 2013) as input. Phyloseq enables handling many data formats. PIME then builds hundreds of randomized decision trees, where each gives a vote for the prediction of the target variable, using a supervised non-parametric machine learning algorithm and combines them into a single model to predict the likelihood of detecting any user defined treatments or variables as source of sample variation (Breiman, 2001). The model performance is indicated by the out-of-bag (OOB) estimate of the error rate calculated by training the algorithm on a subset of samples and tested on the remaining samples. Values can vary between 0 and 1, where 0 and 1 indicate the model has 100% or 0% accuracy, respectively. This overall measurement of accuracy can be interpreted as an estimate of error obtained when the model is applied to new observations. Higher OOB error indicates low accuracy of the model in predicting differences among the categorical variables tested.

This first PIME step is implemented in a function called *pime.oob.error*. This function is run using the dataset without any filtering proposed by PIME. After obtaining the OOB error rate, the user decides whether PIME is adequate for the dataset. For example, an OOB error near zero indicates the prevalence filtering with PIME is not necessary, as the model accuracy is already reasonably good. On the other hand, if OOB error rate is greater than zero, filtering the dataset using PIME might improve the model accuracy. In this case only, the user can proceed and execute the function *pime.split.by.variable*. This step is defined below.

II) Split the dataset by predictor variable and compute prevalence intervals:

The full dataset is split according to the tested categorical variables (e.g. treatment and/or any other factor) defined by the user in the metadata file. Each variable will be used to define data subsets. Those per variable subsets are filtered using different prevalence levels from 5% to 95% with increments of 5% for each level

(see Figure 1 for a simplified schema illustrating this filtering step). Prevalence levels (usually high prevalence levels – e.g. 90%) where samples have zero sequences are not calculated. After removal of taxa that do not match the prevalence criteria, the subsets are merged to compose a new filtered dataset (one per prevalence interval) for subsequent downstream analysis. This step is implemented in these two functions: *pime.split.by.variable* and *pime.prevalence*. The *pime.split.by.variable* function uses the original dataset as input and its output is used as input for *pime.prevalence.* The function *pime.prevalence* keeps, for each treatment group, every OTU/ASV according to the following equation:

$$N_0/N_s > P_i == True$$

Where: N_0 is the number of OTUs/ASVs counts with $Sum > 0$, N_s is the number of samples and P_i is the prevalence interval $P_i = 0.05, \ldots, P_{max} = 0.95$.

III) Computation of OOB error on each prevalence interval and importance of each taxa in the differentiation of microbial communities

Next, Random Forest analyses (Breiman, 2001) are used to determine the level of prevalence that provides the best model to predict differences in the communities, while still including as many taxa as possible in the analysis. After prevalence filtering, performed according to the equation above, the OOB error rate and the number of remaining taxa and sequences are calculated for each prevalence level. The results are provided in a table that allows the user to determine the optimal prevalence interval with high accuracy. This step is implemented in a function called: *pime.best.prevalence.* Within the same function, the contribution of each taxa to the mean decrease in classification accuracy is calculated using the same Random Forests algorithm. High values of mean decrease accuracy indicate the importance of taxa to differentiate two or more microbial communities. The user can access the importance of taxa in each of the prevalence intervals.

IV) Validation

To obtain the likelihood of producing a type I error where PIME predicts the presence of distinct groups where no groups exist, an error detection step is included.

Consider the scenario in which the null hypothesis of “no difference between groups” is false. If we randomly shuffle the labels that identify the sample groups and run the test again, the expected outcome is that the randomized dataset will have a small chance to present distinct groups. Running the test multiple times with the random dataset is expected to produce a high OOB error rate in most cases. This error detection test is implemented in these two functions: *pime.error.prediction* and *pime.oob.replicate*. The first function randomizes the samples labels into arbitrary groupings using 100 random permutations. For each randomized prevalence filtered dataset, the OOB error rate is calculated to determine whether differences in the original groups occur by chance. The second function performs the Random Forest analyses and computes the OOB error for 100 replications in each prevalence interval without randomizing the sample labels. The biological difference among samples is expected to be greater than the differences generated randomly. Thus, the greatest fraction of randomizations should generate high error rates. On the other hand, no improvement in accuracy is expected within the randomized dataset.

Empirical Validation

The PIME workflow was compared against other existing filtering methods and by using empirical tests with 16S rRNA datasets. The performance of PIME was compared against filtering methods based on overall prevalence, low abundance, and low variance. Also, four 16S rRNA datasets were analyzed using PIME to illustrate its usefulness. These include an assessment of: a) the association between diet and saliva microbiome composition (unpublished original research); b) the gut microbiome in subjects at high genetic risk for type 1 diabetes (Davis-Richardson et al., 2014); c) the vaginal microbiome in pregnant women randomized to receive milk with or without probiotic bacterial strains (Avershina et al., 2017); and d) the saliva microbiome compared to the microbiome of the left antecubital fossa of healthy individuals (Human Microbiome Consortium, 2012).

Comparison with other existing filtering methods

Comparisons were performed using a dataset composed by 16S rRNA sequences from microbes extracted from saliva of 125 undergraduate and graduate students from the University of Florida (accessible through BioProject ID PRJNA504439). The following filtering tests were performed: a) filtering the dataset such that the taxa kept in the dataset must be present in at least 20% of the subjects; b) filtering the dataset by abundance to include only those taxa with at least 5 sequences; and c) filtering by low variance such that all taxa in the dataset have variance higher than 20%. Filtered datasets were compared against the prevalence interval of 65% as calculated by PIME as the best prevalence interval where the OOB error was zero. A record of this analysis containing a step-by-step R-code and results is provided in the Supplementary File S1.

Performance evaluation with 16S rRNA datasets

A novel and four published datasets were analyzed with PIME. These datasets covered a broad range of habitats including human and environmental samples. These are used to show that PIME does give predicted results in those cases where we expect to see no differences in the treatment [such as the studies of Avershina et al., (2017) and Davis-Richardson et al., (2014)] and in those cases where we expect large differences between the treatments [such as the saliva microbiome described in this paper and the comparison between saliva and the left antecubital fossa from the human microbiome project (Human Microbiome Consortium, 2012)]. The novel dataset used in this work comprised of 16S rRNA gene sequences from saliva samples obtained from 125 undergraduate and graduate students from the University of Florida. The study assessed the subject's diet as a factor influencing the saliva microbiome. This study was approved by the University of Florida's Institutional Review Board and assigned number IRB201602134. Approximately 224 undergraduate and graduate students taking three courses were invited to anonymously participate in this study as volunteers. A study coordinator was chosen to collect samples and code the samples so that those who did the analysis were unaware of the identity of the volunteers. To assess the diet, the subjects also completed the KIDMED survey (Serra-Majem et al.,

2004). The sampling collection, DNA extraction and library preparation are described below.

Sampling collection, DNA extraction and library preparation

Of the 224 students invited, 125 volunteers obtained the saliva sample collection and provided 2 ml of saliva. The samples were taken from each subject using the GeneFiX™ Saliva DNA Collection device. The collection kit allows immediate stabilization of the DNA. Total DNA was extracted using the GeneFix™ Saliva-prep-2 kit (Cell Projects Ltd, Harrietsham, UK) following the manufacturer's protocol. DNA samples were stored at -20 ºC until use.

To assess the diet, the subjects also completed the KIDMED survey (Lluís Serra-Majem et al., 2004). The KIDMED Index is based on a series of 16 questions, which measures the degree to which a subject adheres to the Mediterranean diet. The KIDMED index has been validated with nutritional data (Ll Serra-Majem, Ribas, García, Pérez-Rodrigo, & Aranceta, 2003) and was much simpler to implement than a diet diary or a serum-based nutrition analysis. Participant's age and gender were also obtained.

The 16S rRNA library preparation as well as the PCR reactions, primers and thermocycling conditions were performed as described previously (Davis-Richardson et al., 2014) and sequenced with Illumina MiSeq: 2x300 cycles run. The raw fastq files were used to build a table of exact amplicon sequence variants (ASVs) with DADA2 version 1.8 (Callahan et al., 2016). Taxonomy was assigned to each ASV using the SILVA ribosomal RNA gene database version v132 (Quast et al., 2012). A detailed R script containing the code used to generate the ASV table is provided in the Supplementary File S2. Downstream analyses were carried out after the normalization of the number of sequences in all samples as recommended by Lemos et al., (2011). The rarefied dataset comprised of 24,900 sequences per sample.

Description of the previously published datasets

The first previously published dataset used here was described by Davis-Richardson et al. (Davis-Richardson et al., 2014) and comprised of partial 16S rDNA sequences from fecal samples of 76 subjects born between 1996 and 2007 at the Turku University Hospital in southwestern Finland. All subjects were at high genetic risk for type 1 diabetes. The cohort was retroactively selected to create an age-

matched genotype-controlled set of subjects for the investigation of the microbiome as an environmental factor influencing the development of Type-1 diabetes. The raw Fastq files were obtained and sequences were processed using DADA2 version 1.8 (Callahan et al., 2016), as described above. Cases were defined as subjects who developed at least two persistent islet cell autoantibody (ICA), IAA, GADA, or IA-2A. Controls were defined as subjects with no detectable islet autoantibodies. Samples from subjects older than one year and post seroconversion were removed.

The second published dataset used here was previously described by Avershina et al. (Avershina et al., 2017). This dataset is comprised of amplified and sequenced 16S rRNA genes from vaginal swab samples collected from a cohort of 256 pregnant women. These subjects were randomized to receive a daily dose of fermented milk containing probiotic bacterial strains, or milk without probiotics. The corresponding author kindly provided an OTU table with 3,000 sequences per sample and the accompanying metadata. This table was used in all downstream bioinformatics and statistical analysis. Only those samples collected at the 36^{th} week of gestation were used in these analyses.

The third previously published dataset comprised of 16S rRNA gene sequences from the V1-V3 hypervariable region downloaded from the NIH Human Microbiome Project (https://www.hmpdacc.org/HMQCP/#data). The final OTU table processed by Qiime (Caporaso et al., 2010) using an OTU-clustering strategy and accompanying metadata were obtained and loaded into the R environment. After removing singletons, only saliva and left antecubital fossa samples were kept. The filtered dataset comprised of 113 saliva samples and 59 left antecubital fossa samples. All were rarefied to 2,000 sequences per sample.

The fourth published dataset comprised of 16S rRNA sequences from soils in a well-controlled microcosm system designed to investigate the individual and interactive effects of moisture and temperature (Lupatini et al., 2019). Specifically, we compare three moisture regimes at 10 ºC using only DNA samples rarefied at 7,100 sequences. A record of all statistical analyses comparing the datasets with and without using PIME including the R-code are included as Supplementary File S3.

3 RESULTS

3.1 Performance of PIME compared against other filtering methods

The performance of PIME was compared with other filtering methods (Figure 2 and 3). After quality filtering the saliva dataset, a total of 4,981,638 high-quality sequences, 400 bp long, were obtained from all subjects. An average 44,258 (sd = 27,743) sequences per sample were obtained. The dataset was rarefied to 24,900 sequences per sample in all analyses commensurate with the lowest number of sequences found in any one sample. Good's coverage (Good, 1953) ranged from 0.97 to 1.00 indicating this number of sequences was sufficient to accurately reflect the microbial diversity in these samples given the low complexity of saliva samples. The optimal prevalence interval calculated by PIME was 65%. This prevalence interval was used to compare the performance of PIME against the other filtering methods. The original dataset, without any filtering, presented 4,555 ASVs and a total of 3,112,500 sequences after rarefaction. Both filtering methods, *prevalence overall* and *PIME*, excluded the highest proportion of ASVs and sequences while filtering by abundance or variance excluded only 22% of ASVs and kept 99.9% of the sequences. Nevertheless, the overall prevalence kept 84% of the sequences while PIME kept 68% of the total number of sequences. Without using the PIME filtering the OOB error obtained while attempting to classify the salivary microbiome according to the three diet categories was 44%. This shows that the overall prevalence without using PIME model had low accuracy in predicting diet according to the microbiota. However, the PIME model had an OOB error of 0% (accuracy of 100%). The analysis of the taxonomic composition at phylum level after filtering the saliva dataset with PIME and other filtering methods are presented in Figure 3. PIME did not skew the phylum distribution but, as expected, removed low prevalent ASVs from particular phyla (e.g. Actinobacteria and Proteobacteria).

3.2 PIME application and effectiveness

Different datasets were used to validate the PIME workflow. PIME computed the OOB error rate from random forests, the number of taxa, and the number of remaining sequences for each prevalence interval from the diet-saliva dataset (Figure 4). Stringent criteria for definition of prevalence lead to greater improvement in accuracy for predicting diet based on the salivary microbiota. The prevalence interval

of 65% provided the best separation of microbial communities (OOB error = zero) while still including the majority of the sequences in the analysis. This prevalence interval was chosen for further analysis, but other intervals of prevalence can also be tested. For instance, the prevalence interval of 25% had OOB error of 7.2%. This indicates that the model is 92.8% accurate, which is a reasonably good model and keeps 88% of the sequences. Those ASVs that contributed to separating the core microbiomes among the diet categories (high, medium, or low diet categories) at 65% prevalence are provided by PIME (Table 1). The table indicates the ability of each variable to classify the microbes according to the three diet categories. The ASVs are ordered as most- to least-important. The more the accuracy of the random forest decreases due to the exclusion of a single ASV, the more important that ASV is, and therefore variables with a large mean decrease in accuracy are more important for classification of the ASVs according to diet. The mean decreased accuracy of the unfiltered dataset presented extremely low values. Negative values or close to zero indicate that the variable does not have a role in the prediction. In other words, the variable is not important to differentiate groups (Table 1). On the other hand, after PIME filtering, the mean decrease accuracy values increased indicating a true contribution of each ASV to classify diet differences among the core microbiomes. Altogether, the results indicated that after PIME filtering differences in the saliva microbiome was partially explained by diet rather than by random distribution patterns. The traditional approach, not accounting for microbial prevalence, was unable to distinguish these differences.

Following this first test, 16S rRNA data from stool of 76 children at high genetic risk for type 1 diabetes (Davis-Richardson et al., 2014) were tested for prevalence differences in those samples from children who remained healthy versus those that became autoimmune. PIME computed the OOB error rate from random forests, the number of taxa, and the number of remaining sequences for each prevalence interval from this dataset described (Figure 4). PIME also calculated prevalence intervals up to 70%. None of the sequences had a prevalence level higher than 70%. As expected, the OOB error rate decreased with higher prevalence intervals. At 60% prevalence the OOB error was zero and the number of remaining sequences was 1,165,304. The importance of each ASV in finding core microbiome differences between cases and controls subjects at 60% prevalence was also determined by PIME (Table 2). Accuracy was improved by applying PIME filtering to the dataset compared to the unfiltered

dataset. Previously, Davis-Richardson et al. (2014) discovered that the relative abundance of *Bacteroides* was significantly higher in autoimmune vs. control subjects. The presence of *Bacteroides* as an important taxa associated with autoimmune subjects was confirmed by PIME and other Amplicon Sequence Variants (ASVs) belonging to *Veillonella* genus were also found associated with autoimmune subjects.

In the third dataset tested, taxa were equally likely to be detected in the probiotic and placebo groups (Avershina et al., 2017). PIME prevalence filtering did not capture any difference between treatments (Figure 4). As the vaginal environment is dominated by *Lactobacillus*, a severe drop in the number of sequences at 5% prevalence interval was observed, The OOB error rate of the overall model obtained by Random Forest analyses suggests that irrespective of the prevalence interval no distinction between probiotic consumption and placebo exists (Supplementary File 3). Those results confirm the author's previous findings and demonstrate that our approach is not prone to type I errors (finding false positive results).

PIME tested the association between saliva microbiome and the left antecubital fossa, using a dataset from the Human Microbiome Project (Consortium, 2012). These two distinctive human microbial habitats were selected as they are expected to harbor very different communities. As predicted, PIME showed that the microbial habitats tested were very distinct. The OOB error rate was 0.005 within the original dataset and zero at all prevalence intervals applied (Figure 4 and Supplementary File 3) indicating the prevalence filtering does not increase the differentiation between these very different microbial habitats.

Finally, to show PIME can also be applied to any environmental survey and not only to human datasets, a database comprised of 16S rRNA sequences from soils was tested (Lupatini et al., 2019). The soil 16S rRNA sequencing was designed to investigate the effects of moisture and temperature under the microbial community (Figure 4 and Table 3). As expected, the OOB error rate decreased with higher prevalence intervals. After PIME filtering, the OOB error rate was about 88%. The best prevalence interval for this dataset was 70% where the OOB error rate was zero. Also, after PIME filtering, the mean decrease accuracy values (a measure of importance of a particular OTU/ASV to explain the model) increased indicating a true contribution of each ASV to detect moisture regime differences among the core microbiomes (Table 3). The mean decrease accuracy values were obtained from the calculations of OTU contribution to prevalence.

Once the best prevalence-filtered dataset is determined, the logical next step is to use the same algorithm (Random Forests) to find the most important OTUs/ASVs responsible for the differences related to a given condition. PIME performs this analysis as described above. However, PIME users might also take advantage of third-party software to further analyze the filtered dataset. To demonstrate this capability, the saliva's microbiome from High and Low KIDMED diet scores were compared using the DESeq2 algorithm (Love, Huber, & Anders, 2014). This example is provided in the Supplementary File 3.

3.4 Likelihood of introducing type I error while building prevalence-filtered datasets

PIME includes an error detection step and the results are presented here (Figure 5). The biological difference among samples is expected to be greater than the differences generated randomly. Thus, as the prevalence interval increases, the OOB error should decrease. As expected, the OOB error rate of samples with true biologically relevant differences (Figures 5A, 5B, 5D and 5E) decreased (or remained constant in low noise datasets – Figure 5C) with the increase in the prevalence interval. On the other hand, random sampling produced OOB error rate always higher than those obtained based on the original dataset. In datasets with no expected biologically relevant differences (Figure 5C), the OOB error did not decrease with higher prevalence intervals. In those cases, the randomized datasets produced higher OOB error rates. Thus, the signal to noise ratio increases with the prevalence intervals generating low OOB error rate values while no improvements in accuracy are observed within the randomized datasets. This error detection analysis showed that no bias was introduced while building prevalence-filtered datasets confirming this workflow is not prone to type I errors.

4 DISCUSSION

Prevalence is a key epidemiological concept where the number of people affected by a disease are counted with respect to the entire population (Noordzij, Dekker, Zoccali, & Jager, 2010; Ward, 2013). PIME was designed based on this concept. Here, the importance of a microbial community found in a single sample is less than if the same community is present in the majority of samples. Under that rationale, a workflow was designed to compare the prevalent populations between treatments. Prevalence has been used in the last as a filter of an entire dataset but never as a means to distinguish treatments. Prevalence differences between treatments are masked when only relative abundance is considered.

Challenges in microbiome data include the presence of many taxa represented sparsely in the dataset. This often results in large variation in distribution patters (also known as over-dispersion). Hence, microbes that are prominent in some subjects/samples and nearly absent in others (Kraal et al., 2014; Li, 2015). The current major challenge for using this information is how to convert it into rational biological conclusions providing control for error rates of false discoveries. Many tools are available to contrast experimental factors but they usually only take into account the microbial abundance and/or presence/absence. Thus, PIME overcomes those challenges by determining per treatment microbial prevalence in the analysis. This approach greatly improves the results by removing a substantial amount between-sample variation within groups that are represented by organisms that are rare in the population. PIME keeps only microbes found in many of the subjects of a population. The prevalence frequency can be chosen by the user after considering the OOB error rate. Thus, PIME can lead to a greater understanding of pathogenesis and the identification of potential probiotic treatments and prevention strategies that are masked by traditional analyses.

The definition of the best prevalence cutoff and the importance of taxa to discriminate treatments are performed by the Random Forests algorithm through the PIME workflow. This machine-learning algorithm has no formal distributional assumptions and can manage skewed and multi-modal data as well as categorical data. It can also manage situations in which the number of predictor variables (OTUs/ASVs) greatly exceeds the number of observations (Cutler et al., 2007). PIME is simple and accurate compared to other machine learning methods (Statnikov et al., 2013) and is applicable for classification of binary and multicategory experiments. Classification by these means is very accurate even with the default parameters,

(Statnikov et al., 2013; Zhou & Gallins, 2019). This ensures a more broad and practical use of PIME. Other methods that are not based on machine learning are able to identify taxa that are indicative of a given condition. This is often called differential abundance analysis. They usually compute p-values, adjusted p-values, False Discovery Rates and Effect Sizes and are based on microbial abundances. Random Forest analysis does not perform this traditional statistical inference. The importance of OTU/ASVs to differentiate treatments or ecological conditions may be used for the purpose of prediction. This machine-learning algorithm also provides an indication of the performance (OOB error rate) for comparisons of two or more microbial communities without OTU/ASV selection. This information is key into PIME's workflow as the estimate of error is used to define the best prevalence cutoff for filtering low prevalent OTUs/ASVs. Taking these considerations into account, PIME is not comparable to other methods designed to perform microbial differential abundance analysis. Still, any other tool can be applied after obtaining the prevalence-filtered dataset by PIME. For example, the filtered dataset can be analyzed using the DESeq2 algorithm (Love et al., 2014) to identify taxa that differ between the saliva microbiome from High and Low KIDMED index (Supplementary File 3).

Several tools designed to support microbiome statistical data analysis include data filtering as one of the first steps. The most commonly used filtering includes the exclusion of low count features (low abundance) using a minimum, yet arbitrary, relative abundance. Such features are very unlikely to be significant in the comparative analysis and likely have low overall prevalence. Arguably filtering those low abundance taxa can reduce the data sparsity issue, improving statistical power. However, when PIME performance is compared with other filtering methods, PIME outperformed all of those other approaches reducing the error rate and detecting microbial community differences where none were seen by other methods. This was illustrated by implementing PIME and other methods to a variety of 16S rRNA datasets. Within all of our tests, PIME confirmed previous findings and improved the results.

PIME does have some limitations. As PIME relies strongly on group prevalence, it is sensitive to the quality of sample groups. Poorly categorized groups comprised of subjects/samples with very different microbial composition may affect the prevalence calculations. Therefore, PIME might not be as effective in suggesting a good prevalence interval for filtering where groups are simply not different (Figure 5C). For

datasets with a very large number of samples, PIME might not find a clear prevalence interval for data filtering as prevalence of a given taxon will likely decline as the number of samples increases. With an increasing number of samples, the chance of sampling different “cores” or subpopulations is also increased. In addition, when there is large heterogeneity within sample groups, coupled with high data sparsity, prevalence computation might not be successful. Also, although this doesn't affect prediction erros, Random Forest models are sensitive to multicollinear variables when informing variable importance. Colinear variables might have inaccurate importance values. For example, if the first chosen variable provides little information, the model may be less accurate. Nevertheless, we have shown using a variety of datasets that PIME can be useful in many circumstances to unveil differences in community structure that are not detected by other methods and it not subject to type 1 errors.

5 Acknowledgements

L.F.W.R. and P.C.T.D. were supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), respectively. The work was supported by grants to E.W.T. from NIH (5R21AI120195-02) and JDRF (1-INO-2018-637-A-N). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (CAPES), Finance Code 001.

6 References

Avershina, E., Slangsvold, S., Simpson, M. R., Storrø, O., Johnsen, R., Øien, T., & Rudi, K. (2017). Diversity of vaginal microbiota increases by the time of labor onset. *Scientific Reports*, *7*(1). doi: 10.1038/s41598-017-17972-0

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. doi: 10.1038/nmeth.3869

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. doi: 10.1038/nmeth.f.303

Consortium, H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, *486*(7402), 207–214. doi: 10.1038/nature11234

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. *Ecology*, *88*(11), 2783–2792. doi: 10.1890/07-0539.1

Davis-Richardson, A. G., Ardissone, A. N., Dias, R., Simell, V., Leonard, M. T., Kemppainen, K. M., … Triplett, E. W. (2014). Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes. *Frontiers in Microbiology*, *5*. doi: 10.3389/fmicb.2014.00678

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Research*. doi: 10.1093/nar/gkx295

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), 996–998. doi: 10.1038/nmeth.2604

Edgar, R. C., & Flyvbjerg, H. (2015). Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics*, *31*(21), 3476–3482. doi: 10.1093/bioinformatics/btv401

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends in Microbiology*, *27*(2), 105–117. doi: 10.1016/j.tim.2018.11.003

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, *40*(3–4), 237–264.

Huse, S. M., Ye, Y., Zhou, Y., & Fodor, A. A. (2012). A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. *PLoS ONE*, *7*(6), e34242. doi: 10.1371/journal.pone.0034242

Kraal, L., Abubucker, S., Kota, K., Fischbach, M. A., & Mitreva, M. (2014). The Prevalence of Species and Strains in the Human Microbiome: A Resource for Experimental Efforts. *PLoS ONE*, *9*(5), e97279. doi: 10.1371/journal.pone.0097279

Lemos, L. N., Fulthorpe, R. R., Triplett, E. W., & Roesch, L. F. W. (2011). Rethinking microbial diversity analysis in the high throughput sequencing era. *Journal of Microbiological Methods*, *86*(1), 42–51. doi: 10.1016/j.mimet.2011.03.014

Li, H. (2015). Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annual Review of Statistics and Its Application*, *2*(1), 73–94. doi: 10.1146/annurev-statistics-010814-020351

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). doi: 10.1186/s13059-014-0550-8

Lupatini, M., Suleiman, A. K. A., Jacques, R. J. S., Lemos, L. N., Pylro, V. S., Van Veen, J. A., … Roesch, L. F. W. (2019). Moisture Is More Important than Temperature for Assembly of Both Potentially Active and Whole Prokaryotic Communities in Subtropical Grassland. *Microbial Ecology*, *77*(2), 460–470. doi: 10.1007/s00248-018-1310-1

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. doi: 10.1371/journal.pone.0061217

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., … Edwards, R. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*(1), 386. doi: 10.1186/1471-2105-9-386

Noordzij, M., Dekker, F. W., Zoccali, C., & Jager, K. J. (2010). Measures of Disease Frequency: Prevalence and Incidence. *Nephron Clinical Practice*, *115*(1), c17–c20. doi: 10.1159/000286345

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. *Applied and Environmental Microbiology*, *84*(7). doi: 10.1128/AEM.02627-17

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., … Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. doi: 10.1093/nar/gks1219

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., … Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. doi: 10.1186/s12915-014-0087-z

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing

Microbial Communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541. doi: 10.1128/AEM.01541-09

Serra-Majem, Ll, Ribas, L., García, A., Pérez-Rodrigo, C., & Aranceta, J. (2003). Nutrient adequacy and Mediterranean Diet in Spanish school children and adolescents. *European Journal of Clinical Nutrition*, *57*(S1), S35–S39. doi: 10.1038/sj.ejcn.1601812

Serra-Majem, Lluís, Ribas, L., Ngo, J., Ortega, R. M., García, A., Pérez-Rodrigo, C., & Aranceta, J. (2004). Food, youth and the Mediterranean diet in Spain. Development of KIDMED, Mediterranean Diet Quality Index in children and adolescents. *Public Health Nutrition*, *7*(07). doi: 10.1079/PHN2004556

Statnikov, A., Henaff, M., Narendra, V., Konganti, K., Li, Z., Yang, L., … Alekseyenko, A. V. (2013). A comprehensive evaluation of multicategory classification methods for microbiomic data. *Microbiome*, *1*(1), 11. doi: 10.1186/2049-2618-1-11

Sze, M. A., & Schloss, P. D. (2016). Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. *MBio*, *7*(4). doi: 10.1128/mBio.01018-16

Tedersoo, L., Nilsson, R. H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., … Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, *188*(1), 291–301. doi: 10.1111/j.1469-8137.2010.03373.x

Ward, M. M. (2013). Estimating Disease Prevalence and Incidence Using Administrative Data: Some Assembly Required. *The Journal of Rheumatology*, *40*(8), 1241–1243. doi: 10.3899/jrheum.130675

Zhou, Y.-H., & Gallins, P. (2019). A Review and Tutorial of Machine Learning Methods for Microbiome Host Trait Prediction. *Frontiers in Genetics*, *10*, 579. doi: 10.3389/fgene.2019.00579

Data Accessibility

The R package, installation instructions and a step-by-step example on how to use PIME are freely available at: https://github.com/microEcology/PIME. The original 16S rRNA gene sequences generated in this work have been deposited in NCBI's Short Raw Archive and are accessible through BioProject ID PRJNA504439.

Author Contributions

LFW Roesch: conceived the project, supervised the code writing, analyzed the data and wrote the paper; PT Dobbler: wrote the R code, analyzed the data and wrote the paper; VS Pylro: supervised the data analysis, assist with the experiment design and wrote the paper; B Kolaczkowski: supervised the code writing and wrote the paper; JC Drew: wrote the paper; EW Triplett: supervised the data analysis, assist with the experiment design and wrote the paper.

The authors declare no conflict of interest

FIGURES

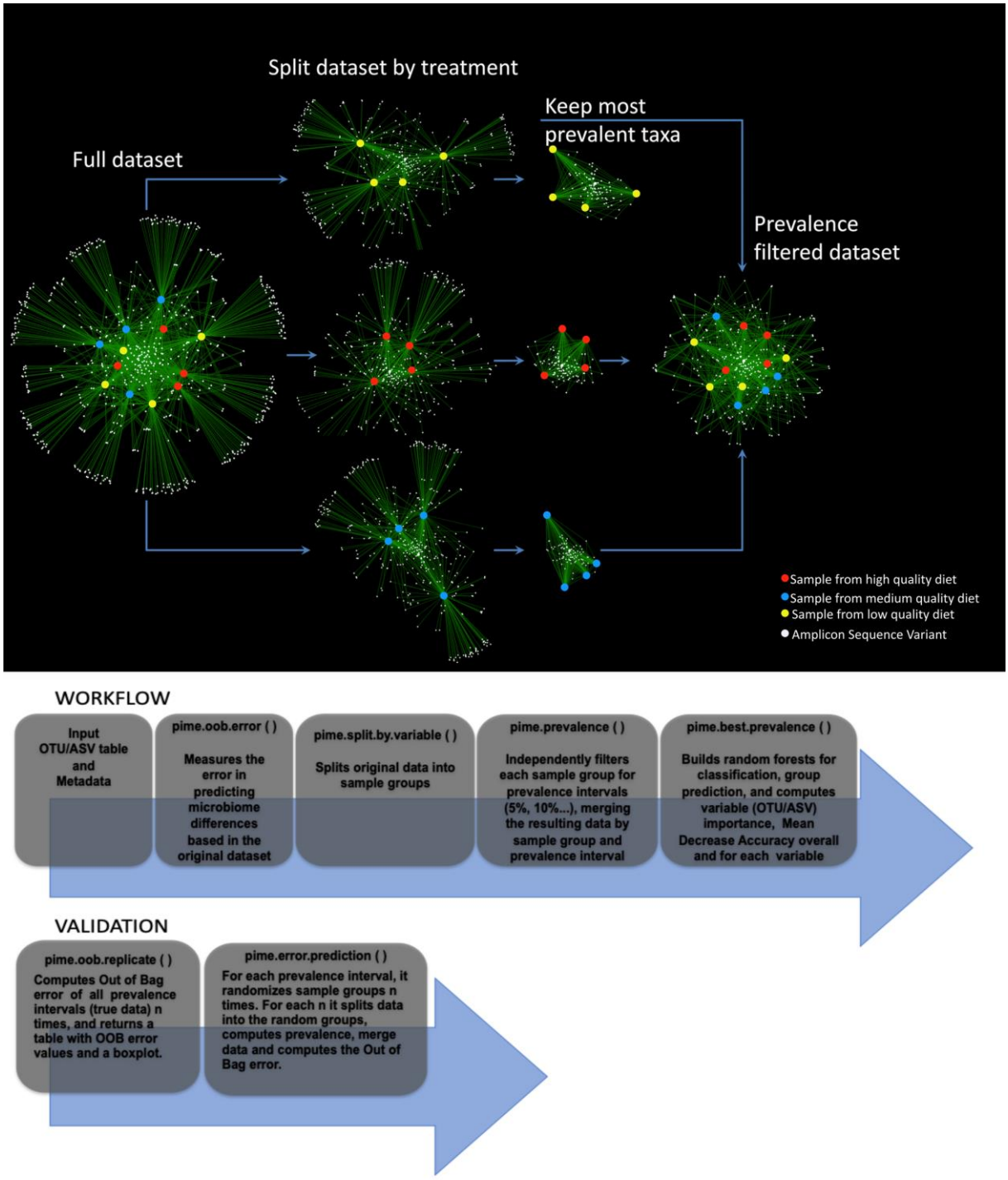


Figure 1. Empirical representation of steps used in PIME. Top panel. Bipartite network illustrating PIME method with a subset of 12 saliva's microbiome samples. Each sample (red, yellow and blue circles) is connected to an ASV (white circles) through edges (green). ASVs observed in more than one sample are connected by at least two edges and are displayed at the center of the network. ASVs present in only one sample are connected by a single edge and are displayed at the border of the network. The first step applied by PIME is to split the full dataset according to the treatments defined by the user.

Within this example red, yellow and blue circles depict three different treatments. At each of the three new groups the low prevalent ASVs are removed. Finally, the subsets are merged to compose a new filtered dataset used in the downstream analysis. Bottom panel. Step-by-step representation of PIME's workflow and validation.

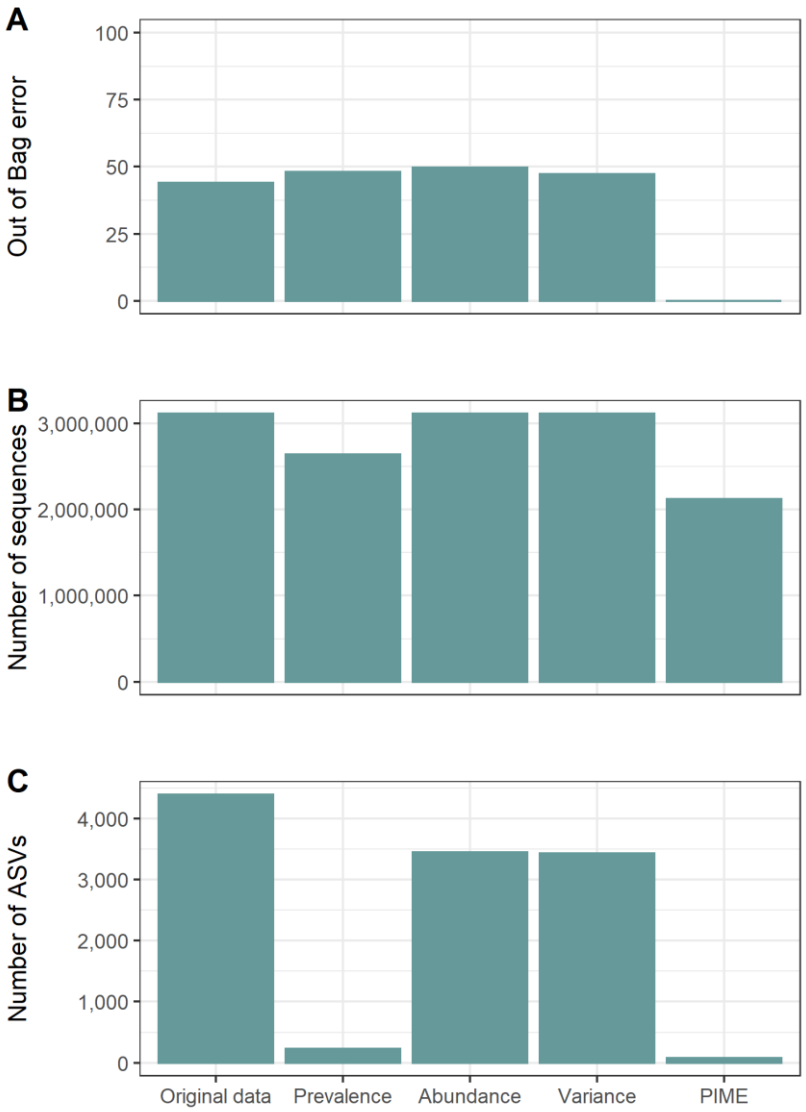


Figure 2. Performance of PIME compared to other filtering methods. A) Out of Bag error rate (OOB error rate); B) total number of sequences; C) Total number of ASVs. Prevalence = filter by overall taxa prevalence in at least 20% of the subjects; Abundance = filter by abundance of at least 5 sequences; Variance = filter by variance higher than 20%. PIME = filter by prevalence interval of 65%. Data was generated by using the saliva dataset and the step-by-step analysis can be found into the Supplementary File 1.

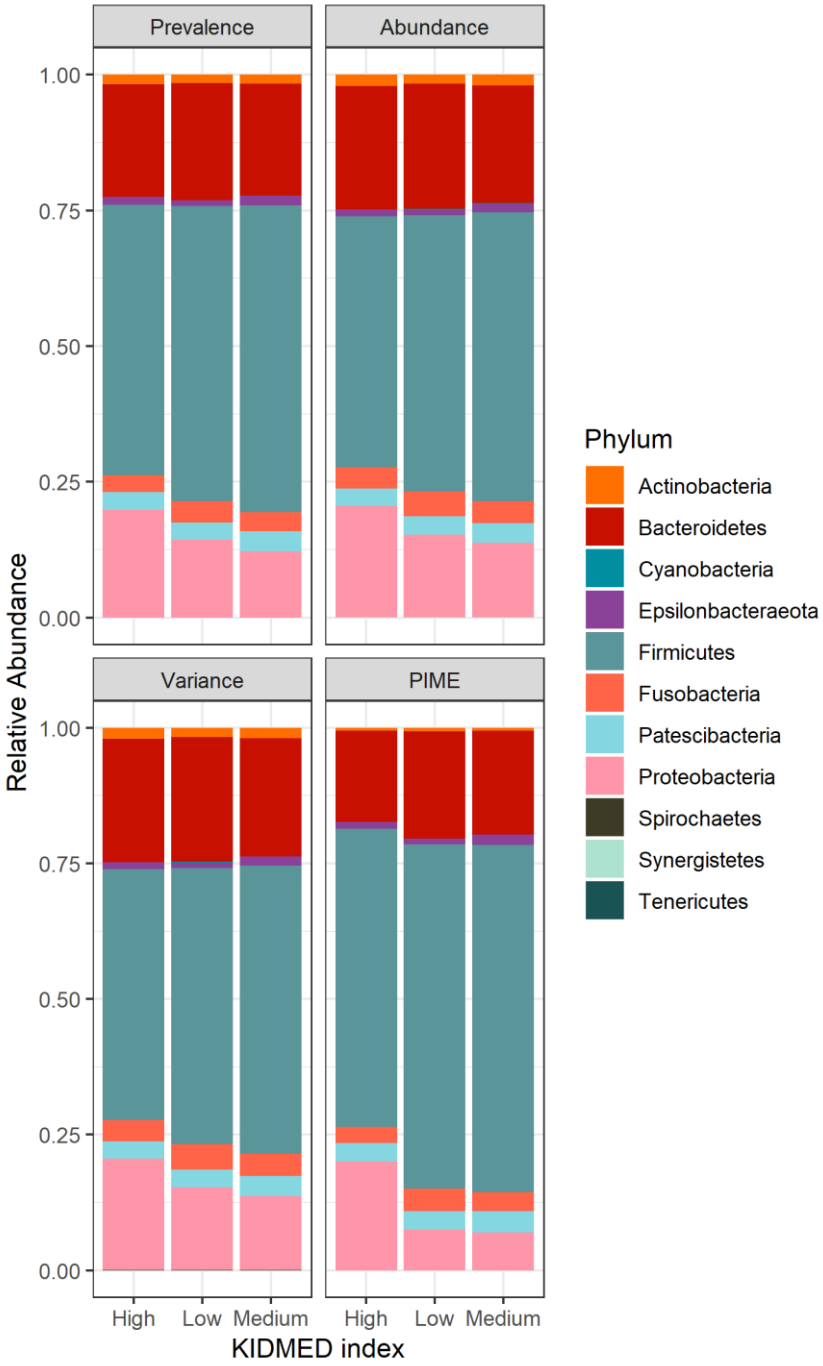


Figure 3. Changes in taxonomic composition at phylum level after filtering the saliva dataset with PIME and other commonly used filtering methods.

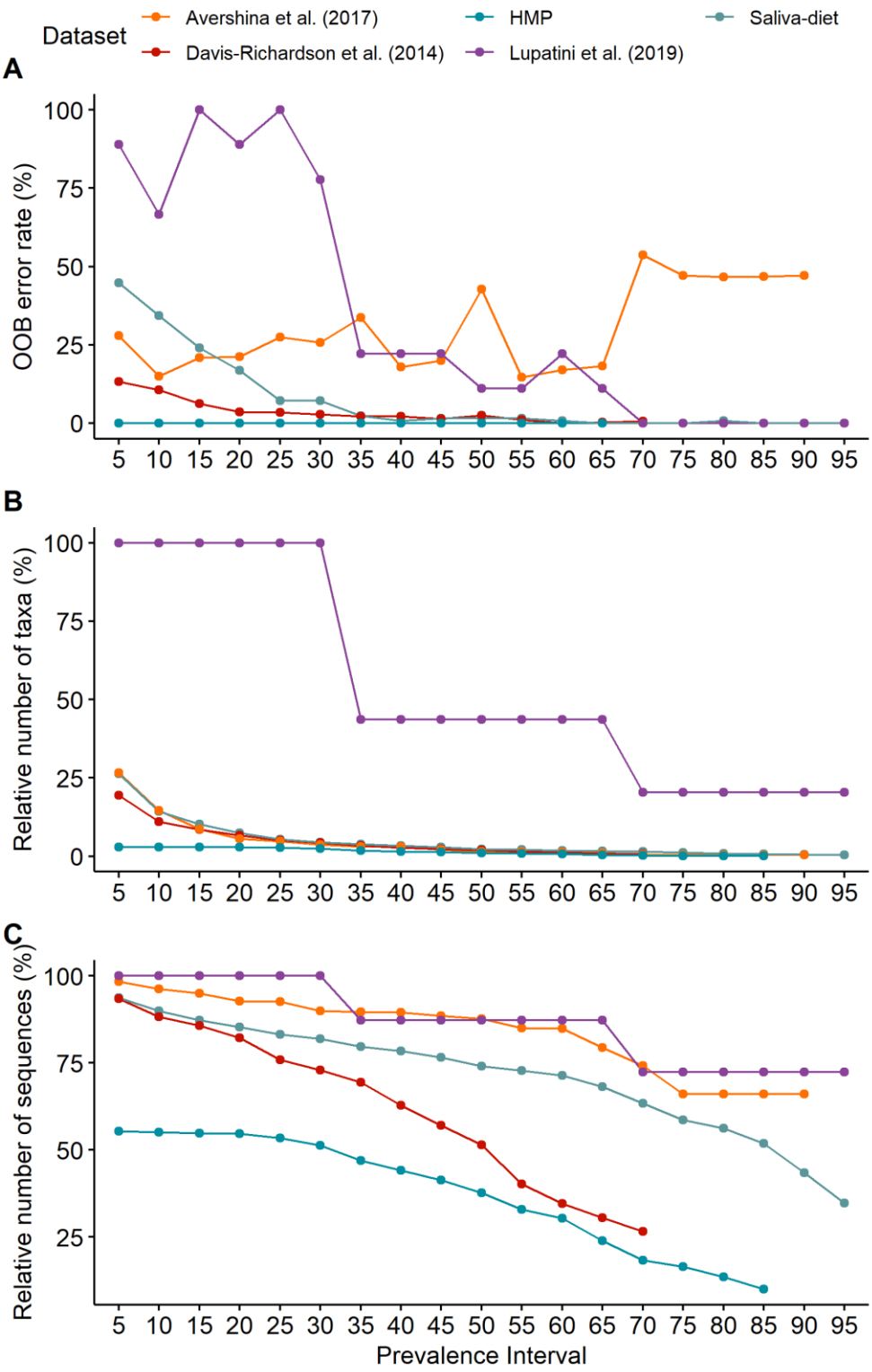


Figure 4. Computations of the out-of-bag error rate from random forests (A), percentage of remaining taxa (B) and percentage of remaining sequences (C) for each prevalence interval from five different 16S rDNA datasets.

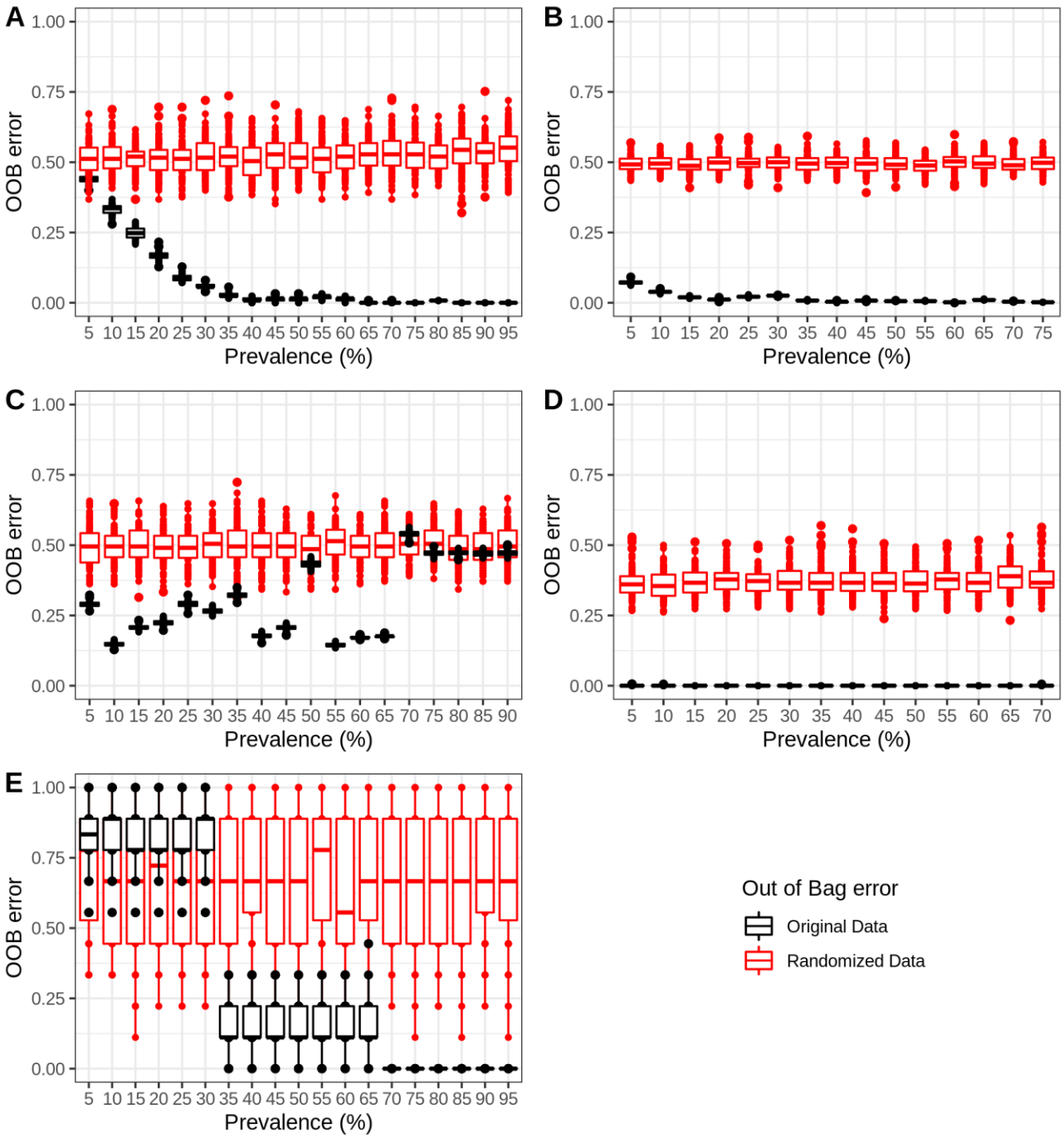


Figure 5. Boxplot depicting the PIME error detection step. Red boxes represent the OOB error rate obtained by randomly shuffling the labels into arbitrary groupings using 100 random permutations and running pime.error.prediction function at each randomization for each prevalence interval. Black boxes represent the OOB error rate against the 100 replications in each prevalence interval against the original sampling labels obtained by running pime.oob.replicate function. (A) Original dataset from salivary microbiome samples. (B) Data from the gut microbial of 76 children at high genetic risk for type 1 diabetes. (C) Data from the vaginal microbiome of pregnant women randomized to receive milk with or without probiotic bacterial strains. (D) Data from the microbiome of saliva and left antecubital fossa of healthy individuals. (E) Data from the soil microcosm system designed to investigate the individual and interactive effects of moisture and temperature. Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles, and the circles indicate outliers.

TABLES

Table 1. Importance of ASVs measured by mean decrease accuracy and the confusion matrix prior and after PIME to differentiate the three diet categories (High, Low and Medium) from the diet-saliva dataset.

Mean Decrease Accuracy				
High	**Low**	**Medium**	**Over all classes**	**Closest microbial relative at genus level**
Unfiltered Dataset				
0.0088	0.0002	0.0041	0.0037	*Neisseria*
0.0013	0.0046	0.0003	0.0017	*Parvimonas*
-0.0002	0.0019	0.0022	0.0016	*Veillonella*
0.0009	0.0025	0.0013	0.0015	*Prevotella*
0.0003	0.0041	0.0006	0.0015	*Parvimonas*
0.0004	0.0022	0.0014	0.0015	*Porphyromonas*
0.0021	0.0014	0.0009	0.0013	*Actinobacillus*
0.0014	0.0021	0.0007	0.0012	*Haemophilus*
0.0024	0.0005	0.0010	0.0011	*Alloprevotella*
0.0007	0.0004	0.0016	0.0011	*Alloprevotella*
Dataset filtered by PIME				
0.0200	0.0287	0.0883	0.0573	*Haemophilus*
0.0490	0.0060	0.0776	0.0504	*Haemophilus*
0.0167	0.0153	0.0567	0.0364	*Prevotella_7*
0.0317	0.0455	0.0318	0.0349	*Gemella*
0.0279	0.0345	0.0260	0.0287	*Prevotella*
0.0236	0.0312	0.0278	0.0276	*Haemophilus*
0.0087	0.0046	0.0483	0.0273	*Capnocytophaga*
0.0200	0.0653	0.0027	0.0239	*Selenomonas_3*
0.0145	0.0250	0.0250	0.0228	*Granulicatella*
0.0188	0.0206	0.0233	0.0216	*Rothia*
Confusion matrix prior PIME				
	High	Low	Medium	Classification error
High	0	1	23	1.0000
Low	0	4	33	0.8918
Medium	0	2	62	0.0312
Confusion Matrix after PIME				
	High	Low	Medium	Classification error
High	24	0	0	0.0000
Low	0	37	0	0.0000
Medium	0	0	64	0.0000

Showing only the first 10 hits. A complete table with the 30 most important ASVs is provided in the Supplementary File 3.

Table 2. Importance of the ASVs measured by mean decrease accuracy and the confusion matrix prior and after PIME from the dataset described by Davis-Richardson et al., (2014).

Mean Decrease Accuracy			
Controls	**Cases**	**Over all classes**	**Closest microbial relative at genus level**
Unfiltered Dataset			
0.0043	0.0043	0.0043	*Bacteroides*
0.0028	0.0061	0.0040	*Bacteroides*
0.0036	0.0039	0.0038	*Bacteroides*
0.0032	0.0045	0.0037	*Bacteroides*
0.0033	0.0041	0.0035	*Bacteroides*
0.0036	0.0033	0.0035	*Bacteroides*
0.0028	0.0047	0.0035	*Bacteroides*
0.0031	0.0035	0.0032	*Bacteroides*
0.0025	0.0041	0.0031	*Bacteroides*
0.0031	0.0033	0.0031	*Bacteroides*
Dataset filtered by PIME			
0.0588	0.0148	0.0422	*Bacteroides*
0.0470	0.0119	0.0339	*Veillonella*
0.0386	0.0118	0.0284	*Bacteroides*
0.0372	0.0073	0.0260	*Veillonella*
0.0375	0.0063	0.0257	*Bacteroides*
0.0366	0.0070	0.0255	*Bacteroides*
0.0356	0.0074	0.0251	*Bacteroides*
0.0372	0.0045	0.0251	*Veillonella*
0.0366	0.0058	0.0250	*Veillonella*
0.0346	0.0076	0.0245	*Bacteroides*
Confusion Matrix prior PIME			
	Controls	Cases	Classification error
Controls	210	14	0.0625
Cases	54	79	0.4060
Confusion Matrix after PIME			
	Controls	Cases	Classification error
Controls	224	0	0.0000
Cases	0	133	0.0000

Showing only the first 10 hits. A complete table with the 30 most important ASVs is provided in the Supplementary File 3.

Table 3. Importance of the OTUs measured by mean decrease accuracy and the confusion matrix prior and after PIME from the dataset described by Lupatini et al., (2019).

Mean Decrease Accuracy				
Moisture regimes (%)			Over all classes	Closest microbial relative at genus level
8	16	23		
Unfiltered Dataset				
0.002	0.0000	0.0040	0.0020	*Reyranella*
0.002	0.0000	0.0020	0.0020	*Streptomyces*
0.002	0.0020	0.0000	0.0020	Unclassified Genus of Class OPB35
0.002	0.0020	0.0000	0.0020	*Candidatus Nostocoida*
0.000	0.0040	0.0020	0.0020	Unclassified Genus of Order *Armatimonadales*
0.000	0.0020	0.0040	0.0020	Unclassified Genus of Class BD7-11
0.000	0.0020	0.0040	0.0018	*Acidothermus*
0.002	0.0040	0.0000	0.0018	Unclassified Genus of Order *Rickettsiales*
0.004	0.0020	0.0000	0.0017	*Bacillus*
0.002	0.0000	0.0020	0.0017	*Jatrophihabitans*
Dataset filtered by PIME				
0.004	0.006	0.002	0.0053	Unclassified Genus of Phylum *Armatimonadetes*
0.004	0.000	0.006	0.0037	Unclassified Genus of Family *Nitrosomonadaceae*
0.000	0.006	0.004	0.0035	Unclassified Genus of Family GR-WP33-30
0.002	0.001	0.004	0.0033	*Pirellula*
0.002	0.003	0.002	0.0033	*Chthonomonas*
0.003	0.000	0.004	0.0033	Pir4_lineage
0.004	0.004	0.000	0.0032	Unclassified Genus of Family TX1A-55
0.002	0.006	0.001	0.0030	*Gemmatimonas*
0.004	0.000	0.006	0.0030	*Sphingomonas*
0.000	0.003	0.004	0.0030	Unclassified Genus of Family *Anaerolineaceae*
Confusion matrix prior PIME				
	8%	16%	23%	Classification error
8%	2	0	1	1.0000
16%	1	2	0	1.0000
23%	2	0	1	0.6667
Confusion Matrix after PIME				
	8%	16%	23%	Classification error
8%	3	0	0	0.0000
16%	0	3	0	0.0000
23%	0	0	3	0.0000

Showing only the first 10 hits. A complete table with the 30 most important ASVs is provided in the Supplementary File 3.

4. CONCLUSÃO

Na primeira parte desta tese, buscou-se analisar a microbiota vaginal de gestantes brasileiras, ao final do terceiro trimestre de gestação, e como está associada a microbiota do recém-nascido no momento do parto.

Dessa forma, foi identificado e caracterizado três tipos de microbiotas vaginais nas gestantes brasileiras, no momento do parto. Dois tipos foram dominados por *Lactobacillus spp.* enquanto outra não apresentou dominância de nenhum micro-organismo. A ausência de outros tipos de microbiotas pode ser devido ao número amostral limitado. Também foi demonstrado associação entre a microbiota vaginal materna e a intestinal do recém-nascido.

Na segunda parte desta tese foi proposto e validado, uma ferramenta capaz de determinar os níveis de prevalências adequados para redução da variabilidade intragrupo. A utilização de níveis de prevalência para filtragem intragrupo, implementado em PIME, permitiu a redução da variabilidade e permitiu identificar associações não identificadas anteriormente.

Por fim, o presente trabalho contribuiu para o melhor entendimento da microbiota materna no momento do parto, bem como forneceu uma nova ferramenta para busca de novas associações para o estudo de microbiomas.

5. REFERÊNCIAS

AVERSHINA, E. et al. Diversity of vaginal microbiota increases by the time of labor onset. **Scientific Reports**, v. 7, n. 1, p. 17558, 14 dez. 2017.

BERG, G. et al. Microbiome definition re-visited: old concepts and new challenges. **Microbiome**, v. 8, n. 1, p. 103, 30 jun. 2020.

BOSKEY, E. R. et al. Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. **Infection and Immunity**, v. 67, n. 10, p. 5170–5175, out. 1999.

CAPORASO, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. **Nature Methods**, v. 7, n. 5, p. 335–336, Maio 2010.

CHO, I.; BLASER, M. J. The human microbiome: at the interface of health and disease. **Nature Reviews Genetics**, 13 mar. 2012.

DHARIWAL, A. et al. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. **Nucleic Acids Research**, v. 45, n. W1, p. W180–W188, 3 jul. 2017.

DOMINGUEZ-BELLO, M. G. et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. **Proceedings of the National Academy of Sciences**, v. 107, n. 26, p. 11971–11975, 29 jun. 2010.

DONOHOE, D. R. et al. The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon. Cell Metabolism, v. 13, n. 5, p. 517–526, 4 maio 2011.

DORNELLES, L. V. et al. Meconium microbiota predicts clinical early-onset neonatal sepsis in preterm neonates. **The Journal of Maternal-Fetal & Neonatal Medicine**, p. 1–9, 7 jun. 2020.

EDGAR, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. **Nature Methods**, v. 10, n. 10, p. 996–998, Outubro 2013.

EISENHOFER, R. et al. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. **Trends in Microbiology**, v. 27, n. 2, p. 105–117, 1 fev. 2019.

FENCHEL, T.; BLACKBURN, H.; KING, G. M. **Bacterial Biogeochemistry: The Ecophysiology of Mineral Cycling**. [s.l.] Academic Press, 2012.

FIERER, N. et al. From Animalcules to an Ecosystem: Application of Ecological Concepts to the Human Microbiome. **Annual Review of Ecology, Evolution, and Systematics**, v. 43, n. 1, p. 137–155, 2012.

GOSALBES, M. J. *et al.* Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants. **Clinical & Experimental Allergy**, v. 43, n. 2, p. 198–211, fev. 2013.

HANSEN, Richard *et al.* First-Pass Meconium Samples from Healthy Term Vaginally-Delivered Neonates: An Analysis of the Microbiota**. PLoS ONE**, v. 10, n. 7, p. e0133320, 28 jul. 2015.

HOOPER, L. V.; GORDON, J. I. Commensal Host-Bacterial Relationships in the Gut. **Science**, v. 292, n. 5519, p. 1115–1118, 11 maio 2001.

HUTTENHOWER, C. et al. Structure, function and diversity of the healthy human microbiome. **Nature**, v. 486, n. 7402, p. 207–214, jun. 2012.

JOHNSON, C. C.; OWNBY, D. R. The infant gut bacterial microbiota and risk of pediatric asthma and allergic diseases. **Translational Research**, Microbiome and Human Disease Pathogenesis. v. 179, p. 60–70, 1 jan. 2017.

KHO, Z. Y.; LAL, S. K. The Human Gut Microbiome – A Potential Controller of Wellness and Disease. Frontiers in Microbiology, v. 9, 2018.

KONOPKA, A. What is microbial community ecology? **The ISME Journal**, v. 3, n. 11, p. 1223–1230, nov. 2009.

KRAAL, L. et al. The Prevalence of Species and Strains in the Human Microbiome: A Resource for Experimental Efforts. **PLOS ONE**, v. 9, n. 5, p. e97279, 14 maio 2014.

LEDERBERG, J.; MCCRAY, A. T. `Ome Sweet `Omics--A Genealogical Treasury of Words. **The Scientist**, v. 15, n. 7, p. 8–8, 2 abr. 2001.

MADAN, J. C. et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. **Archives of Disease in Childhood-Fetal and Neonatal Edition**, v. 97, n. 6, p. F456–F462, 2012.

MCMURDIE, P. J.; HOLMES, S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. **PLOS ONE**, v. 8, n. 4, p. e61217, 22 abr. 2013.

MEYER, F. et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, v. 9, n. 1, p. 386, 19 set. 2008.

MILANI, C. et al. The First Microbial Colonizers of the Human Gut: Composition, Activities, and Health Implications of the Infant Gut Microbiota. **Microbiology and Molecular Biology Reviews**, v. 81, n. 4, p. e00036-17, 1 dez. 2017.

MOLES, Laura *et al.* Bacterial Diversity in Meconium of Preterm Neonates and Evolution of Their Fecal Microbiota during the First Month of Life. **PLoS ONE**, v. 8, n. 6, p. e66986, 28 jun. 2013.

NURIEL-OHAYON, M.; NEUMAN, H.; KOREN, O. Microbial Changes during Pregnancy, Birth, and Infancy. **Frontiers in Microbiology**, v. 7, 2016.
POLLOCK, J. et al. The Madness of Microbiome: Attempting To Find Consensus "Best Practice" for 16S Microbiome Studies. **Applied and Environmental Microbiology**, v. 84, n. 7, 1 abr. 2018.

ROMERO, R. et al. The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. **Microbiome**, v. 2, n. 1, p. 18, 2014a.

ROMERO, R. et al. The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. **Microbiome**, v. 2, n. 1, p. 4, 3 fev. 2014b.

SCHLOSS, P. D. et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7537–7541, 1 dez. 2009.

SENDER, R.; FUCHS, S.; MILO, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. **PLOS Biology**, v. 14, n. 8, p. e1002533, 19 ago. 2016.

SZE, M. A.; SCHLOSS, P. D. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. **mBio**, v. 7, n. 4, 7 set. 2016.

TEDERSOO, L. et al. 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. **New Phytologist**, v. 188, n. 1, p. 291–301, 2010.

TORTORA, G. J.; FUNKE, B. R.; CASE, C. L. **Microbiologia (10a. ed.).** [s.l.] Grupo A - Artmed, 2000.

UEMATSU, S. et al. Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. Nature Immunology, v. 9, n. 7, p. 769–776, jul. 2008.

WEINSTOCK, G. M. Genomic approaches to studying the human microbiota. **Nature**, v. 489, n. 7415, p. 250–256, 12 set. 2012.

WHIPPS, J. M.; LEWIS, K.; COOKE, R. C. Mycoparasitism and plant disease control. In: **Fungi in Biological Control Systems**. [s.l.] Manchester University Press, 1988. p. 269.

WORTHAM, J. M. et al. Chorioamnionitis and Culture-Confirmed, Early-Onset Neonatal Infections. **Pediatrics**, v. 137, n. 1, jan. 2016.

YIU, J. H. C.; DORWEILER, B.; WOO, C. W. Interaction between gut microbiota and toll-like receptor: from immunity to metabolism. Journal of Molecular Medicine, v. 95, n. 1, p. 13–20, 1 jan. 2017.

APÊNDICE A – Doenças hepáticas de armazenamento de glicogênio são associadas a disbiose microbiana

Hepatic glycogen storage diseases are associated to microbial dysbiosis

Karina Colonetti[1,2], Bruna Bento dos Santos[1,2], Tatiéle Nalin[3,4], Carolina Fischinger Moura de Souza[3], Eric W. Triplett[5], Priscila Thiago Dobbler[6], Ida Vanessa Doederlein Schwartz[1,2,3], Luiz Fernando Wurdig Roesch[6]*

[1] Post-Graduation Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

[2] Laboratory of Basic Research and Advanced Investigations in Neurosciences (BRAIN), Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

[3] Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.

[4] Postgraduate Program in Medicine: Medical Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.

[5] Department of Microbiology and Cell Science, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, United States of America.

[6] Interdisciplinary Research Center on Biotechnology-CIP-Biotec, Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil.

*luiz.roesch@gmail.com

DOI: https://doi.org/10.1371/journal. pone.0214582

Received: August 23, 2018

Accepted: March 17, 2019

Published: April 2, 2019

Abstract

Introduction

The gut microbiome has been related to several features present in Glycogen Storage Diseases (GSD) patients including obesity, inflammatory bowel disease (IBD) and liver disease.

Objectives

The primary objective of this study was to investigate associations between GSD and the gut microbiota.

Methods

Twenty-four GSD patients on treatment with uncooked cornstarch (UCCS), and 16 healthy controls had their faecal microbiota evaluated through 16S rRNA gene sequencing. Patients and controls were 3 years of age and not on antibiotics. Faecal pH, calprotectin, mean daily nutrient intake and current medications were recorded and correlated with gut microbiome.

Results

Patients' group presented higher intake of UCCS, higher prevalence of IBD (n = 04/24) and obesity/overweight (n = 18/24) compared to controls (n = 0 and 06/16, respectively). Both groups differed regarding diet (in patients, the calories' source was mainly the UCSS, and the intake of fat, calcium, sodium, and vitamins was lower than in controls), use of angiotensin converting enzyme inhibitors (patients = 11, controls = 0; p-value = 0.001) multivitamins (patients = 22, controls = 01; p-value = 0.001), and mean faecal pH (patients = 6.23; controls = 7.41; p = 0.001). The GSD microbiome was characterized by low diversity and distinct microbial structure. The operational taxonomic unit (OTU) abundance was significantly influenced by faecal pH (r = 0.77; p = 6.8e-09), total carbohydrate (r = -0.6; p = 4.8e-05) and sugar (r = 0.057; p = 0.00013) intakes.

Conclusions

GSD patients presented intestinal dysbiosis, showing low faecal microbial diversity in comparison with healthy controls. Those findings might be due to the disease perse, and/or to the different diets, use of UCSS and of medicines, and obesity rate found in patients. Although the main driver of these differences is unknown, this study might help to understand how the nutritional management affects GSD patients.

Introduction

Hepatic Glycogen Storage Diseases (GSD) are genetic disorders caused by deficient activity of one of the enzymes involved in the glycogenolysis pathway. The global incidence is estimated at 1 case per 20,000–43,000 live births. The most common types of GSD are GSD I, GSD III and GSD Ixα [1].

In GSD I, glucose-6-phosphate cannot be dephosphorylated to free glucose. There are two major subtypes of GSDI: Ia (~80%), caused by mutations in the *G6PC*

gene, and GSD Ib (~20%), caused by mutations in the *SLC37A4* gene. The proteins produced from *G6PC* (catalytic activity) and *SLC37A4* (transporter) work together [2]. GSD Ia involves glycogenolysis and gluconeogenesis, and the clinical manifestations are increased weight, hepatomegaly, failure to thrive, fasting hypoglycaemia, high lactate, hyperuricemia, nephromegaly and hyperlipidaemia [3]. In addition to the features presented in GSD Ia, GSD Ib also presents with susceptibility to recurrent infections, impaired neutrophil and monocyte function, and inflammatory bowel disease (Crohn's-like IBD) [1].

Mutations in the *AGL* gene cause GSD type III, in which the defective glycogen debranching enzyme blocks glycogenolysis, stopping the conversion of glycogen to glucose-1-phosphate [4]. At the same time, gluconeogenesis is enhanced to help maintain endogenous glucose production. Hepatomegaly in type III GSD generally improves with age, but affected individuals may develop chronic liver disease (cirrhosis) and liver failure later in life [5].

GSD IX is caused by the inability of phosphorylase b kinase (PHKA) to break down the glycogen in liver and/or muscle cells. Type IXα glycogenosis is an X-linked disease caused by mutations in the alpha subunit of *PHKA*. The signs and symptoms typically begin in early childhood, but GSD IX is usually milder than the other types [6].

The treatment for the aforementioned types of GSD involves nutritional adjustments primarily, with the periodic and frequent administration of large amounts of uncooked cornstarch (UCCS) and restriction of simple carbohydrates [7] to maintain normoglycaemia and avoid glycogen storage. Usually, higher and frequent doses of UCCS are prescribed for type Ia patients and lower doses for type IX patients. The dose is adjusted according to weight and metabolic demand [8]. GSD III and IX patients may require a hyperproteic diet with fewer restrictions for simple sugars. Sometimes additional medications may be necessary.

During the last decades, our understanding of the human being has changed. We know now that the eukaryote cells encoded by our genome are not the only component of our body. Symbiont prokaryotic cells inhabiting many cavities of our body provide metabolic functions far beyond the scope of our own physiological capabilities [9]. These cells play an important role in health and disease states [10]. The gut microbes are the most studied human associated microbial communities and consists of trillions of microbes and millions of functional genes [11]. Healthy humans present a remarkable microbial diversity but with similar functions indicating that different microbial communities are associated with a healthy microbiome [12]. The gut microbiome can be influenced by diet, lifestyle, drugs and genetics of the host [13], and has been related to several features present in GSD patients including obesity, IBD and liver disease [14]. This work aimed to investigate possible associations between GSD and the gut microbiota.

Methods

This study was a cross-sectional, observational convenience sampling study, which included 24 GSD patients (Ia = 15, Ib = 5, III = 1, IXα = 3) and 16 healthy controls. All patients were recruited from the outpatient clinics of the Medical Genetics Service at Hospital de Clínicas de Porto Alegre (MGS-HCPA), Brazil from Jan/2016 to May/2017. As inclusion criteria, the subjects (patients and controls) were 3 years old

and not on antibiotics. The GSD patients also were required to: a) have a genetic diagnosis of GSD and b) be on treatment with UCCS. The healthy controls were recruited by invitation as they came to routine appointments at Santa Cecília Basic Health Unit, Porto Alegre, Brazil. All subjects received a kit and printed instructions for stool collection, storage, and transport. They were also provided with printed instructions to record three days of dietary information. Each participant collected their own frozen fecal sample and three-day dietary record and submitted them to an outpatient clinic during their next routine check-up. Upon returning to the clinic, each participant answered a brief questionnaire about personal features including weight and height, eating habits, intestinal habits, medicines of recent and/or continuous usage and lifestyle. The study protocol was approved by the Ethics Committee of Hospital de Clínicas de Porto Alegre (HCPA). All participants and/or legal guardians signed an informed consent.

As a routine, GSD patients seen at the MGS-HCPA who are on UCCS therapy also receive a multivitamin prescription. Despite optimum dietary treatment other drugs could also be prescribed, mainly for type I patients, such as allopurinol, to prevent gout and urate nephropathy; angiotensin converting enzyme inhibitors, to slow-down or prevent further deterioration of renal function; citrate, to preventing or ameliorating urolithiasis and nephrocalcinosis, in addition to correcting lactacidaemia; statins to treat hypercholesterolaemia [15]; and mainly for Ib patients, G-CSF to treat neutropenia, neutrophil dysfunction and IBD; and the intestinal anti-inflammatory mesalazine (5-amino-salicylic acid), also to treat IBD [16].

Nutritional assessment, clinical data and statistical analysis

Macro and micronutrients intake by the subjects were estimated from the three-day food records through the Nutribase software (NB16Cloud, CyberSoft, Inc., Phoenix, AZ, USA). The daily nutrient intake of each participant was the sum of the nutrients of each food item. The average of the three-day intake was used for further analysis. Multivitamin consumption and other medications were not included in the nutritional assessment but were considered as variables that potentially were modifying the gut microbial composition, so they were tested by Permutational Multivariate Analysis of Variance. Clinical data, such as IBD and other relevant conditions, were accessed from medical records. BMI-for-age and Z-scores were calculated within the World Health Organization (WHO) AnthroPlus software suite. A qualitative classification for this data followed the WHO criteria [17].

Statistical analysis among the groups was performed using PASW Statistics for Windows software (Vs18.0, 2009, SPSS Inc., Chicago, USA). Numerical variables were compared using the Mann-Whitney U test. Categorical variables were compared using X2, Fisher's exact test or Continuity Correction, when necessary (with statistical significant determined by the threshold $p \leq 0.05$). Statistical analyses with the microbiome feature are described below.

Bacterial DNA extraction, 16SrRNA gene amplifications and sequencing

The bacterial DNA was isolated from 0.3 mg of frozen faecal sample with QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA, USA) (Qiagen) according to

manufacturer instructions and stored at -20˚C until use. The NanoVue system (GE Healthcare, Chicago, IL, USAGE Healthcare) was used to assess the quality of extractions for downstream applications. For the sequencing step, the library was prepared following the procedures described by Barboza et al. [18]. Briefly, region V4 of 16S rRNA gene was amplified with the barcoded bacterial/archaeal primers 515F and 806R [19] in a reaction containing 2U of Platinum Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4 µL 10X High Fidelity PCR Buffer, 2 mM MgSO4, 0.2 mM dNTPs, 0.1 µM of both the 806R barcoded primer and the 515F primer, 25µg of Ultrapure BSA (Invitrogen, Carlsbad, CA, USA) and approximately 50 ng of DNA template in a final volume of 25 µL. After an initial denaturation step of 5 min at 95˚C, 30 cycles of 94˚C for 45 s, 56˚C for 45 s and 72˚C for 1 minute were performed, followed by a final extension step of 10 min at 72˚C. After visualization on agarose gel 1.5%, the PCR products were purified with the Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, USA) and the final concentration of the PCR product was quantified with the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's recommendations. Finally, the reactions were combined in equimolar concentrations to create a mixture composed of 16S gene amplified fragments of each sample. This composite sample was used for library preparation with the Ion OneTouch 2 System using the Ion PGM Template OT2 400 Kit (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed with Ion PGM Sequencing 400 on the Ion PGM System using Ion 318 Chip v2.

16S profiling data analysis

The Fastq files exported from the Ion PGM System were analysed with the BMP Operating System (BMPOS) [20] according to the recommendations of the Brazilian Microbiome Project [21]. Briefly, an Operational Taxonomic Unit (OTU) table was built using reads truncated at 200 bp and quality filtered with a maximum expected error of 0.5. After removing singletons, the sequences were clustered into OTUs at cutoff of 97% similarity, and chimeras were checked and removed to obtain representative sequences for each microbial phylotype. Taxonomic classification was carried out in QIIME version 1.9.1 [22] based on the UCLUST method against the SILVA ribosomal RNA gene database version v132 [23] with a confidence threshold of 80%. Downstream analyses were carried out with dataset rarefied to the minimum library size [24,25] in the R environment [26] using the phyloseq package [27] and vegan package [28]. The online software Microbiome Analyst [29] was used to further detect microbial biomarkers associated with GSD patients. After Cumulative Sum Scaling (CSS) normalization [30], the dataset was analysed by the non-parametric factorial Kruskal-Wallis (KW) sum-rank test followed by Linear Discriminant Analysis [31]. To make sure the biomarkers observed were not only driven by IBDlike patients, we performed one analysis using the full dataset and another analysis excluding all four IBD-like patients and matched controls.

Faecal calprotectin assay and pH measurement

Frozen faecal samples of patients and controls were thawed and aliquoted at room temperature (20˚C) to perform the pH measures and calprotectin assay. To determine the faecal pH, the samples were diluted 1:10 (w/v) in distilled water. After homogenization and incubation for 5 min at room temperature, the faecal pH was

measured by an electronic pH-meter (K39-1014B, KASVI, PR, Brazil) three minutes after complete electrode immersion.

The faecal calprotectin was quantified from 100 mg of faecal sample with the RIDASCREEN Calprotectin test (R-Biopharm AG) according to the manufacturer's instructions. Calprotectin is a calcium-/zinc-binding protein, highly stable and resistant to degradation by intestinal contents (pancreatic secretions, proteases, and bacterial degradation). It is mainly produced by neutrophils in inflammation and has been amply confirmed in intestinal inflammatory diseases [32]. Calprotectin was evaluated to verify gut inflammation across groups and its influence over the number of OTUs. Due to the small sample size of GSD III and IXα, just the subtypes Ia and Ib (groups containing >15% of total sample) were compared. Results for GSD Ia and GSD Ib patients were presented as median (Q1-Q3) and as min-max to GSD III and IXα. To test the correlation among calprotectin and OUT richness, patients who were on mesalazine were excluded from analysis.

Results

The characteristics of the patients and controls are summarized in Table 1. The nutrient intake varied significantly between groups (S1 Table); the largest variation observed was the higher total carbohydrate and calorie intakes in the GSD group due to UCCS usage. The amount of protein consumed (g) and the number of calories derived from proteins did not differ between patients and controls. However, the percentage of total caloric intake from proteins was lower in patients. Patients ingested less fat (g and Kcal/day) and had a lower percentage of fat in the diet. Regarding micronutrients, patients' diet was poor in calcium and sodium, and in vitamins B3, H, D and E in comparison to the control group's diet.

The intakes of macro and micronutrients were similar among all the GSD types, with some kcal variation from carbohydrate intake due the difference in UCCS consumption among groups (S2 Table).

Overall 16S rRNA sequencing results, sequence quality control and control for confounding variables

After quality filtering of the 16S rRNA reads, a total of 1,786,582 high-quality sequences longer than 200 bp were obtained. To analyse whether the number of sequences from each sample was representative of the underlying bacterial community, sequence coverage was calculated (S3 Table). An average of 44,664 sequences per sample was obtained with average sequence coverage of 0.99 at the 3% dissimilarity level. This sequencing depth was sufficient to obtain excellent representation of the microbial community in these samples.

Results for suspected confounding variables that potentially were modifying the gut microbial composition are presented at Table 1 and S1 Table. The gut microbial communities were not affected by sex, age, nor the nutritional status of the subjects tested. Faecal pH was lower in patients (6.23) than in controls (7.41), and this variable affected the presence/absence and abundance of the gut microbes, with a reduced OTU count in lower pH. Only 18% of controls (n = 3) and 41% of patients (n = 10) used

antibiotics within the 6 months prior to data collection. The use of antibiotics within the 6 months prior to sampling did not affect the presence/ absence of microbes ($p = 0.252$) nor microbial relative abundance ($p = 0.179$) in these samples.

Hepatic GSD is associated with an abnormal gut microbial community

The analysis of overall microbial community structure revealed significant differences between patients and controls (Fig 1). According to the PERMANOVA, the microbial community structure between patients and controls differed by the presence and absence of taxa ($r2 = 0.182$; $p = 0.003$) and by their relative abundances ($r2 = 0.166$; $p = 0.001$). The analysis indicated that the relative abundance of taxa contributed 16% of the variation in the microbial community between patients and controls while the presence/absence of specific taxa contributed 18% to that variation.

Microbial diversity as measured by richness of OTUs and by the Shannon diversity index also differed significantly ($p < 0.01$) between patients and controls (Fig 2). On average, control stool samples possessed 184 OTUs while the patients had only 100 OTUs. The average Shannon diversity index was 3.49 and 2.48 in controls and patients, respectively. Together, these beta and alpha diversity analyses indicated that the GSD gut microbiome is characterized by low diversity and distinct microbial structures.

Defining the main taxa associated with the gut microbiota of patients and controls

Specific microbial phylotypes present within the gut community might drive the main differences observed in GSD patients. To find those microbes, biomarker screening analysis was performed at different taxonomic levels. A total of 14 phyla were detected within these samples. However, more than half of the community was dominated by only three phyla: *Bateroidetes* (58% in controls; 47% in patients), *Firmicutes* (34% in controls; 39% in patients) and *Proteobacteria* (5.8% in controls; 10% in patients) (Fig 3). All of the other phyla had very low relative abundances. LEfSe analysis identified three microbial phyla as biomarkers with *Actinobacteria* and *Proteobacteria* overrepresented in patients while *Euryarchaeota* was underrepresented. In particular, *Proteobacteria* presented a very high LDA score (more than 3.9 orders of magnitude), reflecting a marked increase in relative abundance in patients and consistently low abundance in controls. *Firmitutes* had a marginally-significant difference between patients and controls ($p = 0.043$ and LDA score = 4.53 but FDR = 0.07).

At the genus level, nineteen microbial biomarkers were different, both in terms of statistics and biological consistency, between patients and controls (Table 2). Those genera were higher in controls. In patients, those genera were in low abundance and in some cases totally absent. The lack of those microbes might be reflected in the alpha and beta diversity results as mentioned previously (Figs 1 and 2). Besides, *Lactobacillus* and *Escherichia*/*Shigella* were found to be dominant in patients with a very high LDA score (4.36 and 3.89, respectively), highlighting the biological importance of those microbes in GSD. To remove any biases caused by patients with IBD-like symptoms (n = 4), all IBD-like patients and their respective controls were removed from the dataset and a new biomarker analysis was performed (Table 2).

Similar trends as observed within the full dataset were still present in this reduced dataset. However, the *Lactobacillus* genus, found previously in higher abundance in patients was not observed within the dataset without IBD-like patients. On the other hand, Escherichia/Shigella was still found to be more abundant in patients than in controls (LDA score = 3.85).

Correlations between the gut microbiota, diet, faecal pH and gut inflammation

Spearman correlations were calculated between the microbiome, diet, faecal pH and calprotectin (Fig 4).

The faecal pH values varied between patients and controls (Table 1), and this was important for shaping their respective differences in gut microbiomes. Differences were determined with the Euclidian distance matrix (for presence/absence of taxa) and the Bray Curtis distance matrix (for relative microbial abundance). Faecal pH was correlated with the total number of microbial OTUs such that higher faecal pH seemed to support more OTUs.

Microbial richness correlated negatively with total carbohydrate but positively with simple carbohydrates (sugar). Calprotectin seemed to have no influence over the microbiome in terms of the number of OTUs (Fig 4). In addition, there was no correlation between this inflammatory marker and gut microbial richness.

Discussion

This is the first study about the fecal microbiota of GSD patients. In hepatic GSD, high and periodic amounts of UCCS plus dietetic restriction of fast-digestion carbohydrates are the main way to treat the genetic impairment in the glycogenolytic pathway. Our data suggest that the overload of UCCS can lead to low fecal pH by favouring some bacterial genera capable of utilizing complex carbohydrates in detriment of others. The low fecal pH, in turn, also acts as an environmental selection factor to the bacteria in the lumen. Dysbiosis has been associated with IBD and obesity. IBD includes inflammatory bowel diseases of unknown aetiology and has two main forms: ulcerative colitis and Crohn's disease (CD). CD is a chronic disease that can affect any region in the digestive tract but is more likely to involve the small and large intestines and the perianal region [33]. Enteropathy is related to type I patients, and despite GSD Ib patients are classically described as prone to IBD-Crohn's-like due the impaired neutrophil activity, this does not explain why patients with GSD Ia also displayed serologic markers altered for IBD, even if asymptomatic [34]. It's not clear if UCCS is the cause of obesity in GSD patients [35], but the microbiome might be associated with it. Here we discuss why the changes in microbiota could be considered as a factor of influence in the phenotype of these patients and why the UCCS usage, even though not exclusively, is an important factor that contribute to that.

Since the introduction of UCCS treatment for GSD, the focus changed from mortality to morbidity and control of long-term complications [36], such as metabolic syndrome and related symptoms [37,38]. GSD type I patients are prone to obesity, and it is suspected that UCCS contributes to the aforementioned features [35,39]. GSD I patients also are subject of heavier doses of UCCS and more restrict diet in

comparison with types III and IX [35]. Regarding antibiotics, although its usage clearly drives changes in the gut microbial community, subjects who were treated with antibiotics within 6 months prior to data collection, but not during the study itself, were not affected by the previous antibiotic usage.

We found that the phyla *Actinobacteria* and *Proteobacteria* were overrepresented in patients while the *Euryarchaeota* was underrepresented. The microbiome of GSD patients present low diversity and was highly dominated by *Escherichia*/*Shigella*.

One possible driver of the differences in gut microbiomes between patients and controls is UCCS overload, which creates an acidic environment [34,40]. In the human body, acids are generated by regular metabolic activities and through the daily intake of food [41]. Fecal pH was lower in patients than controls and stool acidification might lead to an alteration in the relative abundances of fermenting bacteria, decreasing the conversion of unabsorbable starches to short chain fatty acids (SCFAs) [34].

SCFAs, including butyrate, are compounds made by bacteria in the gut that affect several physiologic functions and serve anti-inflammatory roles [42]. Fecal pH was associated with beta diversity and bacterial families belonging to the *Clostridia* class, an important producer of butyrate in the gut. Several genera of SCFA-producing bacteria—*Coprococcus, Blautia, Anaerostipes, Odoribacter* and *Faecalibacterium*—were decreased in patients. Those genera were also identified in paediatric patients with Crohn's Disease [43]. Besides, *Coprococcus* and *Faecalibacterium* were found to have significantly low abundance in patients with nonalcoholic fatty liver disease, independently of body mass index and insulin resistance [43].

The bacterial species residing within the mucous layer of the colon may influence whether host cellular homeostasis is maintained or inflammatory mechanisms are triggered. A mutualistic relationship between the colonic microbiota, their metabolic products and the host immune system is likely involved [44]. The phylum Proteobacteria was more abundant in patients than in controls while the phylum *Euryarchaeota* was less abundant. *Proteobacteria* is a gram-negative phylum with an outer membrane mainly composed of lipopolysaccharides (LPS), which are known to sustain systemic levels of low-grade inflammation [45]. Higher levels of *Proteobacteria* can be considered a strong marker of dysbiosis [46]. This phylum is prevalent in patients with liver cirrhosis [47]. Several serological markers for IBD were altered in GSD-Ia patients [34], and GSD Ib patients are prone to IBD CD-like. Despite the fact that calprotectin seemed not to influence the number of OTUs gut inflammation (calprotectin >50µg/g) was verified in several patients. GSD type Ib patients have shown a concentration of calprotectin 50µg/g, except for one patient, who had an active IBD diagnosed in the same week. This might be due to a remission state and the use of anti-inflammatory mesalazine by these patients.

In general, dysbiosis can be categorized as a) loss of beneficial organisms, b) excessive growth of potentially harmful organisms and c) loss of overall microbial diversity. These three categories often occur at the same time [48]. Dysbiosis has been implicated in a wide range of diseases, including IBD, liver disease and obesity, that are secondary manifestations in GSD patients [49]. The reason for dysbiosis remains unclear, but the overload of UCCS contributes to those characteristics. The food intake records showed a difference in the intake of calories, mainly due to the administration of UCCS in patients, as well as a difference in microbial signature that is known to be

related to obesity. It is not known whether these microbiome changes are a cause or a consequence of the pathophysiologies. However, correcting the dysbiosis can improve health in some patients [50–52]. Dysbiosis can also provide biomarkers for disease detection and management [53].

Conclusion

In this study, we reported significant alterations in the intestinal environments of GSD patients versus healthy controls. Microbiota can be affected by abiotic and biotic factors, namely pH and inflammation, and the differences in these factors between patients and controls might be linked to both genetic disease and UCCS consumption. Several bacterial taxa were different in GSD patients than in controls, and those groups are consistent with the secondary phenotypic manifestations of GSD. The microbiome patterns of these patients may reinforce immunemetabolic pathways that already are altered by genetic impairment, and may also be a factor in the differential individual response to treatment. Patients may gain health and quality of life from the restoration of gut microbial diversity that has been diminished by high UCCS intake. Future research therefore should investigate ways to manipulate the gut microbiome and clarify the possible effects of doing so.

Acknowledgments

We thank all subjects for allowing us to collect samples and information for the study. PRO-

NEX-FAPERGS, FIPE-HCPA, CAPES, CNPq

Author Contributions

Conceptualization: Karina Colonetti, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Data curation: Karina Colonetti, Bruna Bento dos Santos, Tatiéle Nalin, Priscila Thiago Dobbler, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Formal analysis: Karina Colonetti, Bruna Bento dos Santos, Tatiéle Nalin, Carolina Fischinger Moura de Souza, Eric W. Triplett, Priscila Thiago Dobbler, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Funding acquisition: Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Investigation: Karina Colonetti, Bruna Bento dos Santos, Priscila Thiago Dobbler.

Project administration: Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Resources: Carolina Fischinger Moura de Souza, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Supervision: Ida Vanessa Doederlein Schwartz.

Validation: Eric W. Triplett.

Visualization: Karina Colonetti, Luiz Fernando Wurdig Roesch.

Writing – original draft: Karina Colonetti, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

Writing – review & editing: Bruna Bento dos Santos, Tatiéle Nalin, Carolina Fischinger Moura de Souza, Eric W. Triplett, Priscila Thiago Dobbler, Ida Vanessa Doederlein Schwartz, Luiz Fernando Wurdig Roesch.

References

1. Ozen H. Glycogen storage diseases: New perspectives. World J Gastroenterol. 2007; 13: 2541. https:// doi.org/10.3748/wjg.v13.i18.2541 PMID: 17552001
2. Chou JY, Jun HS, Mansfield BC. Glycogen storage disease type I and G6Pase-β deficiency: etiology and therapy. Nat Rev Endocrinol. 2010; 6: 676–688. https://doi.org/10.1038/nrendo.2010.189 PMID: 20975743
3. Bali DS, Chen YT, Austin S, Goldstein JL, Adam MP, Ardinger HH, et al. Glycogen Storage Disease Type I. 2006 Apr 19 [Updated 2016 Aug 25]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2019. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1312/.
4. Sentner CP, Hoogeveen IJ, Weinstein DA, Santer R, Murphy E, McKiernan PJ, et al. Glycogen storage disease type III: diagnosis, genotype, management, clinical course and outcome. J Inherit Metab Dis. 2016; 39: 697–704. https://doi.org/10.1007/s10545-016-9932-2 PMID: 27106217
5. Kishnani PS, Austin SL, Arn P, Bali DS, Boney A, Case LE, et al. Glycogen Storage Disease Type III diagnosis and management guidelines. Genet Med. 2010; 12: 446–463. https://doi.org/10.1097/GIM. 0b013e3181e655b6 PMID: 20631546
6. Herbert M, Goldstein JL, Rehder C, Austin S, Kishnani PS, Bali DS. Phosphorylase Kinase Deficiency. 2011 May 31 [Updated 2018 Nov 1]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993–2019. Available from: https:// www.ncbi.nlm.nih.gov/books/NBK55061/.
7. Weinstein DA, Wolfsdorf JI. Effect of continuous glucose therapy with uncooked cornstarch on the longterm clinical course of type 1a glycogen storage disease. Eur J Pediatr. 2002; 161 Suppl 1: S35–39. https://doi.org/10.1007/s00431-002-1000-2
8. Kishnani PS, Austin SL, Abdenur JE, Arn P, Bali DS, Boney A, et al. Diagnosis and management of glycogen storage disease type I: a practice guideline of the American College of Medical Genetics and Genomics. Genet Med. 2014; 16: e1–e1. https://doi.org/10.1038/gim.2014.128 PMID: 25356975
9. Sleator RD. The human superorganism—of microbes and men. Med Hypotheses. 2010; 74: 214–215. https://doi.org/10.1016/j.mehy.2009.08.047 PMID: 19836146
10. Zhang YJ, Li S, Gan RY, Zhou T, Xu DP, Li HB. Impacts of gut bacteria on human health and diseases. Int J Mol Sci. 2015; 16: 7493–519. https://doi.org/10.3390/ijms16047493 PMID: 25849657
11. Friedrich MJ. Genomes of Microbes Inhabiting the Body Offer Clues to Human Health and Disease. JAMA. 2013; 309: 1447. https://doi.org/10.1001/jama.2013.2824 PMID: 23571560
12. Consortium HMP. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486: 207–14. https://doi.org/10.1038/nature11234 PMID: 22699609
13. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. Nat Rev Genet. 2017; 18: 690–699. https://doi.org/10.1038/nrg.2017.63 PMID: 28824167
14. Marchesi JR, Adams DH, Fava F, Hermes GDA, Hirschfield GM, Hold G, et al. The gut microbiota and host health: a new clinical frontier. Gut. 2016; 65: 330–339. https://doi.org/10.1136/gutjnl-2015-309990 PMID: 26338727
15. Rake JP, Visser G, Labrune P, Leonard JV, Ullrich K, Smit GPA. Guidelines for management of glycogen storage disease type I—European Study on Glycogen Storage Disease Type I (ESGSD I). Eur J Pediatr. 2002; 161 Suppl 1: S112–119. https://doi.org/10.1007/s00431-002-1016-7

16. Visser G, Rake JP, Labrune P, Leonard JV, Moses S, Ullrich K, et al. Consensus guidelines for management of glycogen storage disease type 1b - European Study on Glycogen Storage Disease Type 1. Eur J Pediatr. 2002; 161 Suppl 1: S120–123. https://doi.org/10.1007/s00431-002-1017-6
17. WHO MULTICENTRE GROWTH REFERENCE STUDY GROUP, Onis M. WHO Child Growth Standards based on length/height, weight and age: WHO Child Growth Standards. Acta Paediatr. 2007; 95:76–85. https://doi.org/10.1111/j.1651-2227.2006.tb02378.x
18. Barboza ADM, Pylro VS, Jacques RJS, Gubiani PI. Seasonal dynamics alter taxonomical and functional microbial profiles in Pampa biome soils under natural grasslands. 2018; 18.
19. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012; 6: 1621–1624. https://doi.org/10.1038/ismej.2012.8 PMID: 22402401
20. Pylro VS, Morais DK, de Oliveira FS, dos Santos FG, Lemos LN, Oliveira G, et al. BMPOS: a flexible and user-friendly tool sets for microbiome studies. Microb Ecol. 2016; 72: 443–447. https://doi.org/10. 1007/s00248-016-0785-x PMID: 27220974
21. Pylro VS, Roesch LFW, Morais DK, Clark IM, Hirsch PR, To´tola MR. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods. 2014; 107: 30–37. https://doi.org/10.1016/j.mimet.2014.08.018 PMID: 25193439
22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010; 7: 335–336. https://doi. org/10.1038/nmeth.f.303 PMID: 20383131
23. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2012; 41: D590– D596. https://doi.org/10.1093/nar/gks1219 PMID: 23193283
24. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW. Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods. 2011; 86: 42–51. https://doi.org/10.1016/j.mimet. 2011.03.014 PMID: 21457733
25. Lemos LN, Fulthorpe RR, Roesch LF. Low sequencing efforts bias analyses of shared taxa in microbial communities. Folia Microbiol (Praha). 2012; 57: 409–413.
26. R Development Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2008. Available: http://www.R-project.org
27. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 2013; 8: e61217. https://doi.org/10.1371/journal.pone. 0061217 PMID: 23630581
28. Oksanen J, Blanchet F G, Kindt R, Legendre P, Minchin P R, O'Hara R B, et al. Vegan: community ecology package. R package vegan, vers. 2.2–1. 2015.
29. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 2017; https://doi.org/10.1093/nar/gkx295
30. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013; 10: 1200. https://doi.org/10.1038/nmeth.2658 PMID: 24076764
31. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011; 12: R60. https://doi.org/10.1186/gb-2011-12-6-r60 PMID: 21702898
32. Bjarnason I. The Use of Fecal Calprotectin in Inflammatory Bowel Disease.: 4.
33. Matsuoka K, Kobayashi T, Ueno F, Matsui T, Hirai F, Inoue N, et al. Evidence-based clinical practice guidelines for inflammatory bowel disease. J Gastroenterol. 2018; 53: 305–353. https://doi.org/10.1007/ s00535-018-1439-1 PMID: 29429045
34. Lawrence NT, Chengsupanimit T, Brown LM, Weinstein DA. High Incidence of Serologic Markers of Inflammatory Bowel Disease in Asymptomatic Patients with Glycogen Storage Disease Type Ia. In: Zschocke J, Baumgartner M, Morava E, Patterson M, Rahman S, Peters V, editors. JIMD Reports, Volume 24. Berlin, Heidelberg: Springer Berlin Heidelberg; 2015. pp. 123–128. https://doi.org/10.1007/ 8904_2015_452 PMID: 26093626
35. dos Santos BB, Nalin T, Grokoski KC, Perry IDS, Refosco LF, Vairo FP, et al. Nutritional Status and Body Composition in Patients With Hepatic Glycogen Storage Diseases Treated With Uncooked Cornstarch—A Controlled Study. J Inborn Errors Metab Screen. 2017; 5: 232640981773301. https://doi.org/10.1177/2326409817733014
36. Moses SW. Historical highlights and unsolved problems in glycogen storage disease type 1. Eur J Pediatr. 2002; 161 Suppl 1: S2–9. https://doi.org/10.1007/s00431-002-0997-6

37. Melis D, Rossi A, Pivonello R, Salerno M, Balivo F, Spadarella S, et al. Glycogen storage disease type Ia (GSDIa) but not Glycogen storage disease type Ib (GSDIb) is associated to an increased risk of metabolic syndrome: possible role of microsomal glucose 6-phosphate accumulation. Orphanet J Rare Dis. 2015; 10. https://doi.org/10.1186/s13023-015-0301-2
38. Derks TGJ, van Rijn M. Lipids in hepatic glycogen storage diseases: pathophysiology, monitoring of dietary management and future directions. J Inherit Metab Dis. 2015; 38: 537–543. https://doi.org/10.1007/ s10545-015-9811-2 PMID: 25633903
39. Jelaska BK, Ostojić SB, Berović N, Kokić V. Continuous glucose monitoring in the treatment of obesity in patients with glycogen storage disease type Ia. Endocrinol Diabetes Metab Case Rep. 2014;2014. https://doi.org/10.1530/EDM-13-0056
40. Kotarski SF, Waniska RD, Thurn KK. Starch hydrolysis by the ruminal microflora. J Nutr. 1992; 122: 178–190. https://doi.org/10.1093/jn/122.1.178 PMID: 1370325
41. Bagchi B. Water in Biological and Chemical Processes: From Structure and Dynamics to Function [Internet]. Cambridge: Cambridge University Press; 2013. https://doi.org/10.1017/ CBO9781139583947
42. Morrison DJ, Preston T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. Gut Microbes. 2016; 7: 189–200. https://doi.org/10.1080/19490976.2015.1134082 PMID: 26963409
43. Da Silva HE, Teterina A, Comelli EM, Taibi A, Arendt BM, Fischer SE, et al. Nonalcoholic fatty liver disease is associated with dysbiosis independent of body mass index and insulin resistance. Sci Rep. 2018; 8. https://doi.org/10.1038/s41598-018-19753-9
44. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. Microb Ecol Health Dis. 2015; 26. https://doi.org/10.3402/mehd.v26.26191
45. Guo H, Diao N, Yuan R, Chen K, Geng S, Li M, et al. Subclinical-Dose Endotoxin Sustains Low-Grade Inflammation and Exacerbates Steatohepatitis in High-Fat Diet–Fed Mice. J Immunol. 2016; 196: 2300–2308. https://doi.org/10.4049/jimmunol.1500130 PMID: 26810228
46. Shin N-R, Whon TW, Bae J-W. Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol. 2015; 33: 496–503. https://doi.org/10.1016/j.tibtech.2015.06.011 PMID: 26210164
47. Chen Y, Yang F, Lu H, Wang B, Chen Y, Lei D, et al. Characterization of fecal microbial communities in patients with liver cirrhosis. Hepatol Baltim Md. 2011; 54: 562–572. https://doi.org/10.1002/hep.24423
48. DeGruttola AK, Low D, Mizoguchi A, Mizoguchi E. Current Understanding of Dysbiosis in Disease in Human and Animal Models: Inflamm Bowel Dis. 2016; 22: 1137–1150. https://doi.org/10.1097/MIB. 0000000000000750 PMID: 27070911
49. Rake JP, Visser G, Labrune P, Leonard JV, Ullrich K, Smit GPA. Glycogen storage disease type I: diagnosis, management, clinical course and outcome. Results of the European Study on Glycogen Storage Disease Type I (ESGSD I). Eur J Pediatr. 2002; 161 Suppl 1: S20–34. https://doi.org/10.1007/s00431002-0999-4
50. Di Luccia B, Crescenzo R, Mazzoli A, Cigliano L, Venditti P, Walser J-C, et al. Rescue of FructoseInduced Metabolic Syndrome by Antibiotics or Faecal Transplantation in a Rat Model of Obesity. Heimesaat MM, editor. PLOS ONE. 2015; 10: e0134893. https://doi.org/10.1371/journal.pone.0134893 PMID: 26244577
51. Vrieze A, Van Nood E, Holleman F, Salojarvi J, Kootte RS, Bartelsman JFWM, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012; 143: 913–916.e7. https://doi.org/10.1053/j.gastro.2012.06.031 PMID: 22728514
52. Carnero-Gregorio M, Molares-Vila A, Corbalan-Rivas A, Villaverde-Taboada C, Rodriguez-Cerdeira C. Effect of VSL#3 Probiotic in a Patient with Glycogen Storage Disease Type Ia and Irritable Bowel Disease-like Disease. Probiotics Antimicrob Proteins. 2018; https://doi.org/10.1007/s12602-017-9372-9
53. Pascal V, Pozuelo M, Borruel N, Casellas F, Campos D, Santiago A, et al. A microbial signature for Crohn's disease. Gut. 2017; 66: 813–822. https://doi.org/10.1136/gutjnl-2016-313235 PMID: 28179361

Figures

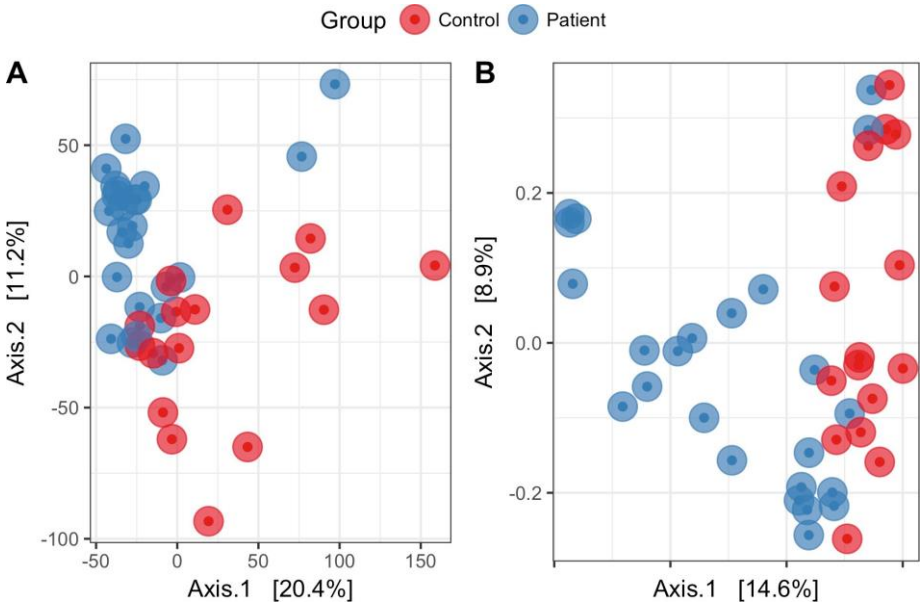


Fig 1. Principal coordinates analysis (PCoA) based on Bray Curtis distance matrix (A) and Euclidean distance matrix (B) show the separation of gut microbiomes between GSD patients and controls. Each point represents a microbial community from one subject; colours indicate the treatment.

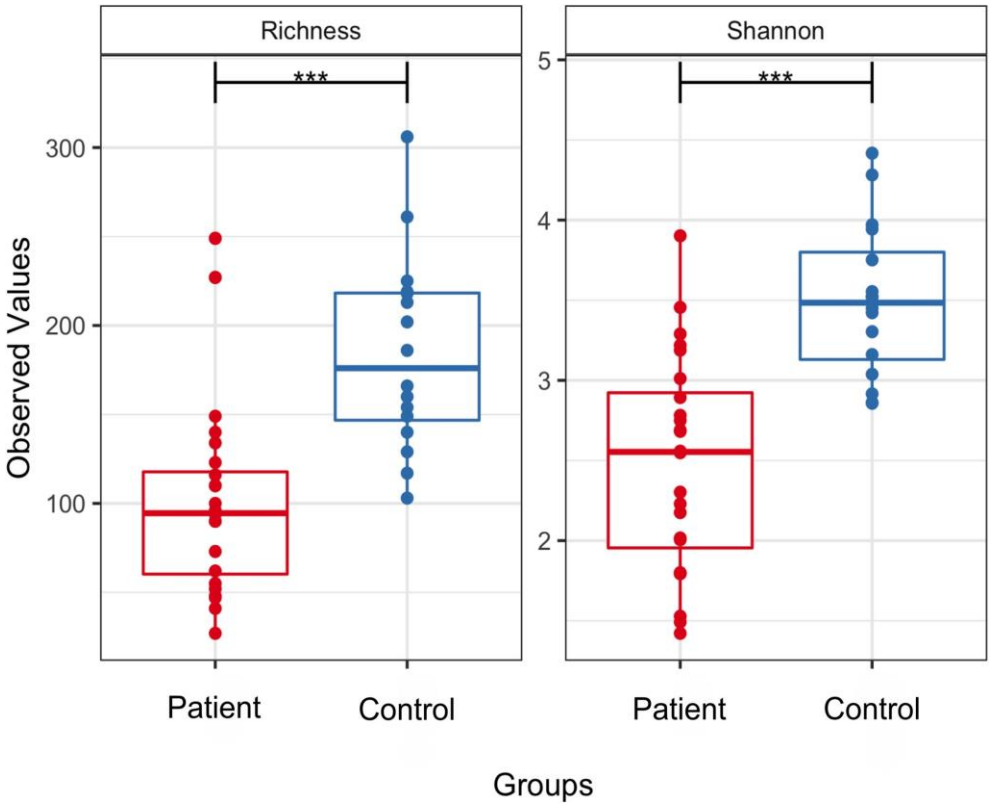


Fig 2. Alpha diversity measurements of microbial communities in the GSD patients and control groups. Each panel represents one alpha diversity measure: Richness = total number of OTUs observed, Shannon = microbial index of diversity. Boxes span the first to third quartiles; the horizontal line within the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles. indicates a statistical difference between treatments at cutoff p 0.001.

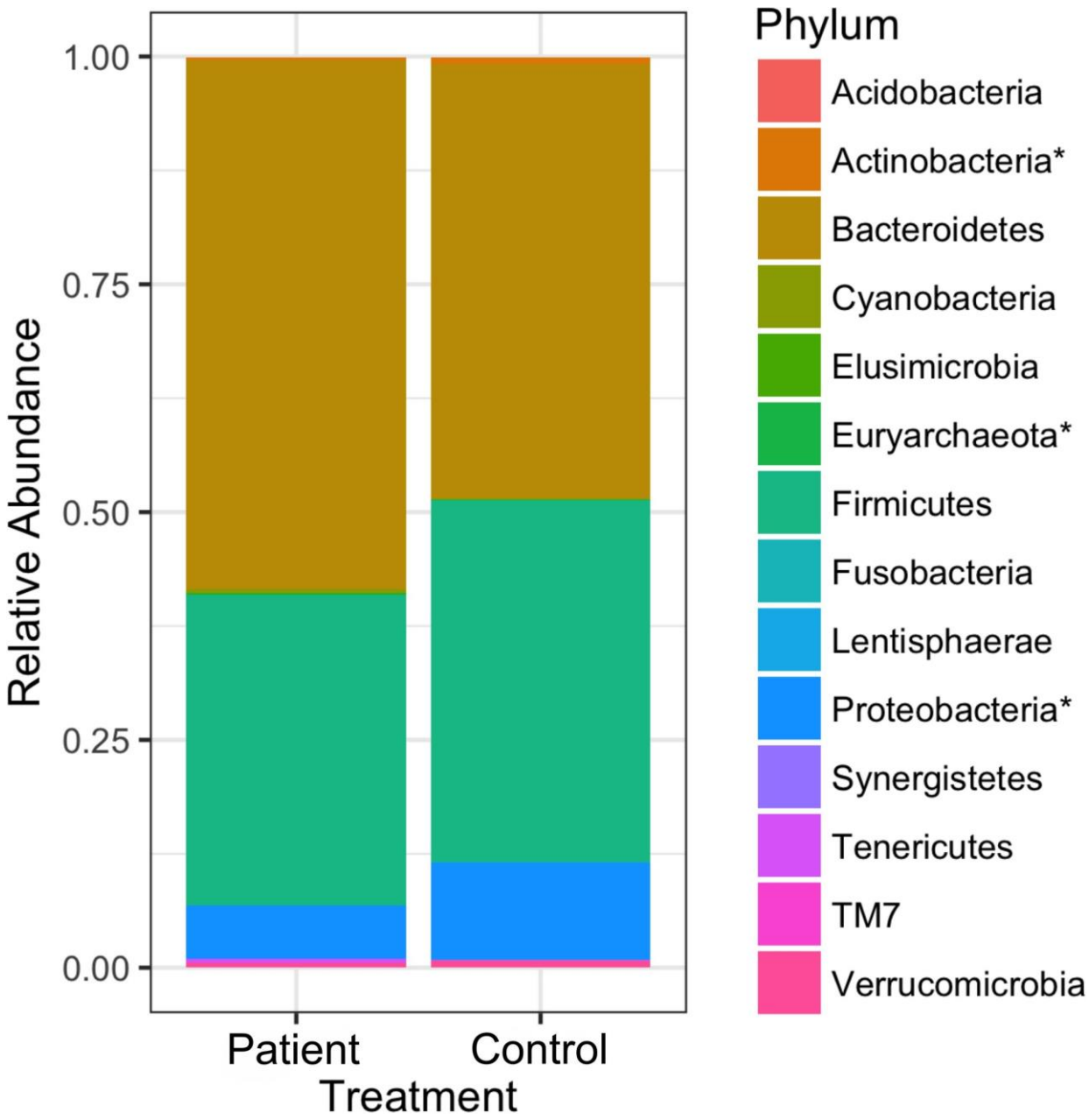


Fig 3. The average relative abundance of phyla found in GSD patients and healthy controls. Phyla followed by an asterisk () are different, both in terms of statistics and biological consistency, between patients and controls at p and FDR 0.05: Euryarchaeota (LDA score = 1.75), Actinobacteria (LDA score = 3.06) and Proteobacteria (LDA score = 3.94). Firmicutes was marginally significantly different with p = 0.064, LDA score = 4.52 and FDR = 0.112.

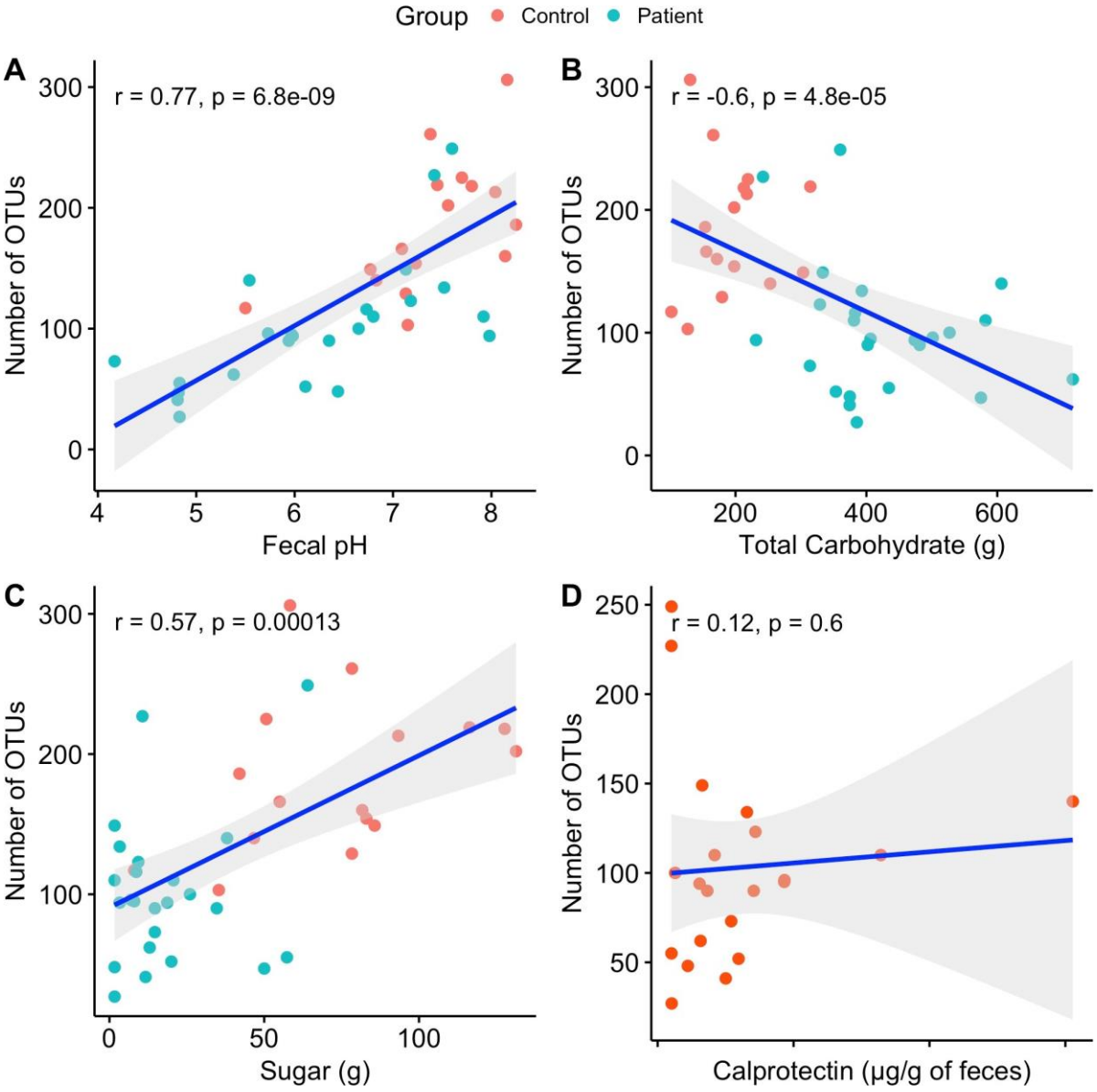


Fig 4. Correlations between the microbiota and diet, faecal pH, and gut inflammation.

Tables

Table 1. Sample characterization, analysis of potential confounding variables and their effect on microbial communities.

Variable[1]	Patients (n = 24)	Controls (n = 16)	p-value	Microbial community difference between patients and controls			
				Euclidian Metric		Bray-Curtis Metric	
				R2	p-value	R2	p-value
Sex (M/F)	14/10	07/09	0.561	0.02942	0.287	0.02964	0.267
Age (yr)	12 (10–19.75)	12.5 (10–23.25)	0.579	0.02895	0.302	0.02775	0.340
Faecal pH	6.23 (5.42–7.16)	7.41 (7.10–7.98)	0.001	0.05938	0.005	0.08507	0.001
Inflammatory Bowel Disease (yes/no)	04/20	00/16	0.136	0.06746	0.009	0.05152	0.003
Abdominal pain complaint (yes/no)	09/15	01/15	0.032	0.05590	0.010	0.04845	0.009
Nutritional status (Obese or Overweight/Normal)	18/06	06/09†	0.044	0.05199	0.004	0.03423	0.121
UCCS intake (g/day)	309.50 (373.7–245.3)	00	0.001	0.03698	0.114	0.05594	0.001
Drugs (yes/no):							
-Allopurinol	4/20	0/16	0.136	0.02477	0.436	0.02426	0.517
-Antibiotic usage (last 6 months)	10/14	3/13	0.241	0.03047	0.252	0.03200	0.179
-ACE inhibitor	11/13	0/16	0.001	0.03351	0.203	0.03919	0.054
-Filgrastim (G-CSF)	5/19	0/16	0.071	0.06654	0.002	0.05377	0.008
-Mesalazine	3/21	0/16	0.262	0.03089	0.290	0.03389	0.109
-Multivitamin	22/2	1/15	0.001	0.04034	0.070	0.05545	0.003
-Potassium Citrate	3/21	0/16	0.262	0.02248	0.516	0.02407	0.551
-Proton Pump Inhibitors	2/22	0/16	0.508	0.03068	0.318	0.03087	0.173
-Statins	1/23	0/16	1.000	0.03312	0.286	0.02542	0.486

UCCS: uncooked cornstarch; ACE: Angiotensin-converting-enzyme inhibitor (enalapril maleate); G-CSF: G-colony stimulating factor. Significant ($p<0.05$) events are highlighted in bold.

1 Numeric variables were reported as medians (Q1-Q3). Due to the not-normal distribution, numeric variables were subjected to the Mann-Whitney test. Qualitative variables were reported as absolute frequency and tested by X2, Fisher's test or Continuity Correction, as appropriate.

† Data for one control was missing. Weight and height were measured when subjects delivered the sample. In this case, a relative drove the sample to the hospital, thus we were unable to do so.

Table 2. Microbial biomarkers differentiating patients with hepatic glycogenosis diseases and healthy controls.

Microbial genus	Patients	Controls	p-values	FDR	LDA score
	Relative abundance (%)				(log 10)
Full dataset	n = 24	n = 16			
Lactobacillus	11.31	0.04	0.009	0.025	4.36
Escherichia/Shigella	6.70	0.96	0.003	0.013	3.89
Alistipes	2.77	9.12	0.005	0.018	-3.22
Subdoligranulum	1.59	1.00	0.012	0.029	2.42
Lachnospiraceae NK4A136 group	1.44	0.89	0.003	0.013	2.48
Faecalibacterium	1.00	3.52	0.016	0.036	-2.98
Ruminococcaceae UCG 002	0.98	3.09	0.001	0.007	-2.79
Bifidobacterium	0.78	0.19	0.004	0.018	3.1
Ruminococcus gnavus group	0.70	0.14	0.007	0.022	3.03
Phascolarctobacterium	0.53	1.31	0.015	0.035	-2.56
Blautia	0.26	0.53	0.002	0.012	-1.55
Odoribacter	0.25	0.53	0.011	0.028	-1.87
Barnesiella	0.22	0.98	0.009	0.025	-2.46
Roseburia	0.18	1.19	0.002	0.011	-2.78
Christensenellaceae R 7 group	0.14	0.80	0.000	0.002	-2.22
Ruminococcaceae UCG 003	0.10	0.60	0.000	0.003	-2.27
Lachnospiraceae UCG 008	0.04	0.26	0.004	0.018	-1.78
Ruminococcaceae UCG 005	0.03	0.25	0.000	0.002	-1.9
Eubacterium hallii group	0.02	0.08	0.000	0.002	-1.39
Anaerostipes	0.01	0.11	0.001	0.009	-1.55
Coprococcus 1	0.01	0.03	0.000	0.005	-0.95
Family XIII AD3011 group	0.01	0.05	0.000	0.002	-1.21
Family XIII UCG 001	0.00	0.03	0.001	0.007	-1.13
Methanobrevibacter	0.00	0.17	0.001	0.007	-1.78
Ruminococcaceae NK4A214 group	0.00	0.08	0.001	0.007	-1.5
Dataset without IBD-like patients	n = 20	n = 14			
Escherichia/Shigella	6.47	0.92	0.003	0.027	3.85
Alistipes	2.97	9.76	0.008	0.039	-3.28
Ruminococcaceae UCG 002	1.12	3.07	0.004	0.028	-1.38
Bifidobacterium	0.81	0.08	0.003	0.027	3.2
Phascolarctobacterium	0.22	1.38	0.004	0.028	-2.74
Christensenellaceae R 7 group	0.17	0.76	0.001	0.016	-2.16
Blautia	0.14	0.39	0.001	0.017	-2.08
Ruminococcaceae UCG 003	0.11	0.61	0.001	0.016	-2.3
Roseburia	0.10	1.15	0.004	0.028	-2.83

Lachnospiraceae UCG 008	0.04	0.19	0.011	0.047	-1.57
Ruminococcaceae UCG 005	0.03	0.20	0.001	0.016	-1.76
Eubacterium hallii group	0.02	0.07	0.000	0.016	-1.32
Anaerostipes	0.01	0.07	0.010	0.047	-1.28
Coprococcus 1	0.01	0.02	0.008	0.039	-0.77
Family XIII AD3011 group	0.01	0.04	0.001	0.017	-1.14
Family XIII UCG 001	0.00	0.03	0.001	0.016	-1.15
Methanobrevibacter	0.00	0.17	0.003	0.027	-1.81
Ruminococcaceae NK4A214 group	0.00	0.08	0.003	0.027	-1.53

*Four IBD-like (Inflammatory Bowel Disease) patients and matched controls were excluded from the dataset to make sure the biomarkers observed were not only driven by these patients.

APÊNDICE B - Influência do leite materno e diferentes proporções da fórmula na microbiota intestinal de recém-nascidos muito prematuros

Influence of own mother's milk and different proportions of formula on intestinal microbiota of very preterm newborns

Adriana Zanella[1], Rita C. Silveira[1], Luiz F. W. Roesch[2], Andréa L. Corso[1], Priscila T. Dobbler[2], Volker Mai[3], Renato S. Procianoy[1]*

[1] Unidade de Neonatologia do Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil,
[2] Centro Interdisciplinar de Pesquisas em Biotecnologia–CIP- Biotec, Campus São Gabriel, Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil,
[3] Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, Emerging Pathogens Institute, University of Florida, Gainesville, Florida, United States of America
* rprocianoy@gmail.com

DOI: https://doi.org/10.1371/journal.pone.0217296

Received: October 24, 2018

Accepted: May 8, 2019

Published: May 20, 2019

Abstract

Objective
To determine the differences in preterm infants' stool microbiota considering the use of exclusive own mother's milk and formula in different proportions in the first 28 days of life.

Methods
The study included newborns with GA 32 weeks divided in 5 group according the feeding regimen: 7 exclusive own mother's milk, 8 exclusive preterm formula, 16 mixed feeding with >70% own mother's milk, 16 mixed feeding with >70% preterm formula, and 15 mixed 50% own mother's milk and preterm formula. Exclusion criteria: congenital infections, congenital malformations and newborns of drug addicted mothers. Stools were collected weekly during the first 28 days. Microbial DNA extraction, 16S rRNA amplification and sequencing were performed.

Results
All groups were similar in perinatal and neonatal data. There were significant differences in microbial community among treatments. Approximately 37% of the variation in distance between microbial communities was explained by use of exclusive own mother´s milk only compared to other diets. The diet composed by exclusive own mother´s milk allowed for greater microbial richness (average of 85 OTUs) while diets based on preferably formula, exclusive formula, preferably maternal milk, and mixed of formula and maternal milk presented an average of 9, 29, 23, and 25 OTUs respectively. The mean proportion of the genus *Escherichia and Clostridium* was always greater in those containing formula than in the those with maternal milk only.

Conclusions
Fecal microbiota in the neonatal period of preterm infants fed with exclusive own mother's milk presented increased richness and differences in microbial composition from those fed with different proportions of formula.

Introduction

The intestinal microbiota is very important for human metabolism, development and behavior [1,2]. Despite several studies on the subject and its connection with high complexity diseases [1,3,4], the studies were based on culture, genetic profile and/or the use of small sample sizes, which makes it clear that the variables responsible for shaping the intestinal microbiota have not been satisfactorily examined [1,5,6]. It is known that the development of infant microbiota depends on medical and dietary factors [1,7], but it is not yet known how such factors influence the microbial overall composition and their associations with the human body [1]. The human body has millions of microorganisms that work in partnership with our own cells to influence the quality of our lifelong health [8,9]. The composition of the childhood intestinal microbiome is influenced by factors such as the type of birth, gestational and postnatal age, ingestion of antibiotics, environment, nutritional exposures, and breastfeeding,

which should be emphasized as an important variable for the assembly of the intestinal microbiota [5,8,10,11,12]. La Rosa et al have shown that the gut microbioma of premature infants admitted to Neonatal Intensive Care Units progresses and bacterial population changes in composition along the time [6]. Despite the influence of breast milk versus formula in the assembly of the gut microbiota, the true impact of own mother's breast milk on the composition of the intestinal microbiome of premature infants is not fully understood [8].

The immature intestinal microbiota of premature newborns is influenced by factors such as postnatal age, gestational age, birth weight and nutritional exposures [8,13,14,15]. In the specific case of breastfeeding, this seems to disguise the influence of birth weight, which suggests a protective function against the intestinal immaturity of the premature newborn at the onset of life [8]. These findings suggest not only the existence of a microbial mechanism underlying the body of evidence that elucidates that breast milk promotes the intestinal health of the premature newborn [16], but also the dynamic interaction of host and dietary factors that help in the colonization and enrichment of specific microbes during the establishment of its intestinal microbiota [8].

Therefore, the aim of this study is to describe the intestinal microbiota of very low birth weight infants in the first 28 days depending on use of mother's own milk or use of formula in different proportions.

Material and methods

This study used a convenience sampling strategy with patients recruited from the Neonatology Section of Hospital de Clínicas de Porto Alegre (HCPA), Brazil from May 2014 to January 2017. Pregnant women with gestation age 32 weeks that provided written informed consent were enrolled at hospital admission for their delivery. The study protocol was approved by the Ethics Committee of Hospital de Clínicas de Porto Alegre (HCPA), and the guardians signed an informed consent form. Exclusion criteria were: 1) HIV or congenital infections, 2) mothers with substance abuse and 3) neonatal congenital malformations. Infants´ weekly stool samples were collected from diapers beginning with the first stool until the 4th week of life. All samples were immediately stored in liquid nitrogen until DNA extraction.

Newborns were divided in five groups according to the feeding: exclusive own mother's milk (LME), exclusive formula (PFL), mixed 50% own mother's milk and 50% formula (MFLM), mixed with formula and 70% or more of own mother's milk (PLM), and mixed with own mother´s milk and 70% or more of formula (PFL). The newborns received daily the same diet for up to 28 days. The different amounts of formula and own mother´s milk was offered separately, at different times of the day, so that at the end of the day the ratio was maintained. The ratio of breast milk to formula was the limitation of the amount of breast milk available. The preterm formula used was Pre Nan. This formula does not contain neither probiotics nor prebiotics.

Obstetrical data and neonatal data were collected prospectively.

DNA extraction and the 16S rRNA library preparation

The laboratory technique as described in detail previously [16].

Microbial DNA was isolated from samples using the QIAmp Fast DNA Stool Mini Kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions with modifications. All DNA samples were kept at -80˚C until use. The V4 region of the 16S rRNA gene was amplified and sequenced using the PGM Ion Torrent platform (Thermo Fisher Scientific, Waltham, MA, USA) with the barcoded bacterial/archaeal primers 515F and 806R [17]. PCR amplification was carried out using barcoded primers linked with the Ion adapter "A" sequence (5'-CCAT CTCATCCCTGCGTGTCTCCGACTCAG-3') and Ion adapter P1 sequence (5'-CCTCTCTATGGGCAGTCGGTGAT-3') to obtain a sequence of primer composed for A-barcode-806R and P1-515F adapter and primers. Each of the 25μL of PCR mixture consisted of 2U of Platinum Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4μL 10X High Fidelity PCR Buffer, 2 mM MgSO4, 0.2 mM dNTP's, 0.1 μM both the 806R barcoded primer and the 515F primer, 25μg of Ultrapure BSA (Invitrogen, Carlsbad, CA, USA) and approximately 50 ng of DNA template. PCR conditions used were: 95˚C for 5 min, 35 cycles of denaturation at 94˚C for 45s; annealing at 56˚C for 45s and extension at 72˚C for 1 min; followed by a final extension step at 72˚C for 10 min.

The resulting PCR products were purified with the Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, USA) and the final concentration of the PCR product was quantified by using the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA) following manufacturer's recommendations. Finally, the reactions were combined in equimolar concentrations to create a mixture composed by 16S gene amplified fragments of each sample. This composite sample was used for library preparation with Ion OneTouch 2 System with the Ion PGM Template OT2 400 Kit Template (Thermo Fisher Scientific, Waltham, MA, USA). The sequencing was performed using Ion PGM Sequencing 400 on Ion PGM System using Ion 318 Chip v2 with a maximum of 40 samples per microchip.

Sequence processing

The 16S rRNA raw sequences were analyzed following the recommendations of the Brazilian Microbiome Project [18]. Briefly, the OTU (Operational Taxonomic Unit) table was built using the UPARSE pipeline [19], in which the reads were truncated at 200 bp and quality filtered using a maximum expected error of 0.5. Filtered reads were dereplicated and singletons were removed. After clustering sequences into OTUs, with a similarity cutoff of 97%, chimeras were checked to obtain representative sequences for each microbial phylotype. Taxonomic classification was carried out using SINTAX [20] against the Ribosomal Database Project (RDP) database [21] with a confidence threshold of 80%. Sampling effort was estimated using Good's coverage [22].

Statistical analysis. All clinical data was analyzed using the software SPSS 21.0 at the significance level of 5%. Quantitative variables with normal distribution were described through means/SD and compared by ANOVA with Tukey test. Quantitative variables with asymmetric distribution were described through median/interquartile range and compared by the Kruskall-Wallis test with Dunn Test.

For comparison among proportions Pearson's chi-square test was used in conjunction with residue analysis adjusted.

For all 16S rRNA downstream analysis, the data set was filtered, keeping only OTUs that were present in at least 30% of the samples in each treatment and rarefied to the same number of sequences [23]. The biom file was imported into R environment [24] and estimations of alpha and beta diversity were calculated using the "phyloseq" packge [25], and further plotted using the "ggpltot2" package.

To accesses the main differences among treatments in this study, a Permutational Multivariate Analysis of Variance (PERMANOVA) [26] as performed using a binomial dissimilarity matrix among samples and the *adonis* function implemented in the vegan package [27].

Differences between treatments were accessed through the STAMP software [28]. P-values were obtained by the two sided White's non-parametric t-test followed by the BenjaminiHochberg FDR correction. A p-value < 0.05 together with effect size filter (difference between proportions effect size <1.00) was applied to determine the most important taxa that differed between treatments.

Analysis along the 28-day period was performed using meta-analysis of effect sizes reported at multiple points using general linear mixed mode [29].

Results

A total of 175 samples from 62 preterm newborns divided in five groups (7 in LME, 8 in FLE, 16 in PLM, 16 in PFL, and 15 in MFLM) were collected and analyzed. The five groups were similar in respect to maternal and obstetrical data (Table 1).

Neonatal data are shown in Table 2.

Overall microbial community differences among diets

The microbial community differences at OTU level (97% similarity cutoff for grouping definition) found among fecal samples from preterm newborns' fed with five different diets during a period of 28 days are presented in Fig 1. The ordination analysis revealed significant differences in microbial community structure among treatments suggesting the feeding of different diets was responsible for the assembly of preterm gut community. Those differences were further confirmed by the permutational multivariate analysis of variance (Table 3).

The overall analyses (all samples) indicate that approximately 31% of the variation in distances among treatments was explained by the different diets provided in each treatment. Pairwise comparisons revealed that diets based on maternal milk assembled microbial communities with large variation within group (e.g. greater differences among microbial communities from subjects feed with maternal milk) while diets based on formula created more similar microbial communities (Fig 1A and 1B) (Table 3). The use of maternal milk was responsible for the greatest variation observed between diets. The highest variation between treatments was observed among the samples under exclusive own mother's milk and samples under exclusive formula.

Approximately 37% of the variation in distance between microbial communities could be explained by the treatment with maternal milk only (LME) compared to the diet based preferentially in formula (PFL).

Besides, the greatest variations in distances between treatments were observed in those comparisons involving the use maternal milk only. Approximately 37% of the variation in distance between microbial communities could be explained by the treatment with maternal milk only compared to the diet based preferentially in formula (Table 3).

In agreement with the ordination analysis, alpha diversity measurements indicated significant differences (p-value < 0.001 according to the Kruskal-Wallis test) among the richness of OTUs within treatments (Fig 2). The diet composed by maternal milk only (LME) allowed for greater microbial richness (average of 85 OTUs). On the other hand, the diet preferably based in formula (PFL) presented the smallest richness (average of 9 OTUs) (Fig 2). The average number of OTUs found within the other diets was similar. The diets based in formula only (FLE) and preferably maternal milk (PLM) presented an average of 29 and 23 OTUs respectively and the diet based in a mixture of formula and maternal milk presented an average of 25 OTUs.

Identification of the main taxonomic differences among diets

Once overall differences among microbial communities found in fecal samples from preterm newborns fed with different diets during 28 days were detected, the next step was to identify the microbial taxa responsible for that difference. To detect differences among treatments at genus level, a pairwise differential abundance based on a two-sided White's non-parametric t-test followed by the Benjamini-Hochberg FDR correction was performed. As LME treatment presented the greatest difference in microbial community among diets, the pairwise comparisons were performed between LME and the other diets (Fig 3).

The mean proportion of the genus *Escherichia* was always greater in treatments containing formula (FLE, PLM, MFLM and PFL) than in the treatment with maternal milk only. Particularly, the diet based on maternal milk presented an increased abundance of *Acinetobacter*,

Bradyrhizobium, *Caulobacter*, *Corynebacterium* and *Paenibacillus*, *Burkholderia*, *Faecalibacterium*, *Sphingomonas* and the unknown genus from the *Microbacteriaceae* family as compared with the other treatments. Compared to the diet based on maternal milk, the diet preferably based on maternal milk (PLM) and the diet composed by a mixture of formula and maternal milk increased the abundance of *Clostridium* and *Escherichia*. The fecal samples from newborns fed with a diet preferable based on formula (PFL) presented greater abundance of *Escherichia*, *Salmonella*, *Enterococcus* and the unknown genus from the *Enterobacteriaceae* family when compared to the LME treatment (Fig 3).

Discussion

The variables influencing the composition of the intestinal microbiota are topic of multiple studies. The assumption is that, once they are known and understood, new strategies can be developed to maintain a state of health [30,31,32]. In this study, we found global differences in the microbial community among the types of milk administered to preterm infants, showing that the greatest microbial richness was found in those who were exclusively fed with own mother´s milk. Approximately 37% of the variation in the distance between microbial communities was explained by treatment with breast milk exclusively, in comparison with diets based preferably on formula.

Knowing the total number of bacteria according to type of feeding allow us an understanding of certain situations which facilitate the development of diseases [8,31,32,33]. In our study, fecal samples of premature infants fed different diets were decisive in the diversity of the microbial community during the 28-day period. All infants were premature and were separated according to the type of diet fed. Feeding them own mother´s milk exclusively allowed for greater microbial richness (mean of 85 OTUs). The formula-based group had the lowest richness (mean of 9 OTUs). These diets based on the exclusive offer of formula and, preferably, breast milk, showed an average of 29 and 23 OTUs respectively; and the diet based on a mixture of formula and breast milk presented an average of 25 OTUs.

Gregory et al studied 30 preterm infants (10 in each group) during the first 60 days of life fed with maternal breast milk, pasteurized donor human milk and preterm infant formula. It was found that those fed with maternal breast milk presented higher diversity [8]. Cacho et al showed that, in vitro, incubation of own mother´s milk with donor breast milk in a certain percentage and for 4 hours, may restore the maternal milk microbiome [31]. We did not mix human milk with formula. We offered them individually in different proportions throughout the day. Our data showed that this unmixed administration of breast milk and formula does not determine the restoration of human milk microbioma.

In our study the average proportion of the *Escherichia* and *Clostridium* genus were always higher in treatments containing formula than in treatment which provided breast milk exclusively. This finding has been reported previously [6]. A possible explanation for the results is found in the transmission carried out from the mother to the child and the external environment after the birth, as has already been indicated in other studies [33,34]. Besides that, there are factors like lactoferrin and glycoproteins in human milk that are protective against pathogen bacteria [35, 36].

In spite of several studies on the topic and the subject's relation with diseases of high complexity [1,3,4], such studies were restricted to the enumeration of such diseases based on culture, genetic profile (16S) and the use of small samples, which makes it clear that the variables that shape the intestinal microbiota have not been satisfactorily examined [1,5,6]. It is a known fact that the development of the microbiota in infants depends on medical and dietary factors (1,7), but it is not yet known how such factors influence the general composition of the microbiota, and how those factors cooperate with each other [1]. Studies based on fecal samples of infants and their mothers contribute for monitoring of each chronological and functional stage during the first year of life [1,34].

Besides contributing to the microbial richness, breast milk also favors the prevention of sepsis, necrotizing enterocolitis (NEC) and other diseases [29,31,33,37].

NEC is one of the main causes of morbidity and mortality in neonatal intensive care units, with most cases occurring among premature infants [32,37]. In another study of our group, we provide evidence of an association between NEC and distortions in the normal development of the microbiota and low diversity in NEC cases [32]. Within this study we found that increased diversity and breast feeding correlate with reduced incidence of NEC [32]. Mai et al reported a microbiota that predisposed to late onset sepsis in preterm infants that was closer to that we found in preterm newborns that were not fed exclusively with own mother´s breast milk [38].

Other studies emphasize the influence of diet according to the ethnic and/or geographic characteristics as variables that influence the development of newborn intestinal microbiota [39,40,41]. In our study, these issues were not addressed because they were not the focus of the study, leaving open perspectives for innovative studies.

The microbiome found in breast milk contributes in the short and long term to the prevention of colonization by pathogens, as it stimulates the production of reactive antibodies and establishes a healthy intestinal microbiome capable of preventing long-term morbidities such as obesity, type 2 diabetes, chronic intestinal inflammation, autoimmune disorders, allergies, irritable bowel syndrome and allergic gastroenteritis [31,42,43,44]. Our study confirms that the intestinal microbiota of preterm infants presents differences according to their diet— whether breast milk or formula—and emphasizes the importance of breast milk in the maintenance of microbial richness of the newborn's microbiota. *Escherichia* and *Clostridium* have been associated to NEC as well as a high proportion of *Proteobacteria* with few numbers with *Firmicutes* [45].

Some situations, however, crossed this research, such as the loss of fecal samples; the difficulty of mothers in breastfeeding preterm infants with breast milk exclusively due to socioeconomic difficulties, for example; and the low fecal volume produced by premature infants. Another important point to consider when analyzing the intestinal microbiota of the newborn, we analyze only the stools and we do not know the microbiota of the proximal colon neither of the small bowel. We could not adjust microbial analysis for birth weight and gestational age because the number of subjects in exclusive breast milk group was not big enough. It is very difficult to have extreme premature newborns fed exclusively with breast milk during the whole period in NICU hence we were not allowed to study different variables. Our objective was to compare exclusive own mother´s milk feeding with different proportion of formula feeding. We studied just extreme premature newborns, all of them included in the same category of prematurity, and those small differences in birth weight and gestational age probably do not have major repercussion on microbioma.

Many studies have focused on the issue of breastfeeding, particularly regarding the newborns' microbiota, to identify its benefits for the development and prevention of diseases throughout their lives. Considering the importance of the topic, this study aimed to describe the intestinal microbiota of preterm newborns according to their nutritional habits establishing modifications of the intestinal microbiota according to the type of enteral diet administered.

Based on our data, it is noticed that global differences of the microbial community are found among the types of diets administered to preterm infants, showing that the greatest microbial richness was found in those who received

exclusive own mother's milk in comparison with those that received different proportion of formula.

Acknowledgments

This study was supported by a grant from Bill and Melinda Gates Foundation, CNPQ and DECIT/Ministério da Saúde do Brasil.

Author Contributions

Conceptualization: Rita C. Silveira, Renato S. Procianoy.

Data curation: Renato S. Procianoy.

Formal analysis: Adriana Zanella, Rita C. Silveira, Luiz F. W. Roesch, Priscila T. Dobbler, Volker Mai, Renato S. Procianoy.

Funding acquisition: Renato S. Procianoy.

Investigation: Andréa L. Corso, Volker Mai, Renato S. Procianoy.

Methodology: Rita C. Silveira, Luiz F. W. Roesch, Priscila T. Dobbler, Volker Mai, Renato S. Procianoy.

Project administration: Andréa L. Corso, Renato S. Procianoy.

Resources: Adriana Zanella, Luiz F. W. Roesch, Priscila T. Dobbler, Renato S. Procianoy.

Software: Luiz F. W. Roesch, Priscila T. Dobbler.

Supervision: Rita C. Silveira, Andréa L. Corso, Renato S. Procianoy.

Validation: Luiz F. W. Roesch, Renato S. Procianoy.

Visualization: Renato S. Procianoy.

Writing – original draft: Adriana Zanella, Rita C. Silveira, Luiz F. W. Roesch, Andréa L. Corso, Volker Mai, Renato S. Procianoy.

Writing – review & editing: Adriana Zanella, Rita C. Silveira, Luiz F. W. Roesch, Andréa L.

Corso, Volker Mai, Renato S. Procianoy.

References

1. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015; 17: 690–703. https://doi.org/10.1016/j.chom.2015.04.004 PMID: 25974306
2. Cabreiro F, Gems D. Worms need microbes too: microbiota, health and aging in Caenorhabditis elegans. EMBO Mol Med. 2013; 5:1300–1310. https://doi.org/10.1002/emmm.201100972 PMID: 23913848
3. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012; 486:207–214. https://doi.org/10.1038/nature11234 PMID: 22699609

4. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et.al. MetaHIT Consortium. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol*. 2014; 32: 834–841. https://doi.org/10.1038/nbt.2942 PMID: 24997786

5. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. R. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA*. 2010; 107: 11971–11975. https://doi.org/10.1073/pnas.1002601107 PMID: 20566857

6. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam M, et al. A Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature*. 2014; 510: 417–421. https://doi.org/10.1038/nature13421 PMID: 24896187

7. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore, et al. Patterned progression of bacterial populations in the premature infant gut. *Proc*. *Natl*. *Acad*. *Sci*. *USA*. 2014; 111: 12522–12527. https://doi.org/10.1073/pnas.1409497111 PMID: 25114261

8. Gregory KE, Samuel BS, Houghteling P, Shan G, Ausubel FM, Sadreyev RI, et al. Influence of maternal breast milk ingestion on acquisition of the intestinal microbiome in preterm infants. *Microbiome*. 2016; 4:68. https://doi.org/10.1186/s40168-016-0214-x PMID: 28034306

9. Houghteling P, Walker WA. Why is initial bacterial colonization of the intestine important to the infant's and child's health? *J Pediatr Gastroenterol Nutr*. 2015; 60: 294–307. https://doi.org/10.1097/MPG.0000000000000597 PMID: 25313849

10. Zeissig S, Blumberg RS. Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease. *Nat Immunol*. 2014; 15: 307–10. https://doi.org/10.1038/ni.2847 PMID: 24646587

11. Brooks B, Firek BA, Miller CS, Sharon I, Thomas BC, Baker R, et al. Microbes in the neonatal intensive care unit resemble those found in the gut of premature infants. *Microbiome*. 2014; 2: 1. https://doi.org/10.1186/2049-2618-2-1 PMID: 24468033

12. Ardeshir A, Narayan NR, Me´ndez-Lagares G, Lu D, Rauch M, Huang Y, et al. Breast-fed and bottle-fed infant rhesus macaques develop distinct gutmicrobiotas and immune systems. *Sci Transl Med*. 2014;6: 252ra120. https://doi.org/10.1126/scitranslmed.3008791 PMID: 25186175

13. Boix-Amoro´s A, Collado MC, Mira A. Relationship between Milk Microbiota, Bacterial Load, Macronutrients, and Human Cells during Lactation. *Front Microbiol*. 2016; 7: 492 https://doi.org/10.3389/fmicb. 2016.00492 PMID: 27148183

14. Khodayar-Pardo P, Mira-Pascual L, Collado MC, Martı´nez-Costa C. Impacto flactati on stage, gestation alage and mode of delivery on breast milk microbiota. *J.Perinatol*.2014; 34: 599–605. https://doi.org/10.1038/jp.2014.47 PMID: 24674981

15. Cabrera-Rubio R, Mira-Pascual L, Mira A, Collado MC. Impact of mode of delivery on the milk microbiota composition of healthy women. *J*. *Dev*. *Orig*. *Health Dis*. 2016; 7: 54–60. https://doi.org/10.1017/ S2040174415001397 PMID: 26286040

16. Roesch LF, Silveira RC, Corso AL, Dobbler PT, Mai V, Rojas BS, et al. Diversity and composition of vaginal microbiota of pregnant women at risk for transmitting Group B Streptococcus treated with intrapartum penicillin. PLoS One. 2017; 12(2):e0169916. https://doi.org/10.1371/journal.pone.0169916 PMID: 28178310

17. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012; 6: 1621–1624. https://doi.org/10.1038/ismej.2012.8 PMID: 22402401

18. Pylro VS, Roesch LF, Morais DK, Clark IM, Hirsch PR, To´tola MR. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *J. Microbiol*. *Methods*. 2014; 107: 30–37. https://doi.org/10.1016/j.mimet.2014.08.018 PMID: 25193439

19. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat*. *Methods*. 2013; 10: 996–998., https://doi.org/10.1038/nmeth.2604 PMID: 23955772

20. Edgar R. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. 2016. https://doi.org/10.1101/074161

21. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, D633–D642. 2014. https://doi.org/10.1093/nar/gkt1244 PMID: 24288368

22. Good IJ. The population frequencies of species and the estimation of population parameters. *Biometrika*. 1953; 40: 237–264

23. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW. Rethinking microbial diversity analysis in the high throughput sequencing era. J. *Microbiol*. *Methods*. 2011; 86: 42–51.

24. R Development Core Team. R: A Language and Environment for Statistical Computing. 2008. Available at: http://www.R-project.org

25. McMurdie PJ, Holmes S. phyloseq: an R Package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE*. 2013; 8:e61217. https://doi.org/10.1371/journal.pone.0061217 PMID: 23630581

26. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol*. 2001; 26: 32–46

27. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. Vegan: Community Ecology Package. 2015. R package vegan, version. 2.2–1.

28. Parks DH, Tyson GW, Hugenoltz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014; 30: 3123–3124. https://doi.org/10.1093/bioinformatics/btu494 PMID: 25061070

29. Musekiwa A, Manda SO, Mwambi HG, Chen DG. Meta-Analysis of Effect Sizes Reported at Multiple Time Points Using General Linear Mixed Model. PLoS One. 2016; 11(10):e0164898 https://doi.org/10.1371/journal.pone.0164898 PMID: 27798661

30. Yasmin F, Tun HM, Konya TB, Guttman DS, Chari RS, Field C J, et al. Cesarean Section, Formula Feeding, and Infant Antibiotic Exposure: Separate and Combined Impacts on Gut Microbial Changes in Later Infancy. *Front Pediatr*. 2017; 5:200. https://doi.org/10.3389/fped.2017.00200 PMID: 29018787

31. Cacho NT, Harrison NA, Parker LA, Padgett K A, Lemas DJ, Marcial GE, et al. Personalization of the Microbiota of Donor Human Milk with Mother's Own Milk. *Front Microbiol*. 2017; 8: 1470. https://doi.org/10.3389/fmicb.2017.01470 PMID: 28824595

32. Dobbler PT, Procianoy RS, Mai V, Silveira RC, Corso AL, Rojas BS, et al. F. W. Low Microbial Diversity and Abnormal Microbial Succession Is Associated with Necrotizing Enterocolitis in Preterm Infants. *Front Microbiol*. 2017; 8: 2243. https://doi.org/10.3389/fmicb.2017.02243 PMID: 29187842

33. Gritz EC, Bhandari V. The human neonatal gut microbiome: a brief review. *Front*. *Pediatr*. 2015; 3: 17. https://doi.org/10.3389/fped.2015.00017 PMID: 25798435

34. Murphy K, Curley D, O'Callaghan TF, O'Shea C-A, Dempsey EM, O'Toole PW, et al. The Composition of Human Milk and Infant Faecal Microbiota Over the First Three Months of Life: A Pilot Study. *Sci Rep*. 2017; 7: 40597. https://doi.org/10.1038/srep40597 PMID: 28094284

35. Liu B, Newburg DS. Human milk glycoproteins protect infants against human pathogens. Breastfeed Med. 2013; 8:354–62. https://doi.org/10.1089/bfm.2013.0016 PMID: 23697737

36. Lees EA, Miyajima F, Pirmohamed M, Carrol ED. The role of Clostridium difficile in the paediatric and neonatal gut—a narrative review. Eur J Clin Microbiol Infect Dis. 2016; 35:1047–57. https://doi.org/10.1007/s10096-016-2639-3 PMID: 27107991

37. Pammi M, Cope J, Tarr PI, Warner BB, Morrow AL, Mai V, et al. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome*. 2017; 5:31. https://doi.org/10.1186/s40168-017-0248-8 PMID: 28274256

38. Mai V, Torrazza RM, Ukhanova M, Wang X, Sun Y, Li N, et al. Distortion in the development of intestinal microbiota associated with late onset sepsis in preterm infants. Plos One 2013; 8(1):e52876 https://doi. org/10.1371/journal.pone.0052876 PMID: 23341915

39. Stearns JC, Zulyniak MA, de Souza RJ, Campbell NC, Fontes M, Shaikh M, et al. Ethnic and dietrelated differences in the healthy infant microbiome. *Genome Med*. 2017; 9: 32. https://doi.org/10.1186/s13073-017-0421-5 PMID: 28356137

40. Fallani M, Amarri S, Uusijarvi A, Adam R, Khanna S, Aguilera M, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology*. 2011; 157: 1385–92.2. https://doi.org/10.1099/mic.0.042143-0 PMID: 21330436

41. Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J*. *Pediatr*. *Gastroenterol*. *Nutr*. 2010; 51: 77–84. https://doi.org/10.1097/MPG.0b013e3181d1b11e PMID: 20479681

42. Goulet O. Potential role of the intestinal microbiota in programming health and disease. *Nutr*. *Rev*. 2015; 73: 32–40. https://doi.org/10.1093/nutrit/nuv039 PMID: 26175488

43. Wallace JG, Gohir W, Sloboda DM. The impact of early life gut colonization on metabolic and obesogenic outcomes: what have animal models shown us? *J*. *Dev*. *Orig*. *Health Dis*. 2016; 7: 15–24. https:// doi.org/10.1017/S2040174415001518 PMID: 26399435

44. Lemas DJ, Yee S, Cacho N, Miller D, Cardel M, Gurka M, et al. Exploring the contribution of maternal antibiotics and breastfeeding to development of the infant microbiome and pediatric obesity. Semin. *Fetal Neonatal Med*. 2016; 21: 406–409. https://doi.org/10.1016/j.siny.2016.04.013 PMID: 27424917

45. Harts LE, Bradshaw W, Brandon DH. Potential NICU environmental influences on the neonate´s microbioma: a systematic review. *Adv Neonatal Care*. 2015; 15: 324–335 https://doi.org/10.1097/ANC.0000000000000220 PMID: 26340035

Figures

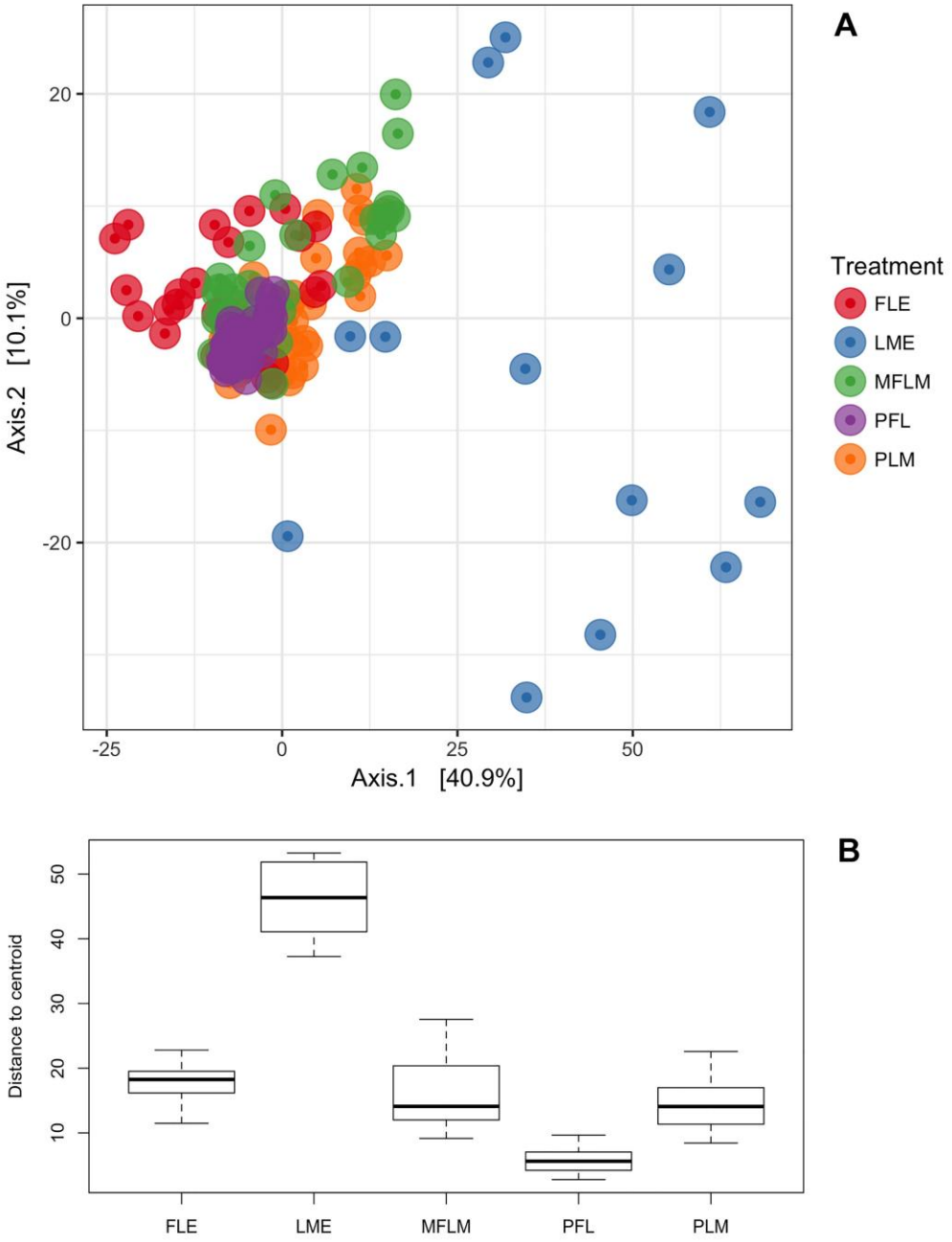


Fig 1. Beta diversity comparisons among microbial communities. A and B. Graph A represents clusters of microbial communities. Each point represents an individual sample, with colors indicating feeding treatments. Graph B represents measurement of multivariate dispersion for each treatment. FLE = exclusive formula; LME = exclusive own mother´s milk; MFLM = 50% formula and 50% own mother's milk; PFL = 70% formula; PLM = 70% own mother's milk.

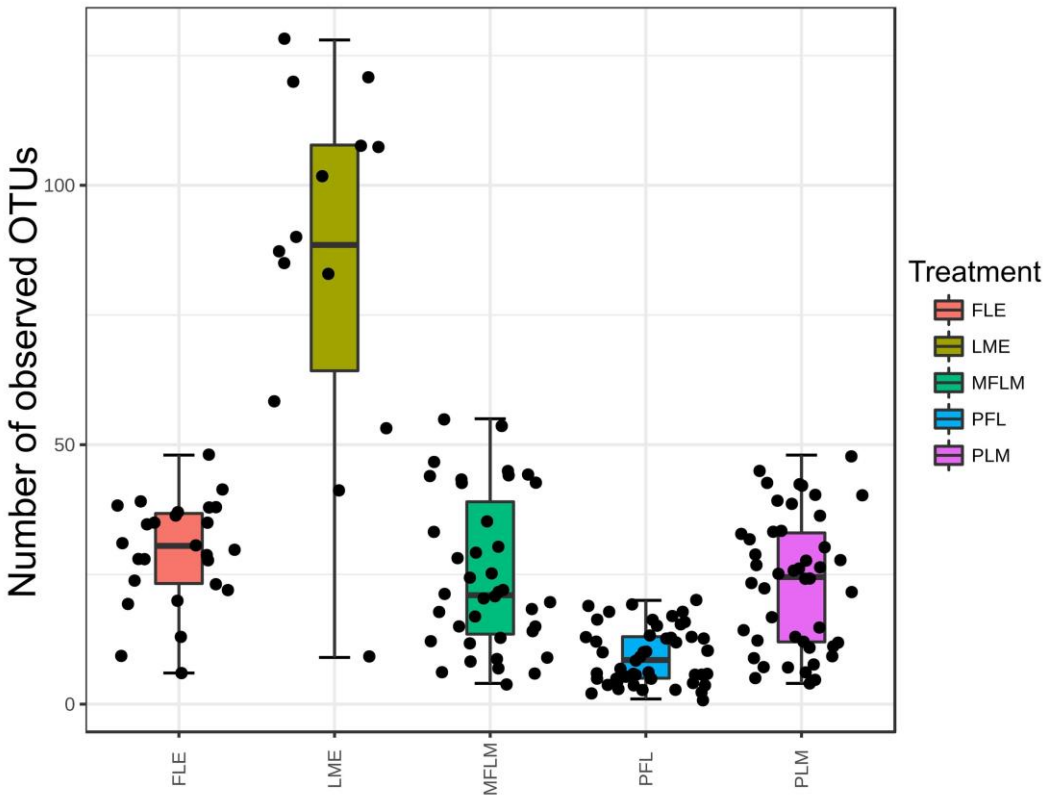


Fig 2. Number of Operational Taxonomic Units measured in fecal samples from preterm babies fed with different diets during 28 days. Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles. Treatments were significant different according to the Kruskal-Wallis test (p-value < 0.001), with greatest diversity in human milk exclusive group.
FLE = exclusive formula; LME = exclusive own mother's milk; MFLM = 50%formula and 50% own mother's milk; PFL = 70% formula; PLM = 70% own mother's milk. All libraries were rarefied at the same number of sequences prior to OTUs analysis.

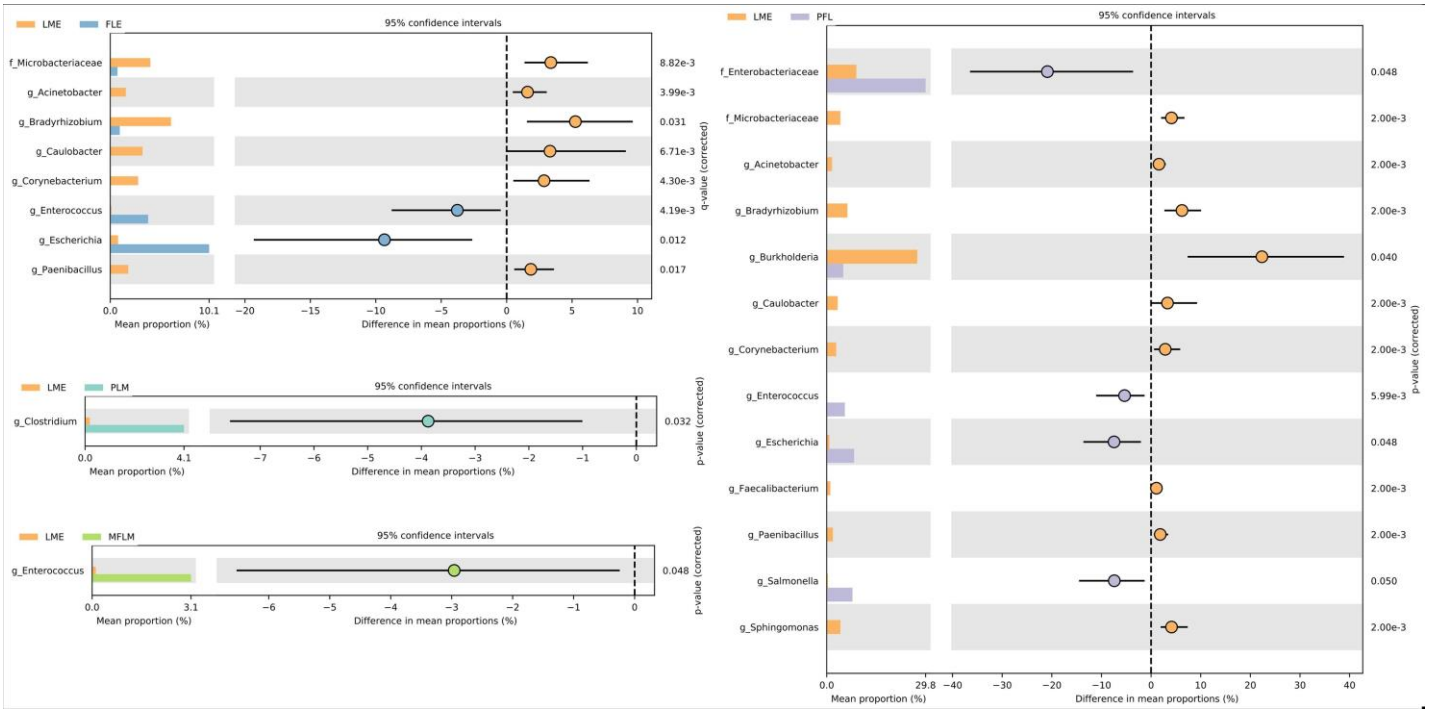


Fig 3. Differential microbial abundance among microbial communities detected in samples from preterm babies fed with different diets during 28 days. p-values were obtained by the two-sided White's non-parametric t-test followed by the Benjamini-Hochberg FDR correction. A p-value of < 0.05 together with effect size filter (difference between proportions effect size <1.00) was applied to determine the most important taxa that differed between treatments. Only statistically significant differences are shown.

Tables

Table 1. Characteristics of the subjects.

Variables	Exclusive breast milk (n = 7)	Exclusive Formula (n = 8)	Predominance of breast milk (n = 16)	Predominance of formula (n = 16)	Mixed (n = 15)	p-value
Maternal age (years)—mean ± SD	24.0 ± 9.1	26.1 ± 6.4	31.6 ± 5.5	24.6 ± 7.4	28.0 ± 7.5	0.052
C-section–n(%)	7 (100)	6 (75)	13 (81.3)	11 (68.8)	12 (80)	0.556
Rupture of membranes (hours)–median (P25-P75)	0 (0–3)	0 (0–0.4)	0.04 (0–37)	6.1 (0–96.5)	0 (0–3)	0.246
Rupture of membranes 18 hours–n(%)	0 (0.0)	1 (12.5)	4 (25)	7 (43.8)	2 (13.3)	0.062
Maternal antibiotic–n(%)	3 (42.9)	5 (62.5)	11 (68.8)	14 (87.5)	8 (53.3)	0.188
Preeclampsia–n(%)						0.119
Yes	5 (71.4)	1 (12.5)	7 (43.8)	6 (37.5)	4 (26.7)	
No	2 (28.6)	6 (75)	9 (56.3)	10 (62.5)	11 (73.0)	
Eclampsia	0 (0.0)	1 (12.5)	0 (0.0)	0 (0.0)	0 (0.0)	
Streptococcus–n(%)	1 (14.3)	2 (25)	4 (25)	5 (31.3)	6 (40)	0.944

Statistically significant association by the test of the residuals adjusted to 5% of significance

The mean, the median standard deviation and the interquartile range were used, using Analysis of Variance (ANOVA) in conjunction with the Tukey test was applied. In case of asymmetry, the Kruskall-Wallis test in set with Dunn were used. Categorical variables were described by absolute and relative frequencies. In the comparison of proportions, Pearson's chi-square test was used in conjunction with residue analysis adjusted

Table 2. Neonatal data according to study group.

Variables	Exclusive breast milk (n = 7)	Exclusive Formula (n = 8)	Predominance of breast milk (n = 16)	Predominance of formula (n = 16)	Mixed (n = 15)	p-value
Birth Weight (g)—mean ± SD	912.1 ± 291.5[a]	1684 ± 430.1[b]	1460 ± 575.4[ab]	1459 ± 413.3[ab]	1332 ± 372[ab]	0.021
Gestational age (weeks)—mean ± SD	27.7 ± 2.7[a]	30.6 ± 1.7[b]	29.9 ± 2.4[ab]	30.5 ± 1.6[b]	29.4 ± 1.7[ab]	0.031
NEC-n (%)	1 (14.3)	0 (0.0)	1 (6.3)	1 (6.3)	2 (13.3)	0.778
Use of antenatal corticosteroid—n (%)	6 (85.7)	7 (87.5)	16 (100)	15 (93.8)	15 (100)	0.185

Early onset sepsis—n (%)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (6.7)	0.527
Use of Antibiotics on the first week—n (%)	6 (85.7)	3 (37.5)	12 (75)	13 (81.3)	8 (53.3)	0.110
Late onset sepsis—n (%)	2 (28.6)	3 (37.5)	7 (43.8)	5 (31.3)	6 (40.0)	0.937
Use of Antibiotics on the 2nd week—n (%)	7 (100)	5 (62.5)	10 (62.5)	10 (62.5)	41 (66.1)	0.396

Statistically significant association by the test of the residuals adjusted to 5% of significance

For the Weight at birth and Gestational age the mean and standard deviation were used. We compare the averages through Analysis of Variance (ANOVA) in conjunction with the Tukey test. In case of asymmetry the Kruskall-Wallis test in conjunction with Dunn were used.

Categorical variables were described by absolute and relative frequencies. In the comparison of proportions, Pearson's chi-square test was used in conjunction with residue analysis adjusted.

Table 3. Permutational multivariate analysis of variance comparing microbial communities found in fecal samples from preterm newborns fed with different diets during 28 days.

	F	R2	Adjusted p-value
All samples			
Diets	19.24	0.311	0.001
Residuals		0.688	
Total		1.000	
Pairwise comparisons			
PLM *vs* FLE	14.17	0.168	0.01
PLM *vs* LME	21.89	0.274	0.01
PLM *vs* MFLM	7.28	0.080	0.01
PLM *vs* PFL	16.14	0.146	0.01
FLE *vs* LME	18.85	0.331	0.01
FLE *vs* MFLM	10.82	0.146	0.01
FLE *vs* PFL	12.92	0.148	0.01
LME *vs* MFLM	18.88	0.270	0.01
LME *vs* PFL	36.76	0.372	0.01
MFLM *vs* PFL	14.78	0.145	0.01

F = F value by permutation. R^2 = shows the percentage of variation explained by diets; p-values were based on 999 permutations and were adjusted by Bonferroni correction.

FLE = exclusive formula; LME = exclusive own mother´s milk; MFLM = 50% formula and 50% own mother´s milk; PFL = 70% formula; PLM = 70% own mother´s milk.

APÊNDICE C – Microbiota de mecônio prediz sepse neonatal precoce clínica em neonatos prematuros

Meconium microbiota predicts clinical early-onset neonatal sepsis in preterm neonates

Laura V. Dornelles[a], Renato S. Procianoy[a], Luiz F. W. Roesch[b], Andrea L. Corso[a,b], Priscila Thiago Dobbler[b], Volker Mai[c] and Rita C. Silveira[a]

[a] Serviço de Neonatologia do Hospital de Clínicas de Porto Alegre, Programa de Pos Graduação em Saúde da Criança e do Adolescente, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil;
[b] Centro Interdisciplinar de Pesquisas em Biotecnologia – CIP-Biotec, Campus São Gabriel, Universidade Federal do Pampa, São Gabriel, Brazil;
[c] Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

Correspondent Author: Rita C. Silveira drarita.c.s@gmail.com

DOI: https://doi.org/10.1080/14767058.2020.1774870
Received 2 November 2018
Revised 26 March 2020
Accepted 24 May 2020

Keywords: Prematurity; early-onset neonatal sepsis; intestinal microbiota; intestinal microbiome; meconium

ABSTRACT

Background: Early-onset neonatal sepsis (EONS) remains one of the leading causes of morbidity and mortality related to premature birth, and its diagnosis remains difficult. Our goal was to evaluate the intestinal microbiota of the first meconium of preterm newborns and ascertain whether it is associated with clinical EONS.
Methods: In a controlled, prospective cohort study, samples of the first meconium of premature infants with a gestational age (GA) ≤ 32 weeks was obtained at Hospital de Clínicas de Porto Alegre and DNA was isolated from the samples. 16S rDNA based microbiota composition of preterm infants with a clinical diagnosis of EONS was compared to that of a control group.
Results: 40 (48%) premature infants with clinical diagnosis of EONS and 44 (52%) without EONS were included in the analysis. The most abundant phylum detected in both groups, *Proteobacteria*, was more prevalent in the sepsis group (p = .034). 14% of variance among bacterial communities (p = .001) correlated with EONS. The genera most strongly associated with EONS were *Paenibacillus, Caulobacter, Dialister, Akkermansia, Phenylobacterium, Propionibacterium, Ruminococcus, Bradyrhizobium*, and *Alloprevotella*. A single genus, *Flavobacterium*, was most strongly associated with the control group.
Conclusion: These findings suggest that the first-meconium microbiota is different in preterm neonates with and without clinical EONS.

Introduction

Early-onset neonatal sepsis (EONS) remains one of the leading causes of morbidity and mortality related to premature birth, and its timely diagnosis remains difficult [1,2]. Currently available laboratory tests (complete blood count, C-reactive protein, procalcitonin, cytokines) have a low positive predictive value (PPV), making it difficult to identify which neonates are infected and leading to excessive use of antibiotics (ABx) in the first week of life. Blood cultures remain the gold standard for diagnosis of early- and late onset neonatal sepsis, but its sensitivity is low – 80% at best [2–4]. Due largely to these challenges, the presumptive or clinical symptom based diagnosis of sepsis (i.e. in patients with risk factors and clinical signs of sepsis, but without blood-culture confirmation) is still very common in neonatal intensive care units (NICUs) [5].

The meconium is not sterile [6,7]. Evidence suggests that the meconial microbiota reflects the intrauterine microbial community [8]. Previous studies have shown that the meconial microbiota is less diverse in neonates that will develop infection [7], suggesting that lack of colonization by several nonpathogenic bacteria facilitates colonization by pathogenic bacteria and may thus increase the risk of infection and bacterial translocation [9,10].

Some studies have shown that an increase in Proteobacteria precedes the diagnosis of late-onset sepsis [9] and necrotizing enterocolitis (NEC) [11]. *Citrobacter koseri* and *Klebsiella pneumoniae*, both belonging to the phylum *Proteobacteria*, have been shown to be dominant in the microbiota of newborns that developed NEC [12]. These microbial signatures might have potential as biomarkers for early diagnosis of this disease. In the same study, NEC cases had a less diverse microbiota, with a lower abundance of *Lactobacillus*, and more interconnected and clustered species compared to controls, which suggests a more closely related community [12].

Mai et al. showed that the fecal microbiota diversity of preterm infants increases with time, and that the intestinal bacterial community in NEC cases is different from

that of controls [11]. Establishing causality of dysbiosis in NEC and EONS may provide novel approaches for manipulating the microbiome of ill preterm infants to resemble that of healthy, full-term neonates and reduce disease risk [13].

Several studies have demonstrated the difficulty of reliably diagnosing EONS, especially in preterm infants. Therefore, a better understanding of the microbial colonization of the gastrointestinal tract of newborns in the immediate postnatal period and its possible association with EONS is essential. Within this context, the present study was designed to evaluate the intestinal microbiota of the first meconium of preterm newborns with gestational age (GA) ≤ 32 weeks and ascertain whether it is associated with clinical early-onset neonatal sepsis. A better understanding of microbiome variation may allow for early detection in patients who are at greater disease risk.

Methods

This cohort study was conducted in a convenience sample of preterm neonates with a gestational age less than or equal to 32 weeks born at Hospital de Clínicas de Porto Alegre (HCPA), Brazil. The exclusion criteria were: congenital malformations or infections; genetic syndromes; HIV-positive mothers; and refusal by parents or legal guardians. After the mother or legal guardian provided written informed consent, the first meconium passed by the infant was collected, stored at 80 C in a cryogenic storage Dewar, and transported to a laboratory at Universidade Federal do Pampa (UNIPAMPA), where DNA extraction and microbial community composition analysis was performed.

The maternal variables of interest were: Mode of delivery (vaginal or cesarean section), time of prolonged rupture of membranes (PROM), urinary tract infection (confirmed by urine cultures), chorioamnionitis (pathological diagnosis), pre-eclampsia (systolic blood pressure 140 mmHg and/or diastolic blood pressure 90 mmHg after GA 20 weeks, with proteinuria >300 mg in a 24-h urine sample, in the absence of preexisting hypertension or renal disease) [14], and antibiotic use. The neonatal variables of interest were: GA (based on obstetric ultrasound performed during the first 12 weeks of pregnancy or, in the absence of such information, on neonatal examination) [15], birth weight, sex, 5-min Apgar score, adequacy for gestational age [16], presence of the respiratory distress syndrome (RDS), bronchopulmonary dysplasia (BPD) (need of supplemental oxygen at 28 days of life) [17], late-onset neonatal sepsis (positive blood cultures after the 72nd hour of life in the presence of clinical signs of infection), NEC (presence of pneumatosis intestinalis and/or pneumoperitoneum and clinical signs), grades 2, 3 or 4 periventricular–intraventricular hemorrhage (PIVH) [18], periventricular leukomalacia (PVL), persistent ductus arteriosus (PDA), discharge, or death.

The preterm infants included were divided into two groups according to the presence or absence of clinical early-onset neonatal sepsis (sepsis group and control group, respectively). We defined clinical EONS when three or more of the following categories were present: 1-maternal risk factors: fever, prolonged rupture of membranes (PROM > 18 h), preterm labor and chorioamnionitis. 2-Clinical signs: apnea, tachypnea, nasal flaring, chest retractions, cyanosis, respiratory effort distress; tachycardia or bradycardia, and poor perfusion or shock; irritability, lethargy, hypotonia,

seizures; abdominal distension, vomiting, diarrhea, feeding intolerance, hepatomegaly; unexplained jaundice, petechiae or purpura, as well as infant “not looking well.” 3-Laboratorial findings: WBC > 25,000 and C-Reactive Protein (CRP > 10 mg/L). This study was approved by the HCPA Research Ethics Committee.

Microbial DNA extraction, amplification, and sequencing

Microbial DNA was isolated from meconium samples using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Valencia, CA, USA), in accordance with manufacturer instructions. DNA quality was quantified by spectrophotometry in a NanoVue™ system (GE Healthcare, Chicago, IL, USA). All DNA samples were stored at 80 C until use. The V4 region of the 16S rRNA gene was amplified and sequenced using the ION PGM™ Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA), with primers 515 F and 806 R [19]. Multiple samples were amplified by polymerase chain reaction (PCR) using barcoded primers linked with the adapter “A” sequence (5^0-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3^0) and “P1” sequence (5^0-CCTCTCTATGG GCAGTCGGTGAT-3^0) to obtain a primer sequence composed for the A-barcode-806R and P1-515F adapter and primers. The PCR reaction volume was 25 lL. Each mix consisted of 2 U Platinum™ Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4 lL 10X High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.1 lM of both primers described above, 25 lg UltraPure BSA (Invitrogen, Carlsbad, CA, USA), and approximately 50 ng of template DNA.

The PCR conditions used were: 95 C for 5 min, 35 cycles at 94 C for 45 s; 56 C for 45 s and 72 C for 1 min, followed by 72 C for 10 min. The resulting PCR products were purified with the Agencourt™ AMPureVR XP Reagent (Beckman Coulter, La Brea, CA, USA) and the final concentration of the PCR product was quantified using the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA), following manufacturer recommendations.

Finally, the reactions were combined in equimolar concentrations to create a mixture composed of amplified fragments of the 16S gene from each sample. This composite sample was used for library preparation with the OneTouch™ 2 Ion system using the ION™ PGM Template 400 OT2 kit (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed using the commercially available ION PGM™ Sequencing 400 kit on an ION PGM™ System, using an Ion 318™ Chip v2, with a maximum of 40 samples per microchip.

Sequence processing for analysis

Fastq files exported from the ION PGM™ system were analyzed following the recommendations of the Brazilian Microbiome Project (BMP) [20], using the BMP Operating System [21]. Briefly, an Operational Taxonomic Unit (OTU) table was compiled using the UPARSE pipeline [22], wherein sequences were truncated at 200 base pairs and quality-filtered using a maximum expected error cutoff of 0.5. Sequences were clustered into OTUs using a 97% similarity cutoff, and chimeric sequences were removed [22].

Taxonomic classification was performed in the QIIME software environment [23], based on the UCLUST method, against the Greengenes 13.5 database [24], with a confidence limit of 80%. The sampling effort was estimated using Good’s coverage formula [25].

Statistical analysis

The data obtained in the study were stored in a Microsoft Excel database compiled for this specific purpose. Data were then processed and analyzed in PASW Statistics for Windows, Version 18.0 (SPSS Inc., Chicago, IL, USA). The results were described as mean and standard deviation (SD) for symmetrically distributed data or as median and interquartile range (IQR) otherwise. Qualitative variables were expressed as absolute and relative frequencies. To compare quantitative variables in relation to the presence or absence of sepsis, Student's t-test (for symmetrically distributed data) or the Mann–Whitney U test (for asymmetric data) were applied. For qualitative variables, the chisquare test was applied. Nonparametric Permutational Multivariate Analysis of Variance (PERMANOVA) was used to adjust to Prolonged Rupture of Membranes (PROM) > 18 h. The significance level for all analyses was established as alpha = 0.05.

All 16S rRNA gene libraries (samples) were normalized by random resampling of sequence data, as recommended by Lemos et al. [26]. The BIOM file containing OTUs clustered at a similarity level of 97% was imported into the R environment (R Development Core Team, 2008), and a compositional dissimilarity matrix was generated based on binomial distances between samples, using the "phyloseq" package [27].

To assess structural differences in microbial community, a dissimilarity matrix between samples using the binomial distance was calculated using principal coordinate analysis (PCoA). The matrix was then used in a nonparametric permutational multivariate analysis of variance (PERMANOVA), with the Adonis function available in the "vegan" package [28], to detect confounding variables. To estimate alpha diversity, the number of OTUs and the Shannon diversity index were calculated and plotted using the "phyloseq" package [27]. These were described as medians and interquartile ranges, and the Mann–Whitney test was used for statistical analysis.

The STAMP v2 statistical package was used to determine differences in relative abundance of microbial phyla between groups. Between-group differences were assessed using the nonparametric White t-test, and confidence intervals were calculated using the bootstrap method. Taxonomic units with a difference between proportions of <1% were excluded from analysis. The linear discriminant analysis (LDA) score was calculated to demonstrate and explain the differences in microbial community between the sepsis and control groups.

Results

One hundred and sixty-eight patients were eligible, provided written informed consent, and were included in the study. Of these, 84 met all of the inclusion and none of the exclusion criteria and provided meconium samples of sufficient quantity and quality to allow the proposed analyses. One meconium sample was obtained from each preterm infant. These 84 patients were divided into two groups: 40 (48%) with a clinical diagnosis of sepsis made up the sepsis group, while 44 (52%) without a diagnosis of sepsis made up the control group.

The characteristics of the studied patients are presented in Table 1, divided into maternal (1 A) and neonatal (1B) variables. The mean birth weight was 1395 g (± 496) in the sepsis group and 1428 g (± 497) in the control group. The mean gestational age was 30 (± 2.4) weeks in the sepsis group and 30.5 (± 2.2) weeks in the control group. There was no difference between the two groups when comparing these characteristics ($p = .69$ and $p = .29$, respectively).

There was no significant difference between groups regarding the maternal and neonatal characteristics, except for time of PROM and RDS. Overall, 15 patients from the sepsis group (37.5%) and five patients from the control group (11.4%) had PROM > 18 h ($p = .03$, Table 1(A)). Twenty-one patients from the sepsis group (52.5%) and 11 from the control group (25%) developed RDS ($p = .03$, Table 1(B)).

After filtering 16S rRNA, a total of 1,920,679 high quality sequences longer than 200 base pairs were obtained. The sampling effort ranged from 87% to 99% [25], indicating that the dataset was representative of the analyzed microbial communities.

Ordination and clustering of microbial communities found in meconium samples of preterm infants with and without sepsis (Figure 1) showed a clear difference in the structure of these communities between the sepsis and no-sepsis groups irrespective of PROM > 18 h. The PERMANOVA statistics controlling for PROM > 18 h, indicated that 14% of the variation among the communities was explained by the presence of clinically diagnosed neonatal sepsis ($p = .001$). To detect possible confounding variables, PERMANOVA was carried out. The results show that none of the variables tested, whether alone or in combination, influenced the meconium microbiota, for example, ABx use by the mother during labor and ABx therapy in the neonate at the time of meconium collection ($p = .06$ and $p = .06$, respectively). The birth weight variable had a statistically significant p-value on this analysis ($p = .022$), but the R^2 value was too low (0.024) for this variable to be considered influential on the meconium microbiota.

Figure 2 shows differences in alpha diversity between the groups with and without a clinical diagnosis of sepsis. The number of OTUs in the control group ranged from 9 to 110. The median was 63.4. The number of OTUs in the sepsis group ranged from 7 to 127. The median was 80.4. The nonparametric Mann–Whitney test revealed a significant difference between the two groups ($p = .005$); however, the variation in number of OTUs observed per patient was large. The mean Shannon diversity index was 2.22 for

the control group and 2.62 for the sepsis group (Figure 2). The nonparametric Mann–Whitney test did not indicate a significant difference between the two groups ($p = .07$).

As there was a statistically significant difference in microbial community structure between the two groups, we used LDA scores to demonstrate and explain this difference between the sepsis and control groups, based on the different microbial genera present in the samples. The results are shown in Table 2. A total of nine genera were more associated with future development of sepsis ($p \leq .05$; FDR ≤ 0.05). In descending order of LDA score, these genera were: *Paenibacillus*, *Caulobacter*, *Dialister*, *Akkermansia*, *Phenylobacterium*, *Propionibacterium*, *Ruminococcus*, *Bradyrhizobium*, and *Alloprevotella*. On the other hand, a single genus was more associated with absence of a clinical diagnosis of sepsis: *Flavobacterium*.

The most abundant phylum in the meconium of preterm neonates was *Proteobacteria*. On comparison of the two groups, we found that this phylum was more prevalent in the sepsis group ($p = .034$) (Figure 3).

Discussion

Our study shows that the intestinal microbiota of preterm infants with a clinical diagnosis of early-onset sepsis is different from that of preterm infants without infection. Overall, 14% of this difference is explained by the presence of clinical EONS. ABx use by the mother during labor and ABx therapy in the neonate at the time of meconium collection did not interfere with the difference in microbiome observed between the two groups. The ABx use by the mother may have prevented sepsis in some cases, and therefore this data was not significant in the difference of the microbiota between the two groups.

Birth weight explained about 2% of the difference in bacterial community between the two groups, which can be considered a negligible contribution in clinical practice. When comparing alpha diversity, we found a significant difference between the two groups, but the variation in the number of OTUs observed per patient in each group was large, allowing us to infer that this measure of diversity may not be a robust indicator of microbial differences between the tested groups.

It is now known that a synergistic interaction exists between the human body and the intestinal microbiota, wherein a balance must be maintained for the proper functioning of several roles, including establishment of the immune response. This immune response may be hematological or local, to strengthen the intestinal mucosal barrier [9,13,29]. In healthy and full term neonates, the gut and immune system regulate the microbial community to ensure proper functioning [30]. However, this balance can be quickly disrupted, especially in situations common in prematurity, such as preterm rupture of membranes, chorioamnionitis, ABx use by the mother, delayed initiation of enteral feeding, infection, and hemodynamic instability [13]. This disruption is known as dysbiosis. Until the present study, the role of dysbiosis in early-onset neonatal sepsis had been poorly described.

Establishing a definitive diagnosis of early-onset neonatal sepsis in preterm infants is challenging. Blood cultures, the gold standard for diagnosis of neonatal sepsis, require that 1–2 ml of blood be drawn to detect microorganisms, as colony counts in the neonatal period are low [31]. In a patient weighing 1000 g, this volume may represent 2–3% of the total blood volume. This reinforces the need to understand whether other indicators, such as a pattern of dysbiosis in the meconial flora, can predict sepsis. Within this context, we found that nine microbial genera were more associated with development of clinical EONS in preterm infants.

Studies show that the microbiome present in the first meconium most likely originates from the amniotic fluid (AF) [8]. Puri et al. compared stool samples from newborns with and without chorioamnionitis/funisitis at weeks 1, 2, and 3 of life and observed a change in the microbiota in those with chorioamnionitis and funisitis only at week 1, allowing a causal association to be established [32]. This is consistent with our study, in which we found a clear difference in meconial microbiota between preterm

infants with and without sepsis, noting that chorioamnionitis is one of the major risk factors for early-onset neonatal sepsis.

Comparison of alpha diversity between different published studies remains a great challenge, as the methods used for collection of the analyzed material are heterogeneous. Both a difference in the microbiota and a reduction in bacterial diversity have been observed in infants who will go on to develop late onset neonatal sepsis as compared to a control group [7,33]. Madan et al. assessed 6 preterm infants and found that those who developed late-onset sepsis had a lower biodiversity in the fecal microbiota since birth [7]. However, none of these studies was able to describe a characteristic microbiome alteration that preceded development of sepsis. In this line, Dobbler et al. reported a reduction in bacterial diversity prior to the diagnosis of NEC [12]. Conversely, two studies found differences in the composition and characteristics of the intestinal microbiota between cases and controls, but no differences in alpha diversity between the two groups. Mai et al. compared the microbiota of preterm infants with and without a diagnosis of late onset sepsis, while Puri et al. compared the microbiota of preterm infants born to mothers with vs. without chorioamnionitis [9,32].

In our study, we found higher alpha diversity in those who developed sepsis, but the two groups showed great variation in the number of OTUs observed per patient, which allows us to infer that this microbiological criterion may not be useful to differentiate patients with a presumptive diagnosis of early-onset sepsis from those without clinical sepsis. In addition, the fecal samples analyzed in the aforementioned studies [9,32] were collected several days after NICU admission, unlike in our study, in which the first meconium passed by each patient was collected. The timing and setting of fecal collection influence the composition of the bacterial community and its colonization pattern, as described by Taft et al. in two different studies [33,34]. Even feces collected at the same hospital in different years showed differences in microbiota, in a previous study of preterm infants who did not develop NEC or late-onset sepsis [34].

We found that the phylum Proteobacteria was more abundant in cases of early-onset sepsis in our sample. This was also the most prevalent phylum in stool samples obtained before a diagnosis of late onset sepsis [9] and NEC [11], and can thus be considered a probable biomarker for diagnosis of these conditions.

Despite the technical difficulties in meconium collection, chief among them the delay in meconium passage and its reduced volume, we were able to collect a substantial number of samples in conditions that permitted reliable analysis. In addition, this was the first study to compare the microbial community present in the meconium of preterm neonates with and without a clinical diagnosis of early-onset sepsis in which the microbiota was analyzed by DNA amplification and 16S rRNA sequencing.

The main limitation of our study is the definition of early-onset neonatal sepsis, which was exclusively clinical, due to the absence of positive blood cultures in the first 72 h of life. It is worth noting that carrying out the same study on a sample limited to neonates with positive blood cultures – only 3% of the general neonatal population [2] – with a significant and representative sample would have required a multicenter design, which is associated with countless biases in microbiome analysis. These biases are related to differences in climate and environment, clinical practice, care patterns, diagnostic and treatment protocols in the population under study, and even

purchasing power and access to health care [33,34]. We need to take in account that a “convenience sampling” can represent a potential bias; however, studies of gut microbiome suffer numerous sources of bias that can impact any step of the pipeline and, in this study we followed strictly adequate methodology [35].

In conclusion, our findings confirm the hypothesis that the first-meconium microbiota of preterm infants with a clinical diagnosis of early-onset neonatal sepsis is different from that of preterm infants without sepsis, and that 14% of this difference is explained by the clinical diagnosis of sepsis. In addition, we observed a predominance of the phylum Proteobacteria in both groups, although it was more significant in cases of early-onset sepsis. These findings are consistent with previous international studies on the intestinal microbiota of preterm infants. As in these studies, we were unable to identify any single specific abnormality associated with diagnosis of sepsis.

The intestinal microbiota is a set of connected individuals and represents a large complex organism [36]. Therefore, one must seek an understanding of the microbiome as a whole and attempt to detect the moment at which an imbalance or change in its composition occurs [12]. Further studies are needed to better elucidate the intestinal microbiota and its function before we can routinely interfere with this complex community.

Author contributions

LVD, RSP, LFWR, ALC and RSC designed the research, acquisitioned, analyzed and interpreted the data, revised critically and approved the final version to be published.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article

Funding

Study funded by Conselho Nacional de Desenvolvimento Cientıfico e Tecnologico (CNPq), Fundac¸~ao Instituto de Pesquisas Economicas (FIPE-Brazil), and the Bill and Melinda^ Gates Foundation.

References

[1] Polin RA. Management of neonates with suspected or proven early-onset bacterial sepsis. Pediatrics. 2012; 129(5):1006–1015.

[2] Lacaze-Masmonteil T, Rosychuk RJ, Robinson JL. Value of a single C-reactive protein measurement at 18 h of age. Arch Dis Child Fetal Neonatal Ed. 2014;99(1): F76–F79.

[3] Zhou M, Cheng S, Yu J, et al. Interleukin-8 for diagnosis of neonatal sepsis: a meta-analysis. PLoS One. 2015;10(5):e0127170.

[4] Manzoni P. Hematologic aspects of early and lateonset sepsis in preterm infants. Clin Perinatol. 2015; 42(3):587–595.

[5] Santos RP, Tristram D. A practical guide to the diagnosis, treatment, and prevention of neonatal infections. Pediatr Clin North Am. 2015;62(2):491–508.
[6] Jimenez E, Marin ML, Martin R, et al. Is meconium from healthy newborns actually sterile? Res Microbiol. 2008;159:187–193.
[7] Madan JC, Salari RC, Saxena D, et al. Gut microbial colonisation in premature neonates predicts neonatal sepsis. Arch Dis Child Fetal Neonatal Ed. 2012;97(6): F456–F462.
[8] Ardissone AN, de la Cruz DM, Davis-Richardson AG, et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. PLoS One. 2014; 9(3):e90784.
[9] Mai V, Torrazza RM, Ukhanova M, et al. Distortions in development of intestinal microbiota associated with late onset sepsis in preterm infants. PLoS One. 2013; 8(1):e52876.
[10] Stewart CJ, Embleton ND, Marrs ECL, et al. Longitudinal development of the gut microbiome and metabolome in preterm neonates with late onset sepsis and healthy controls. Microbiome. 2017;5(1):75
[11] Mai V, Young CM, Ukhanova M, et al. Fecal microbiota in premature infants prior to necrotizing enterocolitis. PLoS One. 2011;6(6):e20647.
[12] Dobbler PT, Procianoy RS, Mai V, et al. Low microbial diversity and abnormal microbial succession is associated with necrotizing enterocolitis in preterm infants. Front Microbiol. 2017;8:2243
[13] Berrington JE, Stewart CJ, Cummings SP, et al. The neonatal bowel microbiome in health and infection. Curr Opin Infect Dis. 2014;27(3):236–243.
[14] Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet. 2005;365(9461):785–799.
[15] Ballard JL, Khoury JC, Wedig K, et al. New Ballard Score, expanded to include extremely premature infants. J Pediatr. 1991;119(3):417–423.
[16] Fenton TR, Kim JH. A systematic review and metaanalysis to revise the Fenton growth chart for preterm infants. BMC Pediatr. 2013;13(1):59.
[17] Jobe AH, Bancalari E. Bronchopulmonary dysplasia. Am J Respir Crit Care Med. 2001;163(7):1723–1729.
[18] Papile LA, Burstein J, Burstein R, et al. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1,500 gm. J Pediatr. 1978;92(4):529–534.
[19] Caporaso JG, Lauber CL, Walters WA, et al. Ultra-highthroughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. Isme J. 2012;6(8): 1621–1624.
[20] Pylro VS, Roesch LF, Morais DK, et al. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods. 2014;107: 30–37.
[21] Pylro VS, Morais DK, de Oliveira FS, et al. BMPOS: a flexible and user-friendly tool sets for microbiome studies. Microb Ecol. 2016;72(2):443–447.
[22] Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods. 2013; 10(10):996–998.
[23] Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335–336.
[24] McDonald D, Price MN, Goodrich J, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. Isme J. 2012;6(3):610–618.

[25] Good IJ. The population frequencies of species and the estimation of population parameters. Biometrika. 1953;40(3-4):237–264.
[26] Lemos LN, Fulthorpe RR, Triplett EW, et al. Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods. 2011;86(1): 42–51.
[27] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8(4):e61217
[28] Oksanen J, Blanchet FG, Kindt R, et al. Vegan: community ecology package. R package vegan, version. 2.2–1. 2015.
[29] Kinross JM, Darzi AW, Nicholson JK. Gut microbiomehost interactions in health and disease. Genome Med.
2011;3(3):14
[30] Wardwell LH, Huttenhower C, Garrett WS. Current concepts of the intestinal microbiota and the pathogenesis of infection. Curr Infect Dis Rep. 2011;13(1): 28–34.
[31] Schelonka RL, Chai MK, Yoder BA, et al. Volume of blood required to detect common neonatal pathogens. J Pediatr. 1996;129(2):275–278.
[32] Puri K, Taft DH, Ambalavanan N, et al. Association of chorioamnionitis with aberrant neonatal gut colonization and adverse clinical outcomes. PLoS One. 2016;
11(9):e0162734.
[33] Taft DH, Ambalavanan N, Schibler KR, et al. Center variation in intestinal microbiota prior to late-onset sepsis in preterm infants. PLoS One. 2015;10(6): e0130604
[34] Taft DH, Ambalavanan N, Schibler KR, et al. Intestinal microbiota of preterm infants differ over time and between hospitals. Microbiome. 2014;2(1):36.
[35] Ingala MR, Simmons NB, Wultsch C, et al. Comparing microbiome sampling methods in a wild mammal: fecal and intestinal samples record different signals of host ecology, evolution. Front Microbiol. 2018;9:803.
[36] Layeghifard M, Hwang DM, Guttman DS.
Disentangling interactions in the microbiome: a network perspective. Trends Microbiol. 2017;25(3): 217–228.

Figures

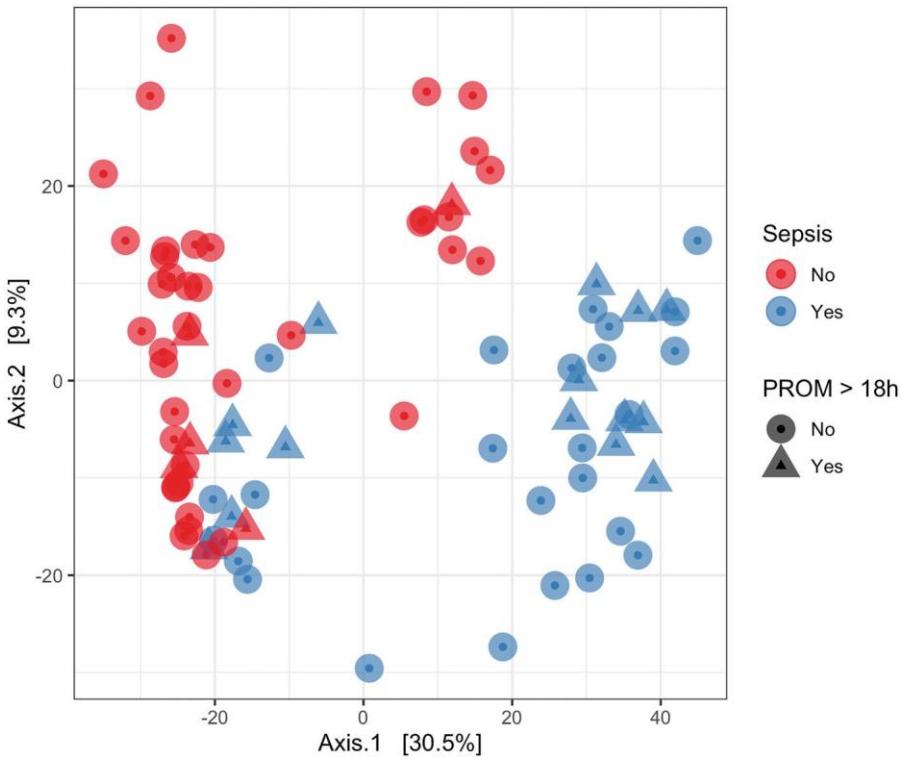


Figure 1. Principal coordinates analysis: principal coordinates analysis (PCoA). Cluster of microbial communities found in meconium samples from preterm neonates with and without sepsis. Each point represents an individual sample; colors indicate the presence or absence of sepsis.

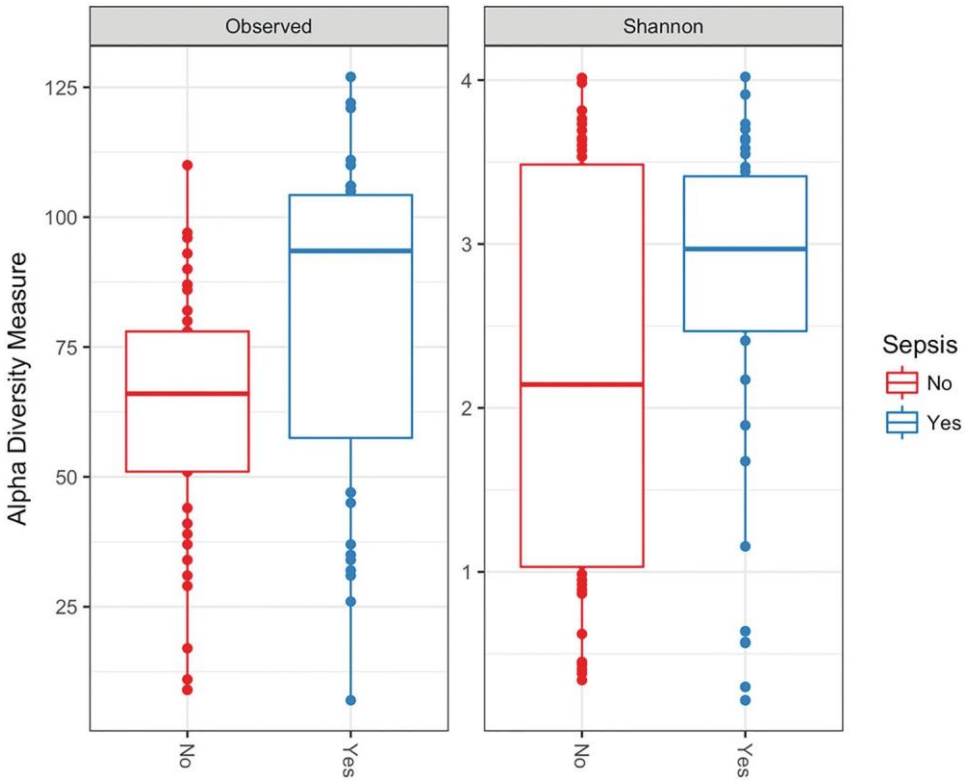


Figure 2. Alpha diversity and Shannon index: diversity of microbial communities found in meconium samples from preterm neonates with and without sepsis. Boxes denote the interquartile range; the

horizontal line inside the boxes represents the median. Whiskers denote variability outside the upper and lower quartiles. Observed, number of Operational Taxonomic Units (OTUs); Shannon, Shannon Diversity Index.

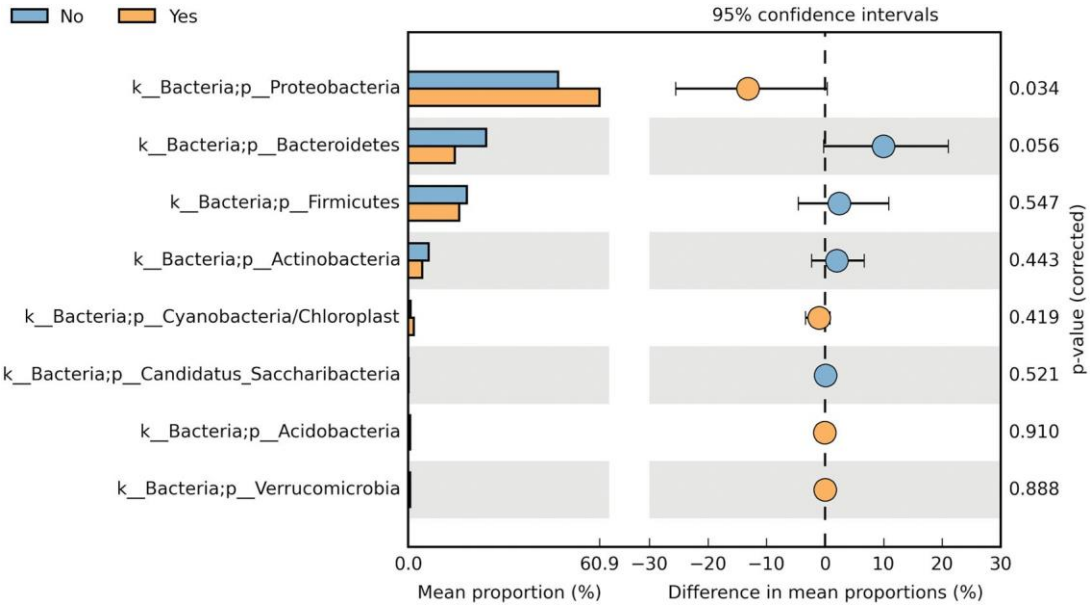


Figure 3. Abundance of microbial phyla: average abundance of phyla present in meconium samples from preterm neonates with and without sepsis. The error bars show the calculated standard deviation. The colored circles represent 95% confidence intervals, calculated using Welch's inverted method. Corrected p-values were calculated using Bonferroni's multiple test correction.

Tables

Table 1. Patient characteristics.

	Sepsis (n = 40)	No sepsis (n = 44)	p-Value
A. Maternal characteristics			
PROM > 18 h	15 (37.5%)	5 (11.4%)	.03 c
Urinary tract infection	9 (22.5%)	5 (11.4%)	.33 c
Chorioamnionitis	17 (42.5%)	11 (26.2%)	.25 c
Pre-eclampsia	11 (27.5%)	13 (29.5%)	.91 c
Antibiotic use	33 (82.5%)	28 (63.6%)	.20 c
Vaginal delivery	15 (37.5%)	12 (27.3%)	.54 c
B. Neonatal characteristics Male sex	24 (60%)	19 (43.2%)	.28 c
Gestational age	30 (±2.4)	30.5 (±2.2)	.29 a
Birth weight (g)	1395 (±496)	1428 (±497)	.69 a
SGA	10 (25%)	18 (40.9%)	.14 c
5-Min Apgar score	8 (7–9)	9 (8–9)	.06 b
RDS	21 (52.5%)	11 (25%)	.03 c
BPD	12 (30%)	11 (25%)	.90 c
Apnea	35 (87.5%)	29 (65.9%)	.11 c
Late-onset sepsis	12 (30%)	12 (27.3%)	.84 c
PIVH	10 (25%)	9 (20.5%)	.91 c
PVL	5 (12.5%)	3 (6.8%)	.66 c
PDA	4 (10%)	10 (22.7%)	.17 c
NEC	3 (7.5%)	7 (15.9%)	.35 c
Death	5 (12.5%)	4 (9.1%)	.94 c

PROM: premature rupture of membranes; BPD: bronchopulmonary dysplasia; RDS: respiratory distress syndrome; NEC: necrotizing enterocolitis; PDA: patent ductus arteriosus; PIVH: periventricular-interventricular hemorrhage; PVL: periventricular leukomalacia; SGA: small for gestational age.

Variables expressed as mean (SD), median (interquartile range), and absolute and relative frequencies. a: T-test; b: Mann–Whitney test; c: chi-square test.

Table 2. Linear discriminant analysis.

Genus	*p*-Value	FDR	LDA Score
A. Genera most associated with the sepsis group			
Paenibacillus	<.0001	<0.001	-1.55
Caulobacter	<.0001	<0.001	-1.13
Dialister	<.0001	<0.001	-0.322
Akkermansia	<.0001	<0.001	-0.382
Phenylobacterium	.000172	0.0017544	-1.95
Propionibacterium	.0020472	0.015447	-0.782
Ruminococcus	.0021201	0.015447	-0.528
Bradyrhizobium	.0025037	0.015961	-1.92
Alloprevotella	.0076397	0.043292	-0.319
B. Genera most associated with the no-sepsis group			
Flavobacterium	.009649	0.04921	0.937

Linear discriminant analysis (LDA) score denotes those genera most likely to explain the difference in microbial community between the sepsis and control groups. FDR: false discovery rate.

APÊNDICE D – Definindo Biomarcadores moleculares para risco de parto prematuro

Defining microbial biomarkers for risk of preterm labor

Anderson Santos de Freitas[1] & Priscila Caroline Thiago Dobbler[1] & Volker Mai[2] & Renato S Procianoy[3] & Rita C. Silveira[3] & Andréa Lúcia Corso[3] & Luiz Fernando Wurdig Roesch[1]

[1] Centro Interdisciplinar de Pesquisas em Biotecnologia – CIP-Biotec, Campus São Gabriel, Universidade Federal do Pampa, São Gabriel, Rio Grande do Sul, Brazil
[2] Department of Epidemiology, College of Public Health and Health Professions and College of Medicine, Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA
[3] Serviço de Neonatologia do Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

Correspondent Author: Luiz Fernando Wurdig Roesch luizroesch@unipampa.edu.br

DOI: https://doi.org/10.1007/s42770-019-00118-x
Received: 25 April 2019
Accepted: 10 July 2019
Published online: 22 July 2019

Keywords: Next-generation sequencing. Vaginal microbiome. Lactobacillus. Brazilian microbiome

Abstract
Preterm birth remains the main contributor to early childhood mortality. The vaginal environment, including microbiota composition, might contribute to the risk of preterm

delivery. Alterations in the vaginal microbial community structure might represent a risk factor for preterm birth. Here, we aimed to (a) investigate the association between preterm birth and the vaginal microbial community and (b) identify microbial biomarkers for risk of preterm birth. Microbial DNA was isolated from vaginal swabs in a cohortof69women enrolled at hospital admission for their delivery. Microbiota was analyzed by high-throughput 16S rRNA sequencing. While no differences in microbial diversity measures appeared associated with the spontaneous preterm and full-term outcomes, the microbial composition was distinct for these groups. Differential abundance analysis showed Lactobacillus species to be associated with full-term birth whereas an unknown Prevotella species was more abundant in the spontaneous preterm group. Although we studied a very miscegenated population from Brazil, our findings were similar to evidence pointed by other studies in different countries. The role of Lactobacillus species as a protector in the vaginal microbiome is demonstrated to be also a protector of spontaneous preterm outcome whereas the presence of pathogenic species, such as Prevotella spp., is endorsed as a factor of risk for spontaneous preterm delivery.

Introduction

According to the World Health Organization, every birth before 37 weeks of pregnancy is considered preterm. Each year, about 15 million babies are born prematurely in the world [1]. Prematurity is the leading cause of mortality before 4 weeks of life and the second until 5 years of age [1, 2]. Preterm birth also leads to disorders related to brain development, the deficit of attention, hyperactivity [3–5], autism [6], and respiratory problems [7]. In the USA, the preterm delivery rate is around 9.6% [8] while in Europe and other developed countries it is between 5 and 9% [9]. In developing countries, especially South Asia and sub-Saharan Africa, the preterm birth rates are above 15% [1].

The epidemiological and clinical natures of preterm birth are not yet fully understood [2, 10, 11]. Nevertheless, preterm delivery is associated with type 2 diabetes [12], weight gain, chronic postpartum hypertension [13], air pollution [14], psychological and social conditions, physical exertion during pregnancy [15], diet, hygiene, and access to health care [2, 11]. Several studies attempted to map the endemicity of this disease and their results indicate a higher incidence in black women, in women under low socioeconomic levels, in smokers, in pregnancies of twins, and/or more advanced age [1, 2, 11, 16–20].

Different microbes also have been correlated with preterm delivery [21, 22], but microbial community-level studies represent a suitable and fast alternative to better understand the relationship between the microbial community and the preterm birth. As women from different ethnic backgrounds have different vaginal microbial communities [23–25], local attempts to detect and associate microbial communities with preterm birth are required [26]. Such regional attempts might sum up with other worldwide initiatives to elaborate a prediction risk assessment plan based on the vaginal microbial community. Within this work we aimed to (a) investigate the association between preterm birth and the vaginal microbial community and (b) identify microbial biomarkers for risk of preterm birth.

Material and methods

Experimental design

This study was carried out with samples collected from women attending the Hospital de Clínicas de Porto Alegre (HCPA). Experimentation used a convenience sampling strategy. Expecting mothers were enrolled at hospital admission for their delivery between May 2014and March 2016. All women provided written informed consent to allow their samples to be used in the study. The ethics committee of HCPA approved the study protocol. Exclusion criteria were: (1) HIV or congenital infections, (2) drug user or alcoholic, (3) urinary tract infections or (4) antibiotic usage in the third trimester of gestation, (5) urogenital infection in the last 3 months, and (6) gestational diabetes.

Sixty-nine pregnant women were analyzed in this study. Twenty-three of them had spontaneous preterm labor (before 33 weeks of gestation), whereas 29 had spontaneous term labor. Another 17 women had non-spontaneous labor but, due to medical reasons affecting pregnancy, had cesarean delivery before 33 gestational weeks. Those subjects called hereinafter “non-spontaneous preterm” group, were used as a second control because they present a microbial community that might not be associated with spontaneous preterm delivery but have a better match in terms of gestational age with the spontaneous preterm group. All pregnant women sampled on this work had vaginal swab (Sterile Specimen Collection Swabs to collect specimens from soft tissue surfaces-Labor swab®) collected up to 4 hours before labor begins, as described by Roesch and colleagues [27]. Collected swab samples were stored at − 80 °C until DNA extraction. The characteristics from the mothers enrolled in this experiment include maternal age, previous pregnancies, gestational age, incidence of chorioamnionitis, preeclampsia, infection by Group B *Streptococcus*, intrapartum penicillin administration, and delivery mode.

Microbial DNA extraction, 16S rRNA amplification and sequencing, and data processing

Microbial DNA was extracted from frozen swab samples as previously described by Roesch et al. [27]. All DNA samples were kept at − 20 °C until use in PCR reactions. Vaginal microbiota was determined by amplification of the V4 region of the 16S rRNA gene and downstream sequencing on the Ion PGM Platform (Thermo Fisher Scientific, Waltham, MA, USA) with the bacterial/archaeal primers 515F and 806R [28]. All samples were PCR-amplified using barcoded primers linked with the Ion adapter “A” sequence (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′) and Ion adapter “P1” sequence (5′-CCTCTCTATGGGCAGTCGGT GAT-3′) to obtain a sequence of primer composed by Abarcode-806R and P1-515F adapter and primers. PCRs were carried out in25μL reactions contained 2 U of Platinum® Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4 μL 10X High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.2 mM dNTP’s, 0.1 μM of both the 806R barcoded primer and the 515F primer, 25 μg of Ultrapure BSA (Invitrogen, Carlsbad, CA, United States) and approximately 50 ng of

DNA template. PCR conditions used were: 95 °C for 5 min; 30 cycles of 94 °C per 45 s denaturation, 56 °C per 45 s annealing, and 72 °C per 1 min extension; followed by 72 °C per 10 min for final extension. Fragments presenting around 400 base pairs from resulting PCR products were purified with the Agencourt® AMPure® XP Reagent (Beckman Coulter, Brea, CA, USA), and final concentration of the PCR products was quantified by using the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, United States) following manufacturer's instructions. Finally, reactions were combined in equimolar concentrations to create a mixture composed of amplicon fragments of each sample. This composite sample was used for library preparation with Ion OneTouch™ 2 System with the Ion PGM™ Template OT2 400 Kit Template (Thermo Fisher Scientific, Waltham, MA, USA). The sequencing was performed using Ion PGM™ Sequencing 400 on Ion PGM™ System using Ion 318™ Chip v2 with a maximum of 40 samples per microchip.

Raw reads were analyzed according to the pipeline proposed by the Brazilian Microbiome Project [29]. A table of operational taxonomic units (OTUs) was constructed by using the UPARSE pipeline [30], with a minimum similarity cutoff value of 97% for clustering and a maximum expected error of 0.5%. Taxonomic classifications were made on QIIME 1.9.0 [31], based on UCLUST method against the SILVA ribosomal RNA gene database version 128[32] with a confidence interval of 95%. Sampling effort was measured by the Good's coverage [33].

Data analyses

Maternal variables were analyzed into the R environment [34]. Numeric variables were summarized as average ± SEM and compared using the Kruskal-Wallis followed by a post hoc Dunn test. Categorical variables were compared using chi square post hoc test.

The 16S rRNA database was analyzed through the phyloseq [35] and the Microbiome [36] packages after removing singletons and rarefying the dataset to the minimum library size. Possible confounding variables were tested by permutational multivariate analysis of variance (PERMANOVA) into the vegan package [37].

Initial insights about general microbial structure were provided by analyses of relative abundance (measured by Kruskall-Wallis post hoc Dunn test) of the most frequent genera and alpha diversity tests.

The differential abundance analysis, applied to find microbial biomarkers of preterm birth, was performed by using the DESeq2 [38] with the raw (non-rarefied) dataset. Briefly, after removing the samples from mothers treated with intrapartum penicillin from the dataset, the OTU table was conglomerated at species level. Taxa not seen more than 3 times in at least 20% of the samples were removed and the number of sequences per OTU was transformed by calculating the geometric mean. Two different contrasts were applied: (a) term birth versus preterm birth and (b) false-preterm birth versus preterm birth. The FDR method was used to control for false discovery rate. Additional correlation analysis between OTUs was tested by using SparCC approach [39].

Results

Maternal variables used for comparison between groups

The characteristics of the three groups are shown in Table 1. Maternal age ranged from 23 to 30 years but was not significantly different between women with subsequent term or preterm labor. Nonetheless, women significantly older composed the non-spontaneous preterm group. The number of previous pregnancies was also similar in all, spontaneous term labor and spontaneous preterm labor groups as well as in nonspontaneous preterm group. Moreover, as expected, gestational age was significantly different between term and preterm groups, but not different between spontaneous preterm and non-spontaneous preterm groups. Women who had preterm labor presented a higher incidence of chorioamnionitis than the term group. This condition did not differ between term and non-spontaneous preterm groups. Only the preterm and the non-spontaneous preterm groups presented cases of infection by Group B *Streptococcus* (GBS). Although the GBS infection rate was higher in the non-spontaneous preterm group, the incidence of GBS infection was not statistically different between spontaneous preterm and non-spontaneous preterm groups. Most cases of non-spontaneous preterm labor presented preeclampsia, whereas there were no cases in the term group and only three cases in the spontaneous preterm labor group. The intrapartum penicillin was used in 15 preterm cases. The term and non-spontaneous preterm groups did not receive prophylactic antibiotics. Intravenous penicillin was administered approximately 4 h before labor in many preterm samples due to a positive test for GBS or due a suspicion of infection in absence of a test. Finally, the three groups differed in terms of vaginal or cesarean delivery. The preterm condition was the major driver of cesarean, especially in the nonspontaneous preterm labor group.

Controlling for confounding variables

Permutational analysis of variance was applied to test the effect of confounding variables on the microbiota analyses (Table 2).

As significant reduction in taxonomic diversity of vaginal microbial community was already observed as pregnancy advances [40], we first attempted to verify the influence of gestational age on the vaginal microbiota. A pairwise analysis among the spontaneous term labor, spontaneous preterm labor, and nonspontaneous preterm labor groups revealed undetectable microbial community differences among those groups in our dataset (Table 2). In all comparisons, R^2 was smaller than 1% and the p value was greater than 0.05. Fifteen out of 23 women from the spontaneous preterm labor group received prophylactic antibiotics during labor, whereas no women from the term group received antibiotics. The R^2 for antibiotic usage was 0.034 and the p value was 0.043 indicating that about 3.4% of the variation in the microbial community between groups was explained by the prophylactic use of antibiotics during labor. Intrapartum antibiotics were administrated only in cases with preterm labor. For this reason, this factor could not be used in a multi-factor design with interactions. All OTUs

under intrapartum antibiotics influence were removed from the dataset prior to diversity and differential abundance analysis.

Different microbial community structure but similar vaginal microbial diversity within treatments

The mean of organisms' abundance was particularly similar between term and non-spontaneous preterm groups as indicated by the Kruskal-Wallis post hoc Dunn test. On the other hand, preterm group presented a low mean of OTUs closest related to the genus *Lactobacillus* among its samples jointly with a tendency to a high mean of OTUs with best hit to *Prevotella* and *Pseudomonas* when compared with terms and non-spontaneous preterm (Fig. 1). Overall, the alpha diversity was low among all samples. The non-parametric Wilcox test indicated no differences in microbial diversity among the three groups tested using either Shannon or inverse of Simpson diversity indexes (Fig. 2).

Defining the main biomarkers associated with term and preterm delivery

To outline the main microbes associated with the term and preterm labor we performed a differential abundance analysis. Pairwise comparisons among spontaneous term and spontaneous preterm groups revealed the abundance of OTUs closest related to two species of *Lactobacillus* associated with the term birth (Table 3). They were *Lactobacillus iners* and *Lactobacillus jensenii*. An unknown species related to the *Prevotella* genus was more abundant in the spontaneous preterm group. Similar tendency was also observed when comparing non-spontaneous preterm labor and spontaneous preterm labor groups. An OTU with the best hit to *Lactobacillus jensenii* was associated with subjects with non-spontaneous preterm labor while two taxa related to the *Prevotella* genus were associated with the spontaneous preterm labor. In an attempt to verify whether *Prevotella* was associated with *Lactobacillus*, we perform a correlation analysis at the genus level by using the SparCC approach [39]. No significant strong correlation (correlation coefficient = − 0.17 and p = 0.09) was found involving either *Lactobacillus* or *Prevotella*.

The results indicated that the absence of high numbers of OTUs classified as *Lactobacillus*, particularly as *Lactobacillus iners* and *Lactobacillus jensenii*, might be the main difference between the vaginal microbial community of pregnant women following term or spontaneous preterm labor.

Discussion

In this work, we attempted to detect biomarkers for preterm labor on the vaginal microbiota of pregnant women. Several studies have described the vaginal microbiota of pregnant women; however, most of them were based in the USA, Canada, Europe, or Mexico [23–25, 41–43] and came to very incongruent results. A North American NGS-based study performed by Romero and colleagues, for example, concluded that there was no difference between abundance and structure of the vaginal microbiome, independent of the type of birth [23]. On the other hand, the efforts by Hyman and collaborators, whom worked with chain-termination sequencing, conclude that mothers

whom deliver prematurely present a high diverse vaginal microbiota [42]. In addition, a Canadian study by Freitas et al. not only correlates high diversity on vaginal microbiota to preterm delivery but also the presence of Mollicutes [43].

Here, we used next-generation sequencing to analyze a Brazilian cohort composed of 69 pregnant women. The unique feature of this work is the high miscegenation rates of the Brazilian population. The aforementioned studies suggest that women from different ethnic backgrounds have

different vaginal microbial communities. Therefore, investigations using cohorts with different ethnic backgrounds are important to better understand the etiology of preterm labor and its relationship with microbes.

Vaginal microorganisms possess a known key role in states of health and disease acting as both generators/stimulators and protectors from diseases [24, 44, 45]. Interactions between the microbiota and human diseases occur in a two-way process. Bacteria can cause diseases as much as states of diseases can cause changes in the normal microbiota. An example is the increase in bacterial pathogen abundance in cases of depression. Gut microbes can produce identical hormones and neurotransmitters produced by humans. In turn, the bacterial receptors for these hormones influence microbial growth [46].

In this context, we presented multiple lines that lead to the presence of different vaginal microbial communities associated with the full-term and spontaneous preterm labor. The first evidence was provided by overall abundance analysis (Fig. 1). Preterm group represented differences when compared with other ones, mainly related to the decrease of general abundance of *Lactobacillus*. The second and most strong evidence was obtained by a differential abundance analysis (Table 3). Lower numbers of OTUs with best hit to species from the genus Lactobacillus were associated with the spontaneous preterm labor while vaginal bacterial communities rich in these microbial species (e.g., spontaneous term and non-spontaneous preterm groups) were associated with the full-term outcome. Non-spontaneous preterm labor presented similar microbial communities composition to those subjects with spontaneous term labor. Indeed, the healthy vaginal microbiota in the Brazilian pregnant woman has low microbial diversity and is dominated by Lactobacillus species [27]. Besides, Lactobacillus species are very often correlated to states of health in the vaginal environment [24, 47–50]. Bacteria from this genus present a fermentative metabolism with lactate and usually acetate, ethanol, CO_2, formate, or succinate as products [51]. These compounds acts lowering the vaginal pH to levels around 4,5 and creating an inhospitable environment for most of pathogenic species [49, 52, 53].

On the other hand, we were able to detect the presence of OTUs closest related to the genus *Prevotella* in association with the spontaneous preterm labor. In fact, many microorganisms, just like Prevotella species, can produce proinflammatory substances that can also lead to a preterm birth [54]. Studies point to adaptation of specific *Prevotella* strains at different niches. The report by Gupta and colleagues, for example, showed 83% of the Prevotella genome may contribute to singletons and flexible sequences and this condition performs a key role in the adaptation to many body sites [55]. Indeed, several works indicate Prevotella strains related dysbiosis in states of disease in highly different body parts, i.e., asthma and bacterial vaginosis [56, 57].

Prevotella is still correlated with inflammatory processes by the activation of Toll-like receptor 2, which leads to production of T helper type 17 cells (Th17) and increase of interleukin 8 (IL-8), interleukin 6 (IL-6), and chemokine (C-C motif) ligand 20 (CCL20) [58]. In addition, the intrauterine infection, which may have originated in the vaginal cavity, might account for 25–40% of preterm births [59]. The most commonly associated bacteria are bacteria from the class *Mollicutes* (*Ureaplasma* species, *Mycoplasma genitalium*, and M. hominis, for example) [60, 61], but many other microbial species have been identified in cases of bacterial vaginosis, including *Prevotella* [62–64]. Those microbes might invade the uterus by migrating from the passage through the cervix from the vagina and infect the amniotic fluid [59]. The metabolism of some of these bacteria may also produce urease, an enzyme that catalyzes the hydrolysis of urea into carbon dioxide and ammonia. Its activity increases the vaginal pH, a stress environmental condition for the mother and the fetus that may influence in a spontaneous premature outcome [65, 66].

Callahan et al. [67] recently presented similar results. The authors studied two cohorts from different locations of the USA. A lower abundance of *L. crispatus* was significantly associated with the preterm birth in both cohorts. In line with our results, the cohort from Birmingham, AL, presented decreased abundance of *L. jensenii* associated with the preterm birth. But contrary to our results, no significant association was detected for *L. iners*. Moreover, among the women with lower levels of *Lactobacillus*, a higher abundance of *Gardnerella* and *Ureaplasma* was associated with the increased risk of preterm labor. According to Baldwin et al. [68], Lactobacillus spp. were markedly decreased when compared with vaginal swabs collected from uncomplicated pregnancy subjects with a matched gestational time. As observed in our dataset, the authors also observed deficiency of *Lactobacillus* and persistence of known pathogenic species, such as *Prevotella* sp., as a risk factor for preterm birth.

In short, reports from the aforementioned studies as well as from this one converge to a pattern of bacteria either pathogenic or related to stress conditions as increased in preterm cases. Considering this fact and the niche adaptation performed by Prevotella spp. [55], we are able to suggest Prevotella as a microbial biomarker for preterm labor in the vaginal microbiota.

Conclusion

The relationship between the vaginal microbes and the spontaneous preterm labor was already described in racially distinct cohorts. In spite of this, to the best of our knowledge, this is the first study to describe and correlate the vaginal microbiota with the spontaneous preterm labor in a Brazilian cohort. This is especially important because: (i) preterm-microbiota associations are population dependent [67] and (ii) the Brazilian population presents high rates of miscegenation. As so, this population cannot be classified using standard stratifications of Caucasian/white and black/African American. Our results add to the ecological theory of the protective effect of *Lactobacillus* and the occurrence of other pathogenic taxa (e.g., *Prevotella*) as a possible risk factor for preterm labor.

Funding information This work was supported by the Bill & Melinda Gates Foundation (OPP1107597), CNPQ, and DECIT/Ministério da Saúde do Brasil. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brasil (CAPES) (Finance Code 001). L.F.W. Roesch and AS de Freitas received research fellowships from the CNPq.

Compliance with ethical standards

The ethics committee of HCPA approved the study protocol.

References

1. Althabe F, Howson CP, Kinney M, Lawn J (2012) World Health Organization. Born too soon: The global action report on preterm birth. http://www.who.int/pmnch/media/news/2012/201204_ borntoosoon-report.pdf. Accessed March 24, 2017
2. Goldenberg RL, Culhane JF, Iams JD, Romero R (2008) Epidemiology and causes of preterm birth. Lancet 371(9606):75– 84
3. Larroque B, Ancel P-Y, Marret S, Marchand L, André M, Arnaud C, Pierrat V, Rozé JC, Messer J, Thiriez G, Burguet A, Picaud JC, Bréart G, Kaminski M (2008) Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. Lancet 371(9615):813– 820
4. Ment LR, Vohr BR (2008) Preterm birth and the developing brain. Lancet Neurol 7(5):378–379
5. Rommel A-S, James S-N, McLoughlin G, Brandeis D, Banaschewski T, Asherson P, Kuntsi J (2017) Association of preterm birth with attention-deficit/hyperactivity disorder–like and wider-ranging neurophysiological impairments of attention and inhibition. J Am Acad Child Adolesc Psychiatry 56(1):40–50
6. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N (2010) Autism spectrum disorders in extremely preterm children. The Journal of Pediatrics 156(4):525–531.e2. https://doi.org/10. 1016/j.jpeds.2009.10.041
7. Rosen CL, Larkin EK, Kirchner HL, Emancipator JL, Bivins SF, Surovec SA, Martin RJ, Redline S (2003) Prevalence and risk factors for sleep-disordered breathing in 8- to 11-year-old children: association with race and prematurity. J Pediatr 142(4):383–389. https://doi.org/10.1067/mpd.2003.28
8. Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Mathews TJ (2017) Births: final data for 2015. Natl Vital Stat Rep 66(1):1
9. Zeitlin J, Szamotulska K, Drewniak N, Mohangoo AD, Chalmers J, Sakkeus L, Irgens L, Gatt M, Gissler M, Blondel B, The EuroPeristat Preterm Study Group (2013) Preterm birth time trends in Europe: a study of 19 countries. BJOG Int J Obstet Gynaecol 120(11):1356–1365. https://doi.org/10.1111/1471-0528.12281
10. Howson CP, Kinney MV, McDougall L, Lawn JE (2013) Born too soon: preterm birth matters. Reprod Health 10(1):S1
11. Slattery MM, Morrison JJ (2002) Preterm delivery. Lancet 360(9344):1489–1497

12. Kajantie E, Osmond C, Barker DJP, Eriksson JG (2010) Preterm birth–a risk factor for type 2 diabetes?: the Helsinki birth cohort study. Diabetes Care 33(12):2623–2625. https://doi.org/10.2337/ dc10-0912
13. Tanz LJ, Stuart JJ, Williams PL, Rimm EB, Missmer SA, Rexrode KM, Mukamal KJ, Rich-Edwards JW (2017) Preterm delivery and maternal cardiovascular disease in young and middle-aged adult women clinical perspective. Circulation. 135(6):578–589. https:// doi.org/10.1161/CIRCULATIONAHA.116.025954
14. Huynh M, Woodruff TJ, Parker JD, Schoendorf KC (2006) Relationships between air pollution and preterm birth in California. Paediatr Perinat Epidemiol 20(6):454–461
15. Steer P (2005) The epidemiology of preterm labour. BJOG Int J Obstet Gynaecol 112(s1):1–3
16. Ahern J, Pickett KE, Selvin S, Abrams B (2003) Preterm birth among African American and white women: a multilevel analysis of socioeconomic characteristics and cigarette smoking. J Epidemiol Community Health 57(8):606–611
17. Ananth CV, Misra DP, Demissie K, Smulian JC (2001) Rates of preterm delivery among black women and white women in the United States over two decades: an age-period-cohort analysis. Am J Epidemiol 154(7):657–665
18. Kindinger LM, Bennett PR, Lee YS, et al. (2017) The interaction between vaginal microbiota, cervical length, and vaginal progesterone treatment for preterm birth risk. Microbiome 5(1). https://doi. org/10.1186/s40168-016-0223-9
19. Kramer MS, Goulet L, Lydon J, Seguin L, McNamara H, Dassa C, Platt RW, Fong Chen M, Gauthier H, Genest J, Kahn S, Libman M, RozenR, MasseA, MinerL, Asselin G, BenjaminA, KleinJ, Koren G (2001) Socio-economic disparities in preterm birth: causal pathways and mechanisms. Paediatr Perinat Epidemiol 15(s2):104–123
20. Smith LK, Draper ES, Manktelow BN, Dorling JS, Field DJ (2007) Socioeconomic inequalities in very preterm birth rates. Arch Dis Child Fetal Neonatal Ed 92(1):F11–F14. https://doi.org/10.1136/ adc.2005.090308
21. Son K-A, Kim M, Kim YM, Kim SH, Choi SJ, Oh SY, Roh CR, Kim JH (2018) Prevalence of vaginal microorganisms among pregnant women according to trimester and association with preterm birth. Obstet Gynecol Sci 61(1):38–47. https://doi.org/10.5468/ ogs.2018.61.1.38
22. Nelson DB, Hanlon A, Hassan S, Britto J, Geifman-Holtzman O, Haggerty C, Fredricks DN (2009) Preterm labor and bacterial vaginosis-associated bacteria among urban women. J Perinat Med 37(2):130–134. https://doi.org/10.1515/JPM.2009.026
23. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Bieda J, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J (2014) The vaginal microbiota of pregnant women who subsequently have spontaneous preterm labor and delivery and those with a normal delivery at term. Microbiome. 2(1):18
24. Hernández-Rodríguez C, Romero-González R, Albani-Campanario M, Figueroa-Damián R, Meraz-Cruz N, Hernández-Guerrero C (2011) Vaginal microbiota of healthy pregnant Mexican women is constituted by four Lactobacillus species and several vaginosisassociated Bacteria. Infect Dis Obstet Gynecol 2011:1–9. https:// doi.org/10.1155/2011/851485

25. MacIntyre DA, Chandiramani M, Lee YS et al (2015) The vaginal microbiome during pregnancy and the postpartum period in a European population. Sci Rep 5:8988. https://doi.org/10.1038/ srep08988
26. Pylro VS, Morais DK, Roesch LFW (2015) Microbiome studies need local leaders. Nature. 528:39
27. Roesch LFW, Silveira RC, Corso AL, Dobbler PT, Mai V, Rojas BS, LaureanoÁM,Procianoy RS(2017) Diversityandcomposition of vaginal microbiota of pregnant women at risk for transmitting group B Streptococcus treated with intrapartum penicillin. PLoS One 12(2):e0169916
28. Caporaso JG, Lauber CL, Walters WA et al (2012) Ultra-highthroughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME Journal 2012(6):1621–1624
29. Pylro VS, Roesch LFW, Morais DK, Clark IM, Hirsch PR, Tótola MR (2014) Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods 107:30–37. https://doi.org/10.1016/j.mimet.2014.08.018
30. Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10(10):996–998. https://doi.org/10.1038/nmeth.2604
31. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of highthroughput community sequencing data. Nat Meth 7(5):335–336. https://doi.org/10.1038/nmeth.f.303
32. QuastC,Pruesse E, YilmazP, Gerken J,SchweerT, YarzaP, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41(Database issue):D590–D596. https://doi.org/10. 1093/nar/gks1219
33. Good IJ (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40(3/4):237. https:// doi.org/10.2307/2333344
34. R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. http://www.R-project.org
35. McMurdie PJ, Holmes S (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8(4):e61217. https://doi.org/10.1371/journal.pone. 0061217
36. Lahti L, Shetty S, Blake T, Salojarvi J (2017) Microbiome r package. Tools Microbiome Anal R
37. Oksanen J, Blanchet FG, Kindt R, et al. (2015) Vegan: community ecology package. R Package Vegan, Vers. 2.2–1. World

    Agroforestry Centre Nairobi, Kenya
38. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15(12). https://doi.org/10.1186/s13059-014-0550-8
39. Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. von Mering C, ed. PLoS Comput Biol 8(9): e1002687. https://doi.org/10.1371/journal.pcbi.1002687

40. Aagaard K, Riehle K, Ma J et al (2012)A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. Ratner AJ, ed. PLoS ONE 7(6):e36466. https://doi.org/10.1371/ journal.pone.0036466
41. Romero R, Hassan SS, Gajer P, Tarca AL, Fadrosh DW, Nikita L, Galuppi M, Lamont RF, Chaemsaithong P, Miranda J, Chaiworapongsa T, Ravel J (2014) The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women. Microbiome. 2(1):4. https://doi. org/10.1186/2049-2618-2-4
42. Hyman RW, Fukushima M, Jiang H, Fung E, Rand L, Johnson B, Vo KC, Caughey AB, Hilton JF, Davis RW, Giudice LC (2014) Diversity of the vaginal microbiome correlates with preterm birth. Reprod Sci 21(1):32–40. https://doi.org/10.1177/ 1933719113488838
43. the VOGUE Research Group, Freitas AC, Bocking A, Hill JE, Money DM (2018) Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. Microbiome 6(1). https://doi.org/10.1186/s40168-018-0502-8
44. Huttenhower C, Gevers D, Knight R et al (2012) Structure, function and diversity of the healthy human microbiome. Nature 486(7402): 207–214. https://doi.org/10.1038/nature11234
45. DiGiulio DB, Callahan BJ, McMurdie PJ et al (2015) Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci 112(35):11060–11065. https://doi.org/10.1073/pnas. 1502875112
46. Galland L (2014) The gut microbiome and the brain. J Med Food 17(12):1261–1272. https://doi.org/10.1089/jmf.2014.7000
47. Gajer P, Brotman RM, Bai G et al (2012) Temporal dynamics of the human vaginal microbiota. Science Translational Medicine 4(132): 132ra52–132ra52. https://doi.org/10.1126/scitranslmed.3003605
48. Ravel J, Gajer P, Abdo Z, et al. Vaginal microbiome of reproductive-age women. Proceedings of the National Academy of Sciences. 2011;108(Supplement_1):4680–4687. https://doi.org/ 10.1073/pnas.1002611107
49. Ma B, Forney LJ, Ravel J (2012) Vaginal microbiome: rethinking health and disease. AnnuRev Microbiol 66(1):371–389.https://doi. org/10.1146/annurev-micro-092611-150157
50. Lamont R, Sobel J, Akins R, Hassan SS, Chaiworapongsa T, Kusanovic JP, Romero R (2011) The vaginal microbiome: new information about genital tract flora using molecular based techniques: vaginal microbiome using molecular tools. BJOG Int J Obstet Gynaecol 118(5):533–549. https://doi.org/10.1111/j.14710528.2010.02840.x
51. Hammes WP, Hertel C (2015) Lactobacillus. In: Whitman WB, Rainey F, Kämpfer P, et al., eds. Bergey's manual of systematics of archaea and bacteria. Chichester, UK: John Wiley & Sons, Ltd: 1–76. https://doi.org/10.1002/9781118960608.gbm00604
52. Boskey ER, Telsch KM, Whaley KJ, Moench TR, Cone RA (1999) Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. Infect Immun 67(10):5170
53. Linhares IM, Summers PR, Larsen B, Giraldo PC, Witkin SS (2011) Contemporary perspectives on vaginal pH and lactobacilli. American Journal of Obstetrics and Gynecology 204(2):120.e1– 120.e5. https://doi.org/10.1016/j.ajog.2010.07.010

54. YangS,ReidG,ChallisJRG,KimSO,GloorGB,BockingAD(2015) Is there a role for probiotics in the prevention of preterm birth? Front Immunol 6. https://doi.org/10.3389/fimmu.2015.00062
55. Gupta VK,ChaudhariNM, Iskepalli S,Dutta C (2015)Divergences in gene repertoire among the reference Prevotella genomes derived from distinct body sites of human. BMC Genomics 16(1):153–153. https://doi.org/10.1186/s12864-015-1350-6
56. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt MF, Cookson WOC (2010) Disordered microbial communities in asthmatic airways. PLoS One 5(1):e8578–e8578. https://doi.org/10.1371/journal. pone.0008578
57. Zozaya-Hinchliffe M, Lillis R, Martin DH, Ferris MJ (2010) Quantitative PCR assessments of bacterial species in women with and without bacterial vaginosis. J ClinMicrobiol 48(5):1812–1819. https://doi.org/10.1128/JCM.00851-09
58. Larsen JM (2017) The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology. 151(4):363–374. https://doi.org/10.1111/imm.12760
59. Goldenberg RL, Hauth JC, Andrews WW (2000) Intrauterine infection and preterm delivery. N Engl J Med 342(20):1500–1507. https://doi.org/10.1056/NEJM200005183422007
60. Kataoka S, Yamada T, Chou K, Nishida R, Morikawa M, Minami M, Yamada H, Sakuragi N, Minakami H (2006) Association between preterm birth and vaginal colonization by mycoplasmas in early pregnancy. J Clin Microbiol 44(1):51–55. https://doi.org/10. 1128/JCM.44.1.51-55.2006
61. Viscardi RM (2010) Ureaplasma species: role in diseases of prematurity. Clin Perinatol 37(2):393–409. https://doi.org/10.1016/j.clp. 2009.12.003
62. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, Cotch MF, Edelman R, Pastorek JG, Rao AV, McNellis D, Regan JA, Carey JC, Klebanoff MA (1995) Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The vaginal infections and prematurity study group. N Engl J Med 333(26):1737–1742. https://doi.org/ 10.1056/NEJM199512283332604
63. Tebes CC, Lynch C, Sinnott J (2003) The effect of treating bacterial vaginosis on preterm labor. Infect Dis Obstet Gynecol 11(2):123– 129. https://doi.org/10.1080/10647440300025509
64. Onderdonk AB, Delaney ML, Fichorova RN (2016) The human microbiome during bacterial vaginosis. Clin Microbiol Rev 29(2): 223–238. https://doi.org/10.1128/CMR.00075-15
65. Humphries C (2017) Detecting diversity. Nature Publishing Group Macmillan Building, 4 Crinan St, London N1 9xw, England
66. Kenny GE, Cartwright FD (1977) Effect of urea concentration on growth of Ureaplasma urealyticum (T strain mycoplasma). J Bacteriol 132(1):144–150
67. Callahan BJ, DiGiulio DB, Goltsman DSA et al (2017) Replication and refinement of a vaginal microbial signature of preterm birth in two racially distinct cohorts of US women. Proc Natl Acad Sci U S A 114(37):9966–9971. https://doi.org/10.1073/pnas.1705899114

68. Baldwin EA, Walther-Antonio M, MacLean AM et al (2015) Persistent microbial dysbiosis in preterm premature rupture of membranes from onset until delivery. PeerJ. 3:e1398. https://doi. org/10.7717/peerj.1398

Figures

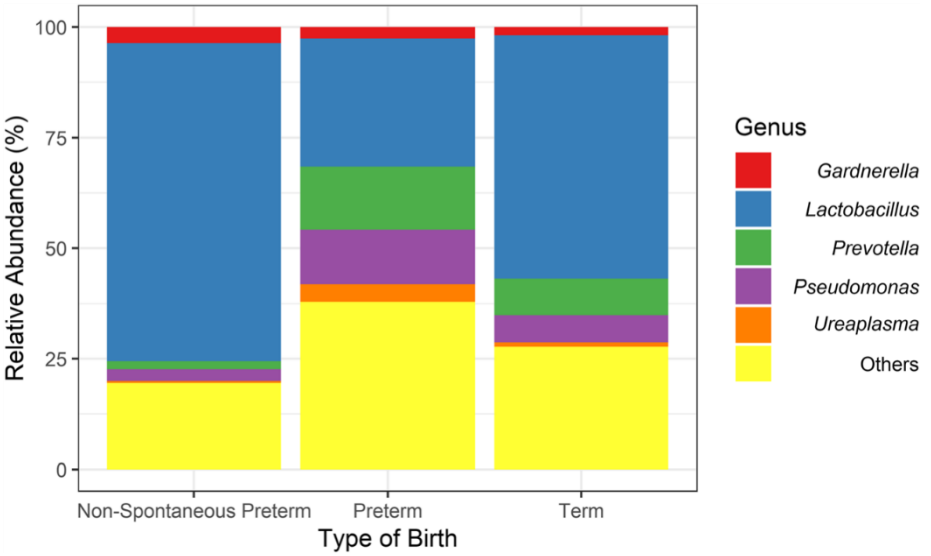


Fig. 1 Percentage of the five most abundant microbial genera found in the three tested groups. The genus Lactobacillus appeared with a low mean in preterm group when compared with term (Kruskal-Wallis post hoc Dunn test, p = 0.003) and with marginally low mean compared with non-spontaneous preterm (p = 0.080). Although *Prevotella* tended to be in high abundance in preterm samples, the test only found such difference when comparing preterm with non-spontaneous preterm group (p = 0.016). *Pseudomonas*, *Ureaplasma*, and *Gardnerella* did not present significant difference neither sample-to-sample or among samples (p > 0.05)

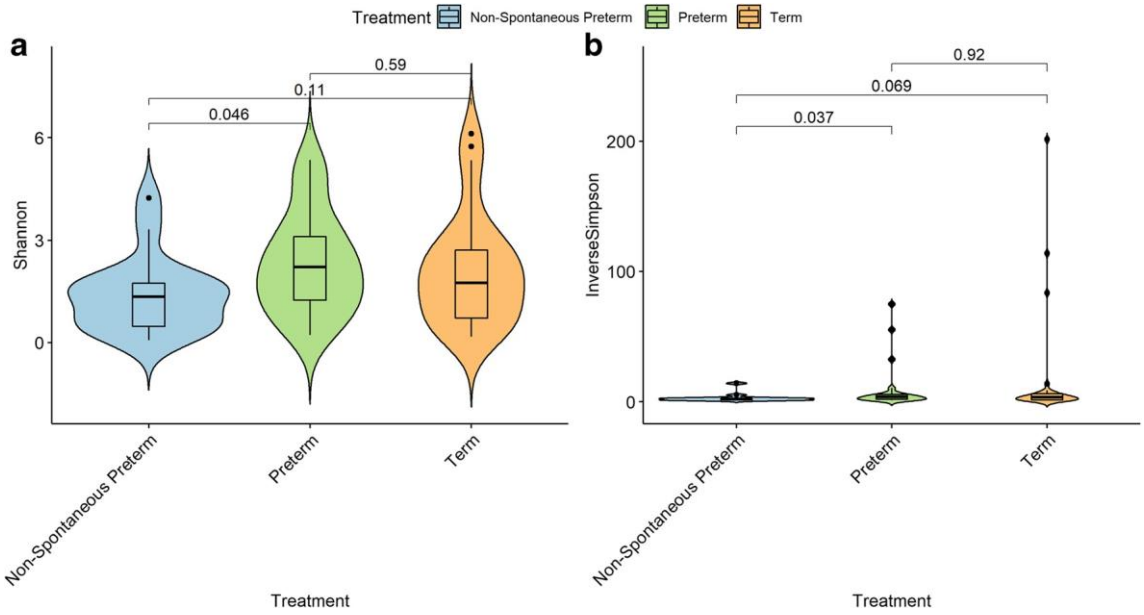


Fig. 2 Alpha diversity measurements of microbial communities in the spontaneous preterm labor and control groups. a Shannon diversity index. b Inverse of Simpson index. Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles, and the single circles indicate outliers. No significant difference was found among the three groups (p > 0.05) according to the non-parametric Wilcoxon test

Tables

Table 1 Maternal variables used for comparison between groups

Variables	Spontaneous term labor (n = 29)	Spontaneous preterm labor (n = 23)	Non-spontaneous preterm labor (n = 17)
Maternal age (years)	25.03 (± 1.13)[a]	23.60 (± 1.28)[a]	30.24 (± 1.75)[b]
Previous pregnancies	2.00 (± 0.17)[a]	1.87 (± 0.34)[a]	2.12 (± 0.28)[a]
Gestational age (weeks)	39.60 (± 0.20)[a]	30.70 (± 0.39)[b]	29.42 (± 0.59)[b]
Chorioamnionitis	0a	8b	2b
Preeclampsia	0a	3a	14[b]
GBS* infection	0a	5b	4b
Intrapartum penicillin	0a	15[b]	0a
Delivery mode (cesarean)	1a	10[b]	17[c]

*GBS, Group B Streptococcus. Numeric variables were summarized as average ± SEM and compared using the Kruskal-Wallis followed by a post hoc Dunn test. Categorical variables were compared using chi square post hoc test. Data followed by the same letter in the line represent groups without significant statistical difference ($p > 0.05$) whereas data followed by different letters in the line represent statistically different groups at the significance level of 95% ($p \leq 0.05$)

Table 2 Permutational analysis of variance (PERMANOVA) of the Bray-Curtis dissimilarities for bacterial OTU community structure used for detection of possible confounding variables associated with preterm labor

Confounding variables	R2	p value
Mother's age	0.029	0.482
Preview pregnancies	0.014	0.748
Gestational age	0.017	0.540
Corioamniotitis	0.011	0.895
Preeclampsia	0.030	0.106
GBS infection	0.013	0.734
Intrapartum penicillin	0.034	0.043
Delivery mode (cesarean/vaginal)	0.014	0.711

Significant value is set in italics. p values are based on 999 permutations

Table 3 Differential abundance analysis depicting vaginal microbial biomarkers associated with term or preterm labor

Base mean	Log_2-fold change	lfcSE	Stat	*p*-value	*p* adj	Closest microbial relative	Increased in
Spontaneous term versus spontaneous preterm							
22769	3.48	1.37	2.54	0.011	0.016	*Lactobacillus iners*	Spontaneous term labor
229	7.84	2.32	3.39	0.001	0.030	*Lactobacillus jensenii*	Spontaneous term labor
106	4.25	1.30	-3.27	0.001	0.031	*Prevotella* sp.	Spontaneous preterm labor
Spontaneous preterm versus non-spontaneous preterm							
229	7.40	2.46	3.00	0.003	0.045	*Lactobacillus jensenii*	Non-spontaneous preterm labor
218	2.90	2.90	-1.96	0.050	0.353	*Prevotella bivia*	Spontaneous preterm labor
106	5.03	5.03	-3.61	0.000	0.026	*Prevotella* sp.	Spontaneous preterm labor

Base mean, the average of the normalized counts taken over all samples; log2-fold change, log2 fold change between the groups; lfcSE, standard error of the log2-fold change; Stat, Wald statistic; p value; Wald test p value; p adj, FDR-adjusted p value

APÊNDICE E – Identificação de fenótipo de depressão usando variantes de sequência de *amplicon* de nucleotídeo único do microbioma intestinal humano

Depression phenotype identified by using single nucleotide exact amplicon sequence variants of the human gut microbiome

Bruce R. Stevens[1,2,3] • Luiz Roesch[4,5] • Priscila Thiago[4] • Jordan T. Russell[4] • Carl J. Pepine[6] • Richard C. Holbert[2] • Mohan K. Raizada[1] • Eric W. Triplett[4]

[1] Department of Physiology and Functional Genomics, University of Florida College of Medicine, Gainesville, FL, USA
[2] Department of Psychiatry, University of Florida College of Medicine, Gainesville, FL, USA
[3] Division of Gastroenterology, Department of Medicine, University of Florida College of Medicine, Gainesville, FL, USA
[4] Department of Microbiology and Cell Science, University of Florida, Gainesville, FL, USA
[5] Centro Interdisciplinar de Pesquisas em Biotecnologia-CIP-Biotec, Universidade Federal do Pampa, São Gabriel, Bagé, Brazil
[6] Division of Cardiovascular Medicine, Department of Medicine, University of Florida College of Medicine, Gainesville, FL, USA

Correspondent author: Bruce R. Stevens stevensb@ufl.edu

DOI: https://doi.org/10.1038/s41380-020-0652-5

Received: 25 July 2019
Revised: 13 January 2020
Accepted: 16 January 2020
Published online: 27 January 2020

Abstract

Single nucleotide exact amplicon sequence variants (ASV) of the human gut microbiome were used to evaluate if individuals with a depression phenotype (DEPR) could be identified from healthy reference subjects (NODEP). Microbial DNA in stool

samples obtained from 40 subjects were characterized using high throughput microbiome sequence data processed via DADA2 error correction combined with PIME machine-learning de-noising and taxa binning/parsing of prevalent ASVs at the single nucleotide level of resolution. Application of ALDEx2 differential abundance analysis with assessed effect sizes and stringent PICRUSt2 predicted metabolic pathways. This multivariate machine-learning approach significantly differentiated DEPR (n = 20) vs. NODEP (n = 20) (PERMANOVA $P < 0.001$) based on microbiome taxa clustering and neurocircuit-relevant metabolic pathway network analysis for GABA, butyrate, glutamate, monoamines, monosaturated fatty acids, and inflammasome components. Gut microbiome dysbiosis using ASV prevalence data may offer the diagnostic potential of using human metaorganism biomarkers to identify individuals with a depression phenotype.

Introduction

Depression is the leading cause of medical disability worldwide [1]. Reliably diagnosing individuals with depression, and understanding the biomedical and neurophysiological mechanisms underlying depression with its attending medical comorbidities, has the potential to radically transform diagnosis and treatment, substantially reduce disability-adjusted-life-year impacts on the economy and quality of life, and mitigate stigma.

Depression is associated with gut dysbiosis that disrupts a microbiome/systemic pathophysiology/brain bidirectional axis, based on preclinical rodent models and human studies [2–16]. Notwithstanding this basic dogma, recent Pubmed meta-analyses of human depression-specific gut microbiome studies [6, 12] have assessed that there is no consensus identifying the particular net gut microbial ecology of taxa and attending metabolomic interactions of microbiota with their hosts' physiology. This extends to lack of accord regarding differences in taxa relative abundances and diversity in depression, as compared with healthy reference subjects, although there is universal agreement that taxon significant differences do indeed exist. The literature is inconclusive regarding possible interplay between particular species and antidepressants use, experimentally limited to rodent models [16–20], although ketamine animal studies implicate a gut microbiome-associated anti-inflammatory aspect to the antidepressant mechanism of this drug [15]. Thus, understanding specific biological mechanisms, identification of microbiome taxa and functional pathway patterns as biomarkers, and development of microbiota-centric medical interventions have been confounded by the literature's wide variation in microbial compositional and abundance dataset analyses. These translational lapses hold for all the mental disorders, and particularly for depression [6].

Next generation sequencing has launched an expansion of microbiome studies, but often at the cost of data relevancy due to sequencing and pipeline shortcomings. The recent bioinformatics literature has challenged status quo microbiome analyses [21–24], establishing—yet it is not yet widely appreciated—that the commonly used high throughput sequencing methods and popular software yield compositional datasets of relative operational taxonomic unit (OTU) abundances are inherently flawed due to differences in 16S rRNA gene sequencing platforms and bioinformatics methodologies, many of which do not account for covariate effects.

Metagenomic shotgun sequencing approaches using composition-based, abundance-based, or combined hybrid binning analyses, can, in many respects, improve on the 16S OTU approach to differentiating cohorts, due to of feature specificity assigned to microbial metabolic pathways and genes [25–28]. Nevertheless, the specificity of closed-database metagenomics in general is not ideal for definitive biomarker identification nor microbiota/host interactive mechanisms, due to limitations incurred by the potential of sequencing errors and to confining microbial compositions within a defined catalog, often altering orthologous groups to conform to artificial enterotypes determined by Dirichlet analysis [29].

Inappropriate use of such compositional data has subsequently resulted in conflicting conclusions in the literature due to pipeline errors and OTU clustering methods. Publications are often widely quoted as authoritative major advances leading the field regarding the role of gut microbiome in mental disorders and comorbid health issues, yet their unchallenged compositional data analyses can be flawed thereby leading to deep propagation of misinformation in the literature [30, 31].

In the present study we avoid the dangers of closed database OTU clustering, and of the downside of shotgun sequencing metagenomics, by instead demonstrating a novel machine-learning pipeline that involves taxa binning and prevalences of exact amplicon sequence variants (ASV) [21]. Prevalence is defined as the proportion of individuals within a specific cohort who share an OTU or taxon at least once, regardless of abundance; it is a frequency of occurrence, in contrast to abundance which is the average fractional representation of a single OTU or taxon only when present. For example, using our (Prevalence Interval for Microbiome Evaluation (PIME) R package [32] a prevalence cutoff of 55% means that the taxa selected at this prevalence interval are found in 55% of the subjects.

We hypothesized that the prevalences of exact ASV within noisy gut microbiome data can readily identify pathophysiology-associated differences in gut signatures of microbial taxa and their metabolic pathway in people living with depression as compared with healthy reference subjects, at high resolution without resorting to OTU relative abundance clustering nor shotgun sequencing.

Materials and methods

Stool samples were obtained from 40 volunteer subjects, median age 34 y.o. distributed as groups of n = 20 subjects (ten females and ten males) meeting DSM-IV criteria for depression (DEPR) as outpatients seeking help at the University of Florida Department of Psychiatry, plus n = 20 healthy reference control subjects (14 females and 6 males) without a diagnosed mental illness (NODEP). Subjects refrained from probiotics and antibiotics during 30 days prior to providing stool samples, and they reported no gastrointestinal disorders. Fifteen of the 20 DEPR subjects had clinically charted antidepressants that included fluoxetine, venlafaxine, duloxetine, lamotrigine, mirtazapine, gabapentin, citalopram, sertraline, or bupropion, while five DEPR subjects had charts that did not specify a particular antidepressant medication by name. The possible effect of antidepressant medication use on metadata clustering was examined, but this resulted in no outliers of any significant differences that influenced data interpretation, as shown with statistical details in “Results” below; therefore, data from all 20 DEPR patients were pooled and included in all subsequent analyses

regardless of specific antidepressant or whether or not their chart specified an antidepressant. The University of Florida Institutional Review Board approved the study (clinicaltrials.govNCT02693327) for which participating subjects provided written informed consent and were remunerated.

Stool samples were collected with OMNIgeneGUT fecal collection kits (DNAGenotek, Ottawa, Ontario, Canada) and stored at −80 °C until DNA extraction. DNA from each sample was extracted from ~200 mg of stool using the E.Z. N.A Stool Extraction Kit following the manufacturer's protocol (Omega Bio-tek, Doraville, CA). Samples were randomized during extraction to avoid processing order bias, with the absence of processing and kit contamination verified by parallel blank negative controls. High DNA quality was determined by spectrophotometry. The region V3–V4 of 16S rRNA gene amplicons was amplified and sequenced with Illumina MiSeq (2 × 300 cycles run) as described previously [33].

The Illumina demultiplexed paired-end sequenced dataset was processed by the R package DADA2 [34] to correct for amplicon errors, to identify chimeras, and to merge paired-ends reads. The end product of DADA2 yielded a total of 2724 unique ASV each trimmed to 400 nucleotides in length, cataloged, and tallied as the number of times each exact ASV was observed for each sample. A phyloseq R object was generated comprised of all 2724 ASVs, a lookup table of taxonomy assignments to each ASVs obtained by using the naive Bayesian classifier method, and the SILVA ribosomal RNA gene database [35] version v132, plus subject sample metadata.

The DADA2-derived ASV dataset was initially analyzed by permutational analysis of variance (PERMANOVA) (Adonis R package), then the ALDEx2 R package [36, 37], as described below. This dataset was then treated using our PIME R package [32], which generated ASV prevalences by machine-learning, as validated by comparison with control Monte Carlo simulations with randomized variations of sequences. This prevalence-filtered dataset was then processed by ALDEx2, and output from the DADA2/PIME/ ALDEx2 workflow resulted in a denoised, filtered dataset comprised of 86 unique ASV sequences each 400 nucleotides in length (Supplementary Information Table 1-SI). PICRUSt2 [38] was employed for stringent predictions of functional metabolic pathways. As advantages over PICRUSt 1.0, the new PICRUSt2 pipeline inputs sequences as single nucleotide resolution ASVs, references ten times more genomes than PICRUSt1, and yields output as MetaCyc [39] pathway abundances referenced to shotgun metagenomics. ASV analyses included ALDEx2 differential abundances with Mann–Whitney and Bland–Altman plots, effects sizes distances, principal component analyses (PCA), and principal coordinates analyses (PCoA), network analysis, and pathway differences' odds ratios. Software included Python and R packages, run either standalone on Mac OS X or Calypso online [40].

Results

Initial comparison of microbial communities from DEPR vs. NODEPR using the entire unfiltered dataset of 2724 unique ASVs was assessed by multivariate PERMANOVA and Bray–Curtis distances, which yielded no significant difference ($p = 0.654$, $R^2 = 0.0289$). A multivariate PERMANOVA-like differential abundance analysis of the 2724 ASV DADA2 dataset was then assessed by employing the ALDEx2 R

package, resulting in the Mann–Whitney plot and Bland–Altman-type plot shown in Fig. 1a, b. According to the pattern and color of ASV points in Fig. 1a, b, no red prevalent data points and no significant differences between the groups were obtained using the full unfiltered 2724 ASV DADA2 dataset; the high number of black points indicates high relative abundance of taxa with low prevalence.

Subsequently, the noisy full dataset of 2724 ASVs was filtered using our PIME R package [32]. PIME removed the within-group variations and captured only biologically significant differences which have high sample prevalence levels. PIME employs a supervised machine-learning algorithm to predict random forests and estimates out-of-bag (OOB) errors, resulting in the Fig. 1c sets of ASV prevalence bins at 5% intervals. Here, high OOB errors indicate a given prevalence dataset bin is noisy representing a high relative abundance of taxa with low prevalence. Therefore in Fig. 1c the minimal OOB error = zero with the highest signal to noise ratio occurring within the 55% prevalence interval based on 292,798 sequence comparisons. This 55% prevalence dataset was comprised of 86 ASVs (Supplementary Information Table 1-SI) that were used for all subsequent downstream analyses.

Using the 55% prevalence 86 ASV dataset, OOB errors were predicted by a Monte Carlo simulation of random forest classifications by running 100 bootstrap aggregations on each prevalence interval. The simulation results shown in Fig. 1d matches Fig. 1c, thus reinforcing the appropriate choice of utilizing the 55% prevalence empirical dataset (Fig. 1c). In order to evaluate the likelihood of introducing bias while building the prevalence filtered datasets, the data were also randomly scrambled from the two subject groups and then run through the PIME error prediction algorithm again using 100 bootstraps. The resulting control randomization OBB errors were not significantly different from predicted value of 0.55 at all prevalence bins, as shown in Fig. 1e, confirming lack of false-positive type I errors. Thus, taken together the prediction simulation (Fig. 1d) and randomization control (Fig. 1e) simulation collectively validate our PIME algorithm [32] outcome of the 86 ASVs (Fig. 1c).

The 55% prevalence dataset of 86 ASVs was then reintroduced into ALDEx2 analysis, as shown in the Mann–Whitney (Fig. 1f) and Bland–Altman-type (Fig. 1g) outcome plots. Unlike Fig. 1a, b, the data in Fig. 1f, g appear as red points representing taxa assigned as differentially abundant at $q < 0.1$, along with non-differentially abundant gray points, but no black points that represent noise of high relative abundance taxa with low prevalence.

PCA [41] was executed using the unfiltered pre-PIME DADA2 dataset of 2724 ASVs and post-PIME-treated 86 ASVs. The pre-PIME 2724 sequence dataset could not be resolved into sample group clusters ($p = 0.238$, $R^2 = 0.02893$) as shown in Fig. 1h. In contrast, in Fig. 1i the post-PIME-treated 86 ASV 55% prevalence dataset yielded PERMANOVA (Adonis) Bray–Curtis $P < 0.001$, $R^2 = 0.531$, with PCA ordination readily resolved cluster differentiation of DEPR vs. NODEP, as shown with all points within the 95% confidence interval (CI) ellipses. These PCA results are consistent with the ADLEx2 Mann–Whitney and Bland–Altman results (Fig. 1f, g) and PERMANOVA described above.

The possible influence of antidepressant medication usage on taxa clustering of the 86 ASV dataset by the DEPR metadata phenotype was examined. This resulted in no outliers from the DEPR metadata by PCA analysis (PERMANOVA $P = 0.355$) (Fig. 1j), nor influence on clustering distances by Bray–Curtis dissimilarity network analyses ($P > 0.05$) (Fig. 1k). Therefore, data from all 20 DEPR patients were pooled and

included as a single cohort in subsequent analyses regardless of specific antidepressant or whether or not their chart specified an antidepressant. Potential influences of subject sex were also assessed; however, data parsed by male/female were not significantly different from random scrambling of sex ($P > 0.05$, data not shown).

Output from PIME/ALDEx2 yielded effect sizes for the 86 unique ASV sequences (Supplementary Information Table 2-SI), along with uncovering certain multiplicities of assigned taxa names assembled at the levels of Family, Genus, and species. The bar graph of Fig. 2a shows the taxa differential effect size values over the cutoff range of ≥ 0.5 for NODEP and ≤ -0.5 for DEPR. ASV sequences of Supplemental Table 2-SI were assigned a unique code used for downstream assessment of the 55% prevalence dataset, regardless of whether multiple sequences could be assigned the same taxonomic name, and then parsed for redundant multiple copies ($n > 1$) or unique ($n = 1$) taxonomic names assigned to the ASVs as shown in Fig. 2b, c. Note the large ASV prevalence representation from the Firmicutes phylum that occurs in both DEPR and NODEP, particularly in *Lachnospiraceae*, *Ruminococcaceae*, and *Bacteroidetes* Families. And also note the diversity of taxa names assigned to the prevalent ASVs is greater for DEPR than for NODEP, as corroborated by Chao1 and Shannon alpha diversity analyses (data not shown). The data of Fig. 2 are further discussed below in the Discussion.

PICRUSt2 treatment of the 86 ASVs from the 55% prevalence dataset predicted 284 MetaCyc microbiome metabolic pathways. Fig. 3a shows the network analysis revealing clustering of gut microbiome pathways common to DEPR, as segregated from clustering common to healthy control NODEP. The possible influence of antidepressant medication use on metabolic pathway clustering by the DEPR metadata phenotype was examined in Fig. 3a, resulting in no outliers within the DEPR cluster by Bray–Curtis dissimilarity network analyses ($P > 0.05$). Therefore, pathway data from all 20 DEPR patients were pooled and considered as a single cohort (purple circles in Fig. 3a) regardless of specific antidepressant (assigned red in Fig. 3a) or whether or not their chart specified an antidepressant (assigned blue in Fig. 3a). Based on PICRUSt2 pathway data of Fig. 3a, odds ratios and AUC were generated, with the top most salient pathway differences shown in the forest plot of Fig. 3b. Notably, these data highlight untoward pathways in subjects living with depression as contrasted with healthy pathways in reference subjects without depression.

Discussion

The machine-learning pipeline of the present study unmasked a novel and useful pattern of gut microbiota taxa variants' prevalences and functionally relevant metabolic pathways associated with depression, as compared with healthy reference subjects. This is the first implementation of advanced error suppression at the level of single nucleotide resolution in compositional gut microbiome assignment to a major mental disorder and attending functional pathways phenotype, via a unique pipeline that incorporates DADA2 error correction [34] combined with de-noising and taxa binning of exact ASV prevalences generated by the R package PIME [32], followed by differential abundance analysis with effect sizes via ALDEx2 [36, 37]. The precedent

for justifying exact sequences to differentiate microbial populations without OTU clustering has been established by large scale population diversity studies [21–24].

Functional metabolic pathways

It is becoming recognized in the literature that the value of microbiome studies lies in the importance of the interactive ecology of microbial intermediary metabolism pathways over differences in taxonomy [42]. Patterns of metabolic pathways in gut microbiota taxa reflect their impact on distinguishing host physiology phenotypes, due to the interplay of microbial metabolism with host metabolome and physiology. The present ASV approach revealed the differentiation of untoward gut microbiota MetaCyc pathways in DEPR, in contrast to healthy pathways in NODEP reference subjects (Fig. 3a, b). These results are consistent with the distinguishing hallmarks of unfavorable shifts in metabolism and host inflammasome dysregulation associated with disruption of gut barrier reported for depression vs. healthy individuals, as we and others [6, 8, 11, 12, 43–47] have reported previously.

In the present results (Fig. 3) DEPR-associated pathways of butyrate degradation and GABA degradation were prominent in the gut microbiome of DEPR. In contrast, NODEP was prominently represented by multiple *Lachnospiraceae NK4A136* and *Lachnospiraceae UCG-001* ASVs (Fig. 2, Table 2-SI), which represent butyrate producing species associated with the physiological and behavioral health benefits of this short chain fatty acid (SCFA) [10, 13, 48, 49]. In a mouse model of depression, *Lachnospiraceae UCG-001* and *Lachnospiraceae NK4A136* abundances were significantly enhanced by the antidepressant fluoxetine in a subgroup of mice that behaviorally responded to fluoxetine, but these species were not enhanced in the subgroup that did not respond to antidepressant treatment [19]. Our pipeline effect size results (Fig. 2, Table 2-SI) also indicated high prevalences of *Roseburia* spp, *Bacteroides* spp, *Faecalibacterium* spp—in particular, *Faecalibacterium prausnitzii*—in both DEPR and NODEP, explainable by the notion that SCFA metabolism is highly strain-specific and diet-dependent [20, 50]. *Faecalibacterium prausnitzii* is the single most common human gut bacterium, with relative abundance dependent on prebiotic diet composition [50, 51].

Inflammation and dysregulation of glutamate, monoamines, and GABA neurotransmission have been associated with the pathophysiology of depression and comorbid neuropathic pain [6, 13, 52–54]. The results (Fig. 2, Table 2-SI) indicated a mixture of taxa representing species that have potential for GABA production (*Parabaceroidetes merdre* and certain strains of *Alstipes* spp, *Bacteriodes* spp, *Eubacterium* spp, and *Escherichia* spp) or GABA consumption (select strains of *Flavonifractor plautii*, *Pseudomonas* spp, and *Acinetobacter* spp) [9, 12], with somewhat greater prevalence of GABA producing taxa in DEPR compared to NODEP. Indeed, the odds ratio data of Fig. 3b favor subjects with depression possessing gut microbiota GABA degradation and reduced biosynthesis via: “L_arginine_putrescine_and_GABA_degradation_superpathway”, L_arginine_and_ornithine_degradation_superpathway”, and “L_argi nine_degradation_AST_superpathway”. Conversely, in healthy reference subjects without depression the Fig. 3b odds ratio favored “L_glutamine_and_glutamate_biosynthesis” which promotes GABA production.

The data of Fig. 3 reflect microbiota alterations in small molecules and other amino acid pathways associated with depression, such that threonine, tryptophan, and leucine that can activate mTOR-mediated intestinal inflammation, while arginine and ornithine suppress gut inflammation [55]. Under proinflammatory gut bacteria conditions, high levels of tryptophan are converted to kynurenine at the expense of reduced serotonin synthesis, whereupon kynurenine crosses blood–brain barrier and is converted to neurotoxic kynurenic and quinolinic acids which have been correlated with depressive symptoms [56]. Microbial degradation of allantoin in DEPR, and D-galactuonate degradation NODEP in our subjects are consistent with animal models [57, 58].

Proinflammatory gut bacteria that generate Kdo2-lipid were favored in DEPR (Fig. 2). Kdo2-lipids are the primary component of LPS that activates host TLR4-MD2 signaling and myeloid differentiation [59], enterobacterial common antigen that is linked to LPS via Kdo2 [60], and iron sequestering biofilm-enhancing enterobactin [61]. The large effect size for Enterobacteriaceae in DEPR (Fig. 2a ,Table 2-S1) is consistent with LPS-related morbidity of strains that disrupt intestinal barrier and invokes inflammation from proinflammatory cytokine responses [50] in humans with depression [62]. The high ASV prevalence of *Roseburia intestinalis* in our NODEP subjects (Fig. 2) is consistent with prior studies showing that the flagellin of this species reduces intestinal inflammation by suppressing IL-17 in the host [63]. The pathway data (Fig. 3) favored enhancement of oleate and palmitoleate, which are inversely correlated with depression [64, 65]. Overall gut microbiota fatty acid beta oxidation was favored in DEPR (Fig. 3b).

Taxa

In addition to functional pathway differences, assigned taxa names and taxa linear discriminant analysis estimates of effect size differences between depressed human subjects vs. healthy control subjects have been used in previous gut microbiome compositional studies based on OTU relative abundances or shotgun sequencing metagenomics [6, 8–10, 44, 47, 48, 66–68]. Our ASV results (Figs. 1j, k and 2a–c, and Table 2-SI) are in accord with the reported general trend in increased OTU relative abundances of taxa associated with human depression and rodent models of depression with respect to *Acidaminococcaceae*, Enterobacteriaceae, *Rikenellaceae*, and *Coriobacteriaceae* Families, and in particular of *Blautia* sp, *Alistipes* sp, *Parabacteroides* spp, *Phascolarctobacterium* sp, *Oscillibacter* spp, *Rosburia* spp, *Flavonifractor* sp, and *Holdemania* sp [10, 11, 44, 45, 48]. Regarding trends for OTU relative abundances depleted in DEPR and increased in NODEP, select species of *Faecalibacterium* spp, *Ruminococcus* spp, *Lachnospiraceae* spp, and *Bacteroides* spp have been reported [9, 10, 44, 48], as also observed in our ASV prevalence results (Fig. 2, Table 2-SI). Previous reports have negatively correlated *Faecalibacterium* spp OTU abundances with magnitude of depression symptom severity [10, 69], while our results (Fig. 2, Table 2-SI) identified separate particular *Faecalibacterium* spp ASVs for DEPR and NODEP. Elevated *Parabacteroidetes* is associated with anhedonia in rat models [70], consistent with our results (Fig. 2a–c, Table 2-SI).

Fig. 2 and Table 2-SI indicated that in NODEP nearly 75% of the prevalent taxa with effect sizes >0.5 are *Lachnospiraceae* spp, with the balance of prevalent taxa

represented by *Bacteroides* spp and *Ruminococcaceae* family. The top prevalence effect size for NODEP (Fig. 2a) was *Faecalibacterium CM04-06* of the *Ruminococcaceae* family. Both DEPR and NODEP yielded large prevalences of ASVs tagged to *Ruminococcaceae*. This is not unexpected because beneficial vs untoward health effects of *Ruminococcaceae* are highly species- and strain-specific and diet-induced, due to variations in their fiber hydrolyzing enzyme profiles [20, 50]. Strain-dependence abundances in DEPR is reportedly associated with elevated *R. flavefaciens* which abrogates effects of antidepressants [18], and elevated select members of *Bacteroidetes* phylum, but with net decreases in *Firmicutes* in both human studies and in rat depression models employing relative OTU abundances alone [45] or in conjunction with LC/MS metabolomics [11]. Getachew et al. [15] reported that antidepressant ketamine reduced levels of *Ruminococcus* spp in rats. These OTU reports are in contrast to the high degree of representation of ASV assigned to Lachnospiraceae, *Ruminococcaceae*, and *Bacteroidaceae* in both DEPR and NODEP (Fig. 2, Table 2-SI).

Our ASV prevalence data (Fig. 2, Table 2-SI) indicated an overall greater diversity of taxa prevalences in DEPR, in concordance with a 16S closed-OTU approach of Jiang et al. [10], but in contrast to other reports of alpha diversity or richness with no difference in humans [44] or reduction with a rat depression model [45]. It has been posited that the many dimensions of “diversity” of a given ecosystem composition are not per se an index of “better” vs. “worse” [71].

Physiological anthropology of depression

Mood disorders exhibit familial transmission, but the exact genetics remain unresolved despite ongoing studies analyzing nearly 200 candidate human maker genes [43]. People are essentially metaorganisms comprised of ~10^{14} prokaryotic cells plus roughly the same number of eukaryotic cells—host physiology is the co-evolutionary consequence of the interplay among human plus bacterial genomes and metabolomics. The present study emphasizes the importance of gut microbiome prevalences on host depression phenotype behavior and metabolic pathways; it is the prevalence—in contrast to relative abundance—of certain bacterial metabolic functions steered by microbiome genes that ultimately shapes host-microbiota relationships. Specific human genetic loci shape heritable patterns of gut microbiome taxa prevalences in the host [72]. The *Christensenellaceae* family is the single most heritable gut microbial taxon, typically correlated with various healthy phenotypes [72]. Notably, our results (Table 2-SI) indicated ASV prevalence of *Christensenellaceae* R-7 group in the NODEP cohort (effect size 0.47), in contrast to DEPR. These data taken together with PCA discriminatory taxa clustering of NODEP subjects vs. DEPR (Fig. 1i–k) collectively imply the heritability of resistance to depression. Thus, anthropological group cohesion cultural factors such as food and dietary habits, mating preferences that sustain closed groups, and environmental communal exposure to a common set of commensal bacteria may propagate bacterial species of depression. Or conversely, perhaps certain gut microbiota may have usurped human hosts as unwitting prokaryotic propagation vessels by shaping mood and sickness behavior as an evolutionary survival advantage by withdrawing their hosts from environmental harms or from competing infectious pathogenic bacteria.

Conclusion

In conclusion, the present study describes a novel gut microbiome machine-learning approach to potentially differentiate people with depression from healthy reference controls. The process employs DADA2, ALDEx2, PIME, and PICRUSt2 R packages to evaluate prevalent ASVs from human gut microbiome 16S rRNA amplicon sequences at the level of single nucleotide resolution. This machine-learning technique approach may reduce pitfalls of OTU relative abundance clustering and shotgun metagenomics. By employing prevalent ASVs, this study led to an ability to distinguish 20 individuals with depression from 20 healthy reference subjects. Results are supported by multivariate analyses' PERMANOVA $P < 0.001$, effect sizes ≥ 0.5, PCA ordination, network analyses, and odds ratios, which collectively conformed to current dysbiosis and pathophysiologic hypotheses of depression associated with neurocircuit-relevant neurotransmitter pathways, inflammation, and gut–brain dysregulation. Furthermore, the differential patterns of unique ASVs assigned to taxonomy and metabolic pathways associated in individuals with depression and healthy controls were generally consistent with prior OTU relative abundance and metagenomics studies, with disparities attributable to the taxonomic MetaCyc assignment of species- and strain-specific microbiota metabolomics. This is the first published report using this gut microbiome machine-learning approach and its utility as a high throughput sequencing technique of the gut microbiome to identify individuals with depressive symptoms different from healthy reference subjects. Its application in the clinical setting may lend to personalizing treatments based on ASV in patients with depression by decreasing neurobiological heterogeneity, as based on the current DSM-5 diagnostic framework. Larger studies are needed to delineate the extent to which different symptoms of depression and influences of antidepressants may correspond with functional metaorganisms tethered to underlying neurobiological dysfunction.

Code availability

PIME is available at https://rdrr.io/github/microEcology/ pime/; DADA2 is available at https://www.bioconductor. org/packages/release/bioc/html/dada2.html; ALDEx2 is at https://bioconductor.org/packages/release/bioc/html/ ALDEx2.html; and picrust2 is at https://github.com/ picrust/picrust2.

Acknowledgements We thank Seungbum Kim and Elaine Richards for useful discussions. This study was supported by University of Florida Clinical and Translational Science Institute grant (BRS, CJP) from the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR, 427; National Institute of Health (NIH) grants HL33610, HL56921 (MKR, CJP); UM1 HL087366 (CJP); Gatorade Trust through funds distributed by the University of Florida, Department of Medicine (CJP); PCORI-OneFlorida Clinical Research Consortium CDRN1501–26692 (CJP); and internal funds from the University of Florida Department of Physiology and Functional Genomics (BRS, MKR).

References

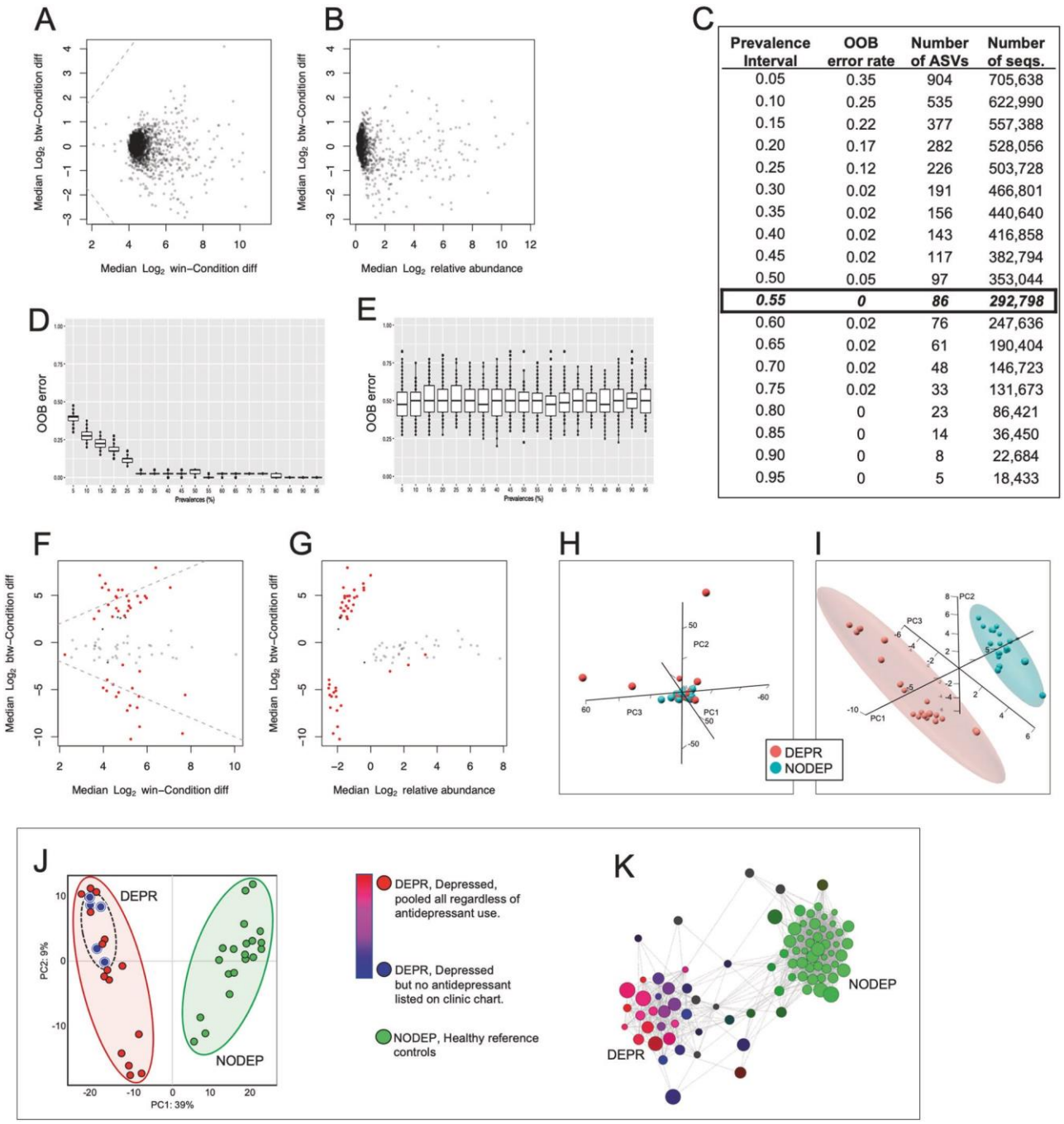
1. WHO. World Health Organization Fact Sheet, Depression. 2018. https://www.who.int/news-room/fact-sheets/detail/depression.
2. Heiss CN, Olofsson LE. The role of the gut microbiota indevelopment, function and disorders of the central nervous system and the enteric nervous system. J Neuroendocrinol. 2019;31: e12684.
3. Warner BB. The contribution of the gut microbiome to neurodevelopment and neuropsychiatric disorders. Pediatr Res. 2019;85:216–24.
4. Stefano GB, Pilonis N, Ptacek R, Raboch J, Vnukova M, KreamRM. Gut, microbiome, and brain regulatory axis: relevance to neurodegenerative and psychiatric disorders. Cell Mol Neurobiol. 2018;38:1197–206.
5. Skonieczna-Zydecka K, Marlicz W, Misera A, Koulaouzidis A,Loniewski I. Microbiome-the missing link in the gut-brain axis: focus on its role in gastrointestinal and mental health. J Clin Med. 2018;7:E521.
6. Winter G, Hart RA, Charlesworth RPG, Sharpley CF. Gutmicrobiome and depression: what we know and what we need to know. Rev Neurosci. 2018;29:629–43.
7. Kelly JR, Clarke G, Cryan JF, Dinan TG. Brain-gut-microbiotaaxis: challenges for translation in psychiatry. Ann Epidemiol. 2016;26:366–72.
8. Zalar B, Haslberger A, Peterlin B. The role of microbiota indepression—a brief review. Psychiatr Danub. 2018;30:136–41.
9. Strandwitz P, Kim KH, Terekhova D, Liu JK, Sharma A,Levering J, et al. GABA-modulating bacteria of the human gut microbiota. Nat Microbiol. 2018;4:396–403.
10. Jiang H, Ling Z, Zhang Y, Mao H, Ma Z, Yin Y, et al. Alteredfecal microbiota composition in patients with major depressive disorder. Brain Behav Immun. 2015;48:186–94.
11. Yu M, Jia H, Zhou C, Yang Y, Zhao Y, Yang M, et al. Variationsin gut microbiota and fecal metabolic phenotype associated with depression by 16S rRNA gene sequencing and LC/MS-based metabolomics. J Pharm Biomed Anal. 2017;138:231–9.
12. Cheung SG, Goldenthal AR, Uhlemann AC, Mann JJ, Miller JM,Sublette ME. Systematic review of gut microbiota and major depression. Front Psychiatr. 2019;10:34.
13. Vuong HE, Yano JM, Fung TC, Hsiao EY. The Microbiome andHost Behavior. Annu Rev Neurosci. 2017;40:21–49.
14. Pearson-Leary J, Zhao C, Bittinger K, Eacret D, Luz S, Vigderman AS, et al. The gut microbiome regulates the increases in depressive-type behaviors and in inflammatory processes in the ventral hippocampus of stress vulnerable rats. Mol Psychiatry. 2019. https://doi.org/10.1038/s41380-019-0380-x. [Epub ahead of print].
15. Getachew B, Aubee JI, Schottenfeld RS, Csoka AB, ThompsonKM, Tizabi Y. Ketamine interactions with gut-microbiota in rats: relevance to its antidepressant and anti-inflammatory properties. BMC Microbiol. 2018;18:222.
16. Sun L, Zhang H, Cao Y, Wang C, Zhao C, Wang H, et al.
Fluoxetine ameliorates dysbiosis in a depression model induced by chronic unpredicted mild stress in mice. Int J Med Sci. 2019;16:1260–70.
17. Romijn AR, Rucklidge JJ, Kuijer RG, Frampton C. A doubleblind, randomized, placebo-controlled trial of Lactobacillus helveticus and Bifidobacterium longum for the symptoms of depression. Aust N. Z J Psychiatr. 2017;51:810–21.
18. Lukic I, Getselter D, Ziv O, Oron O, Reuveni E, Koren O, et al.Antidepressants affect gut microbiota and Ruminococcus flavefaciens is able to abolish their effects on depressive-like behavior. Transl Psychiatr. 2019;9:133.
19. Lyte M, Daniels KM, Schmitz-Esser S. Fluoxetine-inducedalteration of murine gut microbial community structure: evidence for a microbial endocrinology-based mechanism of action responsible for fluoxetine-induced side effects. PeerJ 2019;7: e6199.
20. Hassan AM, Mancano G, Kashofer K, Frohlich EE, Matak A,Mayerhofer R, et al. High-fat diet induces depression-like behaviour in mice associated with changes in microbiome, neuropeptide Y, and brain metabolome. Nutr Neurosci. 2019;22:877–93.
21. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variantsshould replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43.

22. Lovell D, Pawlowsky-Glahn V, Egozcue JJ, Marguerat S, BahlerJ. Proportionality: a valid alternative to correlation for relative data. PLoS Comput Biol. 2015;11:e1004075.
23. Glassman SI, Martiny JBH. Broadscale ecological patterns arerobust to use of exact sequence variants versus operational taxonomic units. Msphere. 2018;3:e00148–18.
24. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J,
Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nature. 2017;551:457–63.
25. Garrido-Cardenas JA, Manzano-Agugliaro F. The metagenomicsworldwide research. Curr Genet. 2017;63:819–29.
26. Cao HT, Gibson TE, Bashan A, Liu YY. Inferring humanmicrobial dynamics from temporal metagenomics data: Pitfalls and lessons. Bioessays. 2016;39. https://doi.org/10.1002/bies. 201600188. Epub online.
27. Sedlar K, Kupkova K, Provaznik I. Bioinformatics strategies fortaxonomy independent binning and visualization of sequences in shotgun metagenomics. Comput Struct Biotechnol J. 2017;15: 48–55.
28. Franzosa EA, Huang K, Meadow JF, Gevers D, Lemon KP,Bohannan BJ, et al. Identifying personal microbiomes using metagenomic codes. Proc Natl Acad Sci USA. 2015;112: E2930–8.
29. Tang ZZ, Chen G. Zero-inflated generalized Dirichlet multinomial regression model for microbiome compositional data analysis. Biostatistics. 2018;20:698–713.
30. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ.Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
31. Gloor GB, Reid G. Compositional analysis: a valid approach toanalyze microbiome high-throughput sequencing data. Can J Microbiol. 2016;62:692–703.
32. Roesch LFW, Dobbler PT, Pylro VS, Kolaczkowski B, Drew JC,Triplett EW. PIME: A package for discovery of novel differences among microbial communities. Mol Ecol Resour. 2019. https:// doi.org/10.1111/1755-0998.13116. [Epub ahead of print].
33. Davis-Richardson AG, Ardissone AN, Dias R, Simell V, LeonardMT, Kemppainen KM, et al. Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes. Front Microbiol. 2014;5:678.
34. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ,Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
35. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P,et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
36. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, EdgellDR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome. 2014;2:15.
37. Gloor GB. ALDEx2: ANOVA-like differential expression tool forcompositional data. 2018. https://bioconductor.org/packages/relea se/bioc/vignettes/ALDEx2/inst/doc/ALDEx2_vignette.pdf.
38. Douglas GM, Beiko RG, Langille MGI. Predicting the functionalpotential of the microbiome from marker genes using PICRUSt. Methods Mol Biol. 2018;1849:169–77.
39. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A,Krummenacker M, et al. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2018;46:D633–D39.
40. Zakrzewski M, Proietti C, Ellis JJ, Hasan S, Brion MJ, Berger B,et al. Calypso: a user-friendly web-server for mining and visualizing microbiome-environment interactions. Bioinformatics. 2017;33:782–3.
41. R Core Team. R: A language and environment for statisticalcomputing. R Foundation for Statistical Computing, Vienna, Austria. 2018. https://www.R-project.org/.
42. Visconti A, Le Roy CI, Rosa F, Rossi N, Martin TC, Mohney RP,et al. Interplay between the human gut microbiome and host metabolism. Nat Commun. 2019;10:4505.
43. Shadrina M, Bondarenko EA, Slominsky PA. Genetics factors inmajor depression disease. Front Psychiatr. 2018;9:334.
44. Naseribafrouei A, Hestad K, Avershina E, Sekelja M, LinlokkenA, Wilson R, et al. Correlation between the human fecal microbiota and depression. Neurogastroenterol Motil. 2014;26: 1155–62.
45. Kelly JR, Borre Y, Patterson COB, El Aidy E, Deane J S, et al.Transferring the blues: depression-associated gut microbiota induces neurobehavioural changes in the rat. J Psychiatr Res. 2016;82:109–18.

46. Johansson R, Carlbring P, Heedman A, Paxling B, Andersson G.Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. PeerJ. 2013;1:e98.
47. Editoriaol. Links between gut microbes and depression strengthened. Nature. 2019;566:7.
48. Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J,Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol. 2019;4:623–32.
49. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, et al. Gutmicrobiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatr. 2016;21:786–96.
50. Hiippala K, Jouhten H, Ronkainen A, Hartikainen A, KainulainenV, Jalanka J, et al. The potential of gut commensals in reinforcing intestinal barrier function and alleviating inflammation. Nutrients. 2018;10:E988.
51. Miquel S, Martin R, Rossi O, Bermudez-Humaran LG, Chatel JM,Sokol H, et al. Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol. 2013;16:255–61.
52. Peirce JM, Alvina K. The role of inflammation and the gut microbiome in depression and anxiety. J Neurosci Res.
2019;97:1223–41.
53. Liu CS, Adibfar A, Herrmann N, Gallagher D, Lanctot KL. Evidence for Inflammation-Associated Depression. Curr Top Behav Neurosci. 2017;31:3–30.
54. Halaris A. Inflammation and depression but where does the inflammation come from? Curr Opin Psychiatr. 2019;32:422–8.
55. He F, Wu C, Li P, Li N, Zhang D, Zhu Q, et al. Functions andsignaling pathways of amino acids in intestinal inflammation. Biomed Res Int. 2018;2018:9171905.
56. Waclawikova B, El Aidy S. Role of microbiota and tryptophanmetabolites in the remote effect of intestinal inflammation on brain and depression. Pharmaceuticals. 2018;11:E63.
57. Ahn YJ, Park SJ, Woo H, Lee HE, Kim HJ, Kwon G, et al. Effectsof allantoin on cognitive function and hippocampal neurogenesis. Food Chem Toxicol. 2014;64:210–6.
58. Chung WSF, Meijerink M, Zeuner B, Holck J, Louis P, Meyer
AS, et al. Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. FEMS Microbiol Ecol. 2017;93. https://doi.org/10.1093/femsec/fix127.
59. Wang XY, Quinn PJ, Yan AX. Kdo(2)-lipid a: structural diversityand impact on immunopharmacology. Biol Rev 2015;90:408–27.
60. Mitchell AM, Srikumar T, Silhavy TJ. Cyclic enterobacterialcommon antigen maintains the outer membrane permeability barrier of Escherichia coli in a manner controlled by YhdP. MBio 2018;9:e01321–18.
61. Wang H, Zeng X, Mo Y, He B, Lin H, Lin J. Enterobactin-specific antibodies induced by a novel enterobactin conjugate vaccine. Appl Environ Microbiol. 2019;85:e00358–19.
62. Stevens BR, Goel R, Seungbum K, Richards EM, Holbert RC,Pepine CJ, et al. Increased human intestinal barrier permeability plasma biomarkers zonulin and FABP2 correlated with plasma LPS and altered gut microbiome in anxiety or depression. Gut. 2018;67:1555–7.
63. Zhu C, Song K, Shen Z, Quan Y, Tan B, Luo W, et al. Roseburia intestinalis inhibits interleukin17 excretion and promotes regulatory T cells differentiation in colitis. Mol Med Rep. 2018;17:7567–74.
64. Wolfe AR, Ogbonna EM, Lim S, Li Y, Zhang J. Dietary linoleicand oleic fatty acids in relation to severe depressed mood: 10 years follow-up of a national cohort. Prog Neuropsychopharmacol Biol Psychiatry. 2009;33:972–7.
65. Fernandes MF, Mutch DM, Leri F. The relationship between fattyacids and different depression-related brain regions, and their potential role as biomarkers of response to antidepressants. Nutrients. 2017;9:E298.
66. Bailey MT, Dowd SE, Galley JD, Hufnagle AR, Allen RG, LyteM. Exposure to a social stressor alters the structure of the intestinal microbiota: implications for stressor-induced immunomodulation. Brain Behav Immun. 2011;25:397–407.
67. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF,Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352:565–9.

68. Goll J, Thiagarajan M, Abubucker S, Huttenhower C, Yooseph S,Methe BA. A case study for large-scale human microbiome analysis using JCVI's metagenomics reports (METAREP). PLoS One. 2012;7:e29044.
69. Evans SJ, Bassis CM, Hein R, Assari S, Flowers SA, Kelly MB,et al. The gut microbiome composition associates with bipolar disorder and illness severity. J Psychiatr Res. 2017;87:23–9.
70. Yang C, Fang X, Zhan G, Huang N, Li S, Bi J, et al. Key role ofgut microbiota in anhedonia-like phenotype in rodents with neuropathic pain. Transl Psychiatr. 2019;9:57.
71. Shade A. Diversity is the question, not the answer. ISME J.2017;11:1–6.
72. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159:789–99.

Figures



Prevalence Interval	OOB error rate	Number of ASVs	Number of seqs.
0.05	0.35	904	705,638
0.10	0.25	535	622,990
0.15	0.22	377	557,388
0.20	0.17	282	528,056
0.25	0.12	226	503,728
0.30	0.02	191	466,801
0.35	0.02	156	440,640
0.40	0.02	143	416,858
0.45	0.02	117	382,794
0.50	0.05	97	353,044
0.55	***0***	***86***	***292,798***
0.60	0.02	76	247,636
0.65	0.02	61	190,404
0.70	0.02	48	146,723
0.75	0.02	33	131,673
0.80	0	23	86,421
0.85	0	14	36,450
0.90	0	8	22,684
0.95	0	5	18,433

Fig. 1 ASV prevalences and metadata analyses. **a** Mann–Whitney plot of ALDEx2 output using complete DADA2 dataset of 2724 Illumina demultiplexed paired-end unique DNA sequences (Supplementary Information Table 1-SI) without PIME treatment. **b** Bland–Altman MA-type plot of same dataset described for **a**. **c** PIME prevalence bins and out-of-bag (OOB) errors for intervals 5–95%. Based on the criterion of greatest number of random forest sequence combinations at the minimal OBB error = zero, the 55% interval with 86 AVSs (listed in Supplementary Information Table 2-SI) was ultimately employed for all subsequent downstream analyses. **d** PIME OOB error predictions at each prevalence interval, showing box plots of treatments randomly assigned to the 86 ASV dataset samples. These simulated predictions match the empirical data of **c**. **e** Scrambled control validation of the PIME simulations of **d**, assessed by running randomized variations of OOB errors for each prevalence bin, resulting in box plots with OOB errors at all bins hovering at the predicted value of 0.55. Each plot in **d** and **e** was generated by machine-learning using the 55% prevalence dataset of 86 ASVs run through

100 bootstrap iterations of Monte Carlo simulations of random forest classifications on each prevalence interval in 5% increments. **f** Mann–Whitney plot of pipeline output from DADA2 that derived from the PIME 55% prevalence dataset of 86 ASV seqs, then subsequently processed with ALDEx2. **g** Bland–Altman MA-type plot of same dataset described for **d**. **a**, **b**, **f**, and **g** each point represents a unique microbial taxon exact amplicon sequence variant (ASV). Red points represent taxa assigned as differentially abundant at $q < 0.1$; gray points are abundant, but not nondifferentially abundant; black points are rare and not differentially abundant. Based on **f** and **g**, the 55% interval with 86 AVSs was ultimately employed for all subsequent downstream analyses. **h** Principal component analysis (PCA) of entire 2724 sequence dataset from the pipeline of DADA2 plus ALDEx2 but not treated using PIME, revealing no metadata clustering by PERMANOVA (Adonis) Bray–Curtis $p = 0.654$, $R^2 = 0.02893$. **i** PCA of the 55% prevalence dataset of highly prevalent 86 ASVs from the complete DADA2/ ALDEx2/PIME pipeline, revealing significant clustering of metadata by PERMANOVA (Adonis) Bray–Curtis with $P < 0.001$, $R^2 = 0.531$, with group clusters shown within predicted 95% CI ellipses. Each dataset in **h**, **i** was log2 transformed, centered and scaled, and run with Bray–Curtis distances with the pca3d subroutine of the prcomp R package [41]. **j** Lack of antidepressant influence on DEPR cohort taxa clustering by PCA. Metadata were assigned as: five DEPR subjects with no specific antidepressant listed on their chart (blue), 20 DEPR subjects pooled regardless of antidepressant use (red; 15 DEPR subjects with charted use of an antidepressant plus five DEPR with no specifically listed antidepressant), or 20 NODEP subjects (green). There were no outliers from the clustered 20 subject pooled DEPR cohort (PERMANOVA $P = 0.355$) which were collectively isolated from NODEP ($P < 0.001$) relating to the 86 prevalent taxa. **k** Bray–Curtis dissimilarity network analysis of 86 prevalent taxa and lack of antidepressant influence on cohort distances. Pearson correlation algorithm was employed with positive taxa nodes placed with dissimilarity ordination distances connected by principal coordinates analysis (PCoA) edge placement (false discovery rate, FDR $P < 0.05$), with similarity cutoff at 0.25. Node sizes and colors are proportional to relative magnitude within the dataset. Note taxa clustering and purple color blend resulting from the overlay of DEPR subjects whose charts listing an antidepressant (red) on DEPR subjects whose charts did not list a specific antidepressant (blue), and of which pooled DEPR metadata were collectively segregated from clustered NODEP (green).

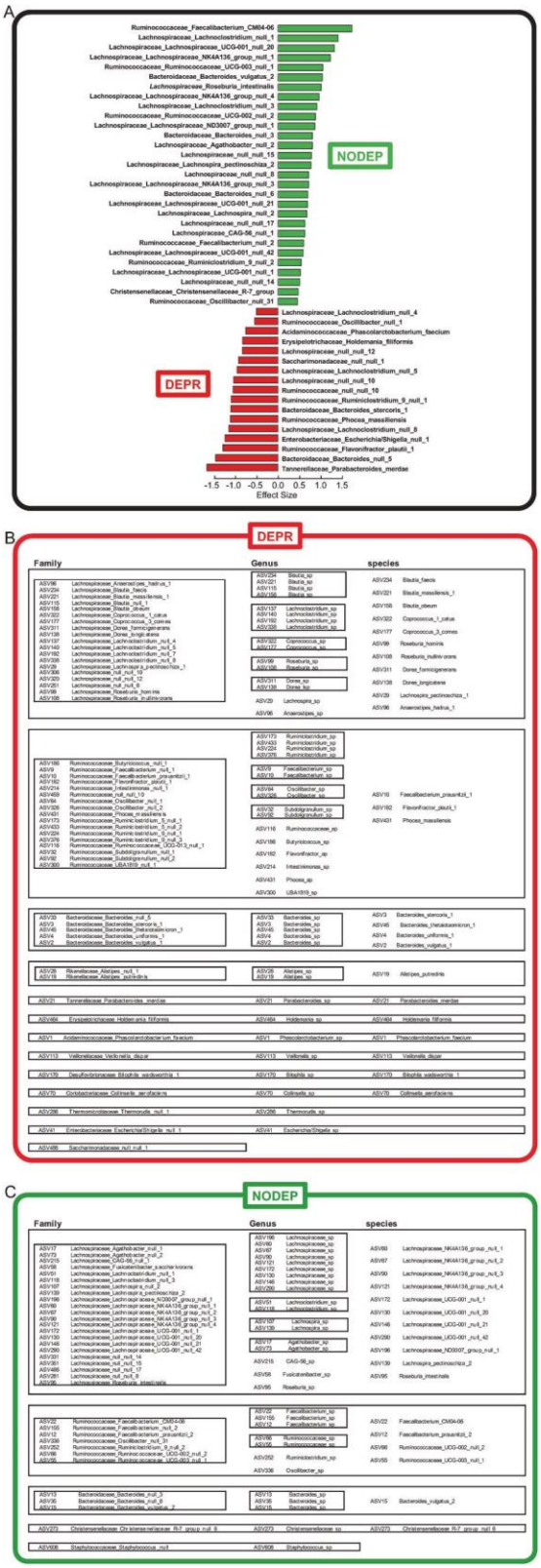
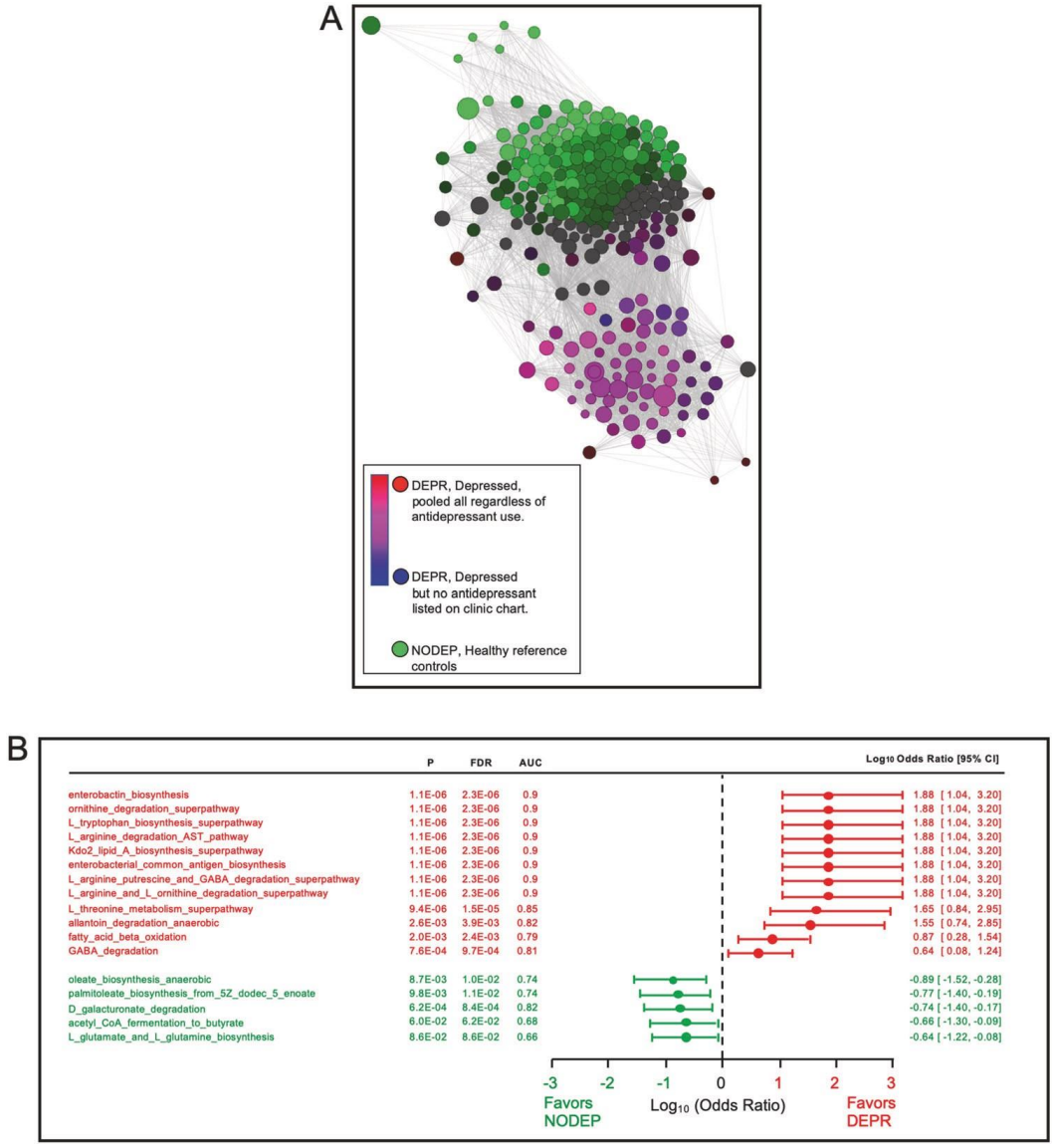


Fig. 2 ASV and taxa prevalences for DEPR vs. NODEP. **a** .ALDEx2 effect sizes for taxons assigned from ASVs. Displayed cutoffs are effect size ≥0.5 (NODEP) or ≤−0.5 (DEPR). **b** DEPR ASVs and taxons. **c** NODEP ASVs and taxons. **a–c** The DADA2/ALDEx2/ PIME pipeline output 55% prevalence dataset taxa names were assigned to each 86 unique ASV, regardless of whether multiple sequences could be assigned the same taxonomic name, based on the naive Bayesian classifier method and the SILVA ribosomal RNA gene database [35] version v132. Redundant multiple copies (N > 1) or unique (N = 1) taxonomic names for the ASVs are shown at the levels of Family, Genus, and species. The full set of all ALDEx2 effect sizes ASV sequences are listed in Supplementary Information Table 1-SI and Table-2SI.

Fig. 3 Metabolic pathway segregation of DEPR vs. NODEP. **a** Bray–Curtis dissimilarity network



analysis of microbiome metabolic pathways, and lack of antidepressant influence on cohort distances. Data are 284MetaCyc pathways predicted by PICRUSt2 using the DADA2/ALDEx2/PIME 55% bin prevalence taxa dataset of 86 ASVs. Pearson correlation algorithm was employed with pathway nodes placed by PCoA, showing significant (FDR $P < 0.05$) positive associations connected by edges, with similarity cutoff at 0.25. Node sizes and colors are proportional to relative magnitude within the dataset. Note pathway clustering and purple color blend resulting from the overlay of DEPR subjects whose charts listed an antidepressant (red) on DEPR subjects whose charts did not list a specific antidepressant (blue), and of which pooled DEPR metadata were collectively segregated from clustered NODEP. **b**. Odds ratios in forest plot of select microbiome metabolic pathway data of Fig. 2a. Based on the PICRUSt2 MetaCyc pathway data, odds ratios and AUC were generated, with salient pathway differences shown. Note the untoward pathways associated with depression pathophysiology phenotype in DEPR, in contrast to healthy pathways in NODEP.

APÊNDICE F – A microbiota intestinal é disbiótica em pacientes com homocistinúria clássica?

Is the gut microbiota dysbiotic in patients with classical homocystinuria?

Gustavo Mottin Rizowy [a, b], Soraia Poloni [b, c], Karina Colonetti [a, b], Karina Carvalho Donis [a, d], Priscila Thiago Dobbler [e], Sandra Leistner-Segal [b, c, d], Luiz Fernando Wurdig Roesch [e], Ida Vanessa Doederlein Schwartz [a, b, c, d, *]

[a] PostGraduate Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[b] Laboratory of Basic Research and Advanced Investigations in Neuroscience (BRAIN), Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil
[c] Post-Graduation Program in Medicine: Medical Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[d] Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil
[e] Interdisciplinary Research Center on Biotechnology (CIP-Biotec), Universidade Federal do Pampa, Sao Gabriel, Rio Grande do Sul, Brazil~

* Corresponding author. Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Rua Ramiro Barcelos, 2350, Porto Alegre, RS, 90035-903, Brazil. E-mail address: ischwartz@hcpa.edu.br (I.V.D. Schwartz).

DOI: https://doi.org/10.1016/j.biochi.2020.02.013

Received 31 October 2019
Accepted 20 February 2020
Available online 24 February 2020

Keywords: Classical homocystinuria; Gut microbiota; Dysbiosis; CBS deficiency; Homocysteine

Abstract

Classical homocystinuria (HCU) is characterized by increased plasma levels of total homocysteine (tHcy) and methionine (Met). Treatment may involve supplementation of B vitamins and essential amino acids, as well as restricted Met intake. Dysbiosis has been described in some inborn errors of metabolism, but has not been investigated in HCU. The aim of this study was to investigate the gut microbiota of HCU patients on treatment. Six unrelated HCU patients (males = 5, median age = 25.5 years) and six age-andsex-matched healthy controls (males = 5, median age = 24.5 years) had their fecal microbiota characterized through partial 16S rRNA gene sequencing. Fecal pH, a 3-day dietary record, medical history, and current medications were recorded for both groups. All patients were nonresponsive to pyridoxine and were on a Met-restricted diet and presented with high tHcy. Oral supplementation of folate (n = 6) and pyridoxine (n = 5), oral intake of betaine (n = 4), and IM vitamin B12 supplementation (n = 4), were reported only in the HCU group. Patients had decreased daily intake of fat, cholesterol, vitamin D, and selenium compared to controls ($p < 0.05$). There was no difference in alpha and beta diversity between the groups. HCU patients had overrepresentation of the *Eubacterium coprostanoligenes* group and underrepresentation of the *Alistipes*, *Family XIII UCG-001*, and *Parabacteroidetes* genera. HCU patients and controls had similar gut microbiota diversity, despite differential abundance of some bacterial genera. Diet, betaine, vitamin B supplementation, and host genetics may contribute to these differences in microbial ecology.

1. Introduction

Classical homocystinuria (HCU; OMIM 236200), or cystathionine beta-synthase (CBS; EC 4.2.1.22) deficiency, is an inborn error of metabolism (IEM) which predominantly affects the transsulfuration pathway. CBS is a key enzyme for the transsulfuration pathway, because it irreversibly catalyzes the conversion of homocysteine (Hcy) into cystathionine, using pyridoxal phosphate (an active form of pyridoxine) as a cofactor. CBS is expressed mainly in the liver, but also in the pancreas, kidneys, and brain [1e3].

The pathophysiology of HCU is still not fully understood. The spectrum of clinical manifestations is broad, ranging from paucisymptomatic patients to a very severe multisystem disease. The main organ systems affected are the eyes (ectopia lentis and/or severe myopia), skeleton (abnormally high stature, long limbs, osteoporosis, bone deformities), central nervous system (CNS) (developmental delay/intellectual disability, seizures, psychiatric and behavioral problems, extrapyramidal signs), and vascular system (thromboembolic events). The HCU phenotype is closely related to pyridoxine responsiveness; usually, pyridoxineresponsive patients have a milder phenotype and a later onset than nonresponsive patients [3]. Pyridoxine responsiveness is commonly defined according to plasma total homocysteine (tHcy) concentration achieved after a challenging test with pyridoxine, with tHcy decreasing to below 50 mmol/L [3,4].

The aim of HCU treatment is to reduce tHcy to a safe level; nevertheless, normal concentrations are usually unachievable, especially in nonresponsive patients. Treatment consists of pyridoxine supplementation (for responsive patients) and a

Methione(Met)-restricted diet and/or betaine (for nonresponsive patients). As a consequence of the Met-restricted diet, nutritional supplementation with a metabolic formula (a Met-free mixture of L-amino acids and micronutrients) is necessary; in addition, folate and vitamin B12 levels should be monitored and supplemented if necessary. Treatment must be lifelong [3].

The gut microbiota is composed of trillions of microorganisms [5]. Recent studies have shown an impact of these microorganisms on physiological processes, and changes in microbiota have been linked to a range of diseases [6,7], including CNS [8,9] and vascular disorders [10e12]. Additionally, health-illness continuum in humans is associated with differences in microbial communities and their functions [13].

Diet and genetics are known to influence the gut microbiome [14,15]. The restrictive diet, use of a synthetic L-amino acids formula and vitamin supplementation used in HCU treatment may affect the gut microbiota [16]. Within this context, the present study aims to investigate possible relationships between HCU treatment and the gut microbiota.

2. Methods

2.1 Experimental design

This observational, cross-sectional study used a convenience sampling strategy, which included six HCU patients and six ageand-sex-matched healthy controls. The study protocol was approved by the Ethics Committee of Hospital de Clínicas de Porto Alegre (HCPA) (project number 2015-0218). All participants and/or their legal guardians provided written informed consent. Patients were recruited from outpatient clinics of the Medical Genetics Service (MGS), HCPA, Rio Grande do Sul, Brazil. Control subjects were recruited from the Santa Cecília Basic Health Unit, Porto Alegre, Brazil, which is located nearby to HCPA; subjects who performed routine medical appointments at this Basic Health Unit were invited to participate in the study. The inclusion criteria for patients were: a) having a genetic diagnosis of HCU (e.g., presence of biallelic pathogenic mutations in the CBS gene); b) being on treatment; c) having no other chronic disease; and d) no current use of antibiotics. The inclusion criteria for controls were a) no relation to HCU patients, b) no chronic or acute diseases, and c) no current use of antibiotics.

Clinical information, such as treatment strategies and tHcy and Met levels, was obtained byreviewing medical records. Besides that, a specific questionnaire was applied to every participant, which included questions on the use of medicines (antibiotics, laxatives), and of probiotics and fermented milk, and fibers supplementation in the last 6 months. Additionally, the use of dietary fiber supplements was also recorded. The criteria used to define metabolic control was based on the 1-year median of tHcy measurements before the stool collection. If the median tHcy value was within target, the patient was considered to have good metabolic control. The target level of tHcy is 50 mmol/L for pyridoxine-responsive patients, and <100 mmol/L for nonresponsive patients [17].

All subjects received a kit which consisted of a Styrofoam box containing a sterile container in which to store the stool sample, a gel ice pack, and a sterile spatula to collect the sample. The participants also received printed instructions for stool collection, storage, and transport. They were instructed to collect stool in their own homes on the day before their medical appointments at HCPA, store the specimen in

a household freezer (20 C), and deliver it the next day, on ice, during the scheduled appointment. All participants were also provided printed instructions and a sheet to record 3 days of dietary information. The participants were instructed to write down everything they had eaten over the 3-day period, as well as the amount and manner of preparation of all foods and beverages. Stool samples were preferably to be collected on the third day of the food record. Upon returning to the clinic, each participant answered a questionnaire about personal features, including dietary and bowel habits and current medications.

2.2 Nutritional assessment

Patients and controls were instructed to complete a 3-day dietary record (day 3 = the day when the stool sample was collected; days 1 and 2 = before the collection of the stool sample). Macroand micronutrient intake were analyzed using Nutribase™ software (NB16Cloud, Cybersoft Inc. Phoenix, AZ, USA). As there is no Brazilian table of the Met content on foods, daily Met intake could not be estimated; hence, only total protein intake was calculated. Daily nutrient intake was computed considering the average of the three days of food records. Vitamin supplements and other medications were not included in the nutritional analysis.

2.3 pH measurement, bacterial DNA extraction, 16S rRNA geneamplification and sequencing

Frozen stool samples of participants were thawed and aliquoted at room temperature (20 C) for pH measurement [18]. To measure the fecal pH, samples were diluted 1:10 (w/v) in distilled water. The dilute was homogenized and incubated for 5 min at room temperature, and fecal pH was measured with an electronic pH meter (K39e1014B, KASVI, PR, Brazil) after complete immersion of the electrode for 3 min.

Bacterial DNA was isolated from 300 mg of frozen stool sample using the QIAmp Fast DNA Stool Mini Kit (Qiagen, Valencia, CA, USA), following manufacturer instructions. DNA quality was determined by spectrophotometry in a NanoVue™ system (GE Healthcare, Chicago, IL, USA). The sequencing library was prepared following the procedures described by Barboza et al. [19]. The gut microbial community was determined by amplification of the V4 region of 16S rRNA with the barcode bacterial/archeal primers 515F and 806R [20]. PCR reactions were carried out with 2U of Platinum Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4 mL 10X High Fidelity PCR Buffer, 2 mM $MgSO_4$, 0.2 mM dNTPs, 0.1 mM of each barcoded primer, 25 mg of Ultrapure BSA (Invitrogen, Carlsbad, CA, USA) and approximately 50 ng of DNA template in a final volume of 25 mL. PCR conditions were 95 C for 5 min, 30 cycles at 94 C for 45 s, 56 C for 45 s, and 72 C for 1 min, followed by a final extension step of 10 min at 72 C. After visualization on a 1.5% agarose gel, the PCR products were purified with Agencourt AMPure XP Reagent (Beckman Coulter, Brea, CA, USA). The final concentration of the PCR product was quantified with a Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA), following manufacturer instructions, and combined in equimolar ratios to create a mixture composed of amplified 16S gene fragments of each sample. Ultimately, this composite was used for library preparation in the Ion One-Touch 2 System, using Ion PGM Template OT2 400 Kit (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed with Ion PGM Sequencing 400 on the Ion PGM System, using 318 Chip kit v2.

2.4. 16S profiling data analysis

The 16S rRNA reads were analyzed following the recommendations of the Brazilian Microbiome Project [21], for an efficient removal of sequencing artifacts that might exacerbate biases due to the presence of chimeric sequences and sequence errors. Briefly, an Operational Taxonomic Unit (OTU) table was built using the UPARSE pipeline [22]. Reads were truncated at 200 bp and qualityfiltered with a maximum expected error of 0.5. Filtered reads were dereplicated and singletons were removed. The sequences were clustered into OTUs at a 97% cutoff, following the UPARSE pipeline. Taxonomic classification was carried out in QIIME version 1.9.1 [23], based on the UCLUST method, against SILVA ribosomal RNA gene database version v1327 [24] with 80% boundary confidence. Downstream analyses were performed with dataset rarefy to the minimum library size [25,26] in the R software environment [27], using the phyloseq [28], vegan [29], and ALDEx2 [30,31] packages. Differential abundant microbes were selected based on the effect size rather than on p-values only, as proposed by Gloor et al. [31]. Spearman correlation was used to evaluate the association between HCU biochemical markers (tHcy and Met levels) and OTU richness.

2.5. Statistical analysis

Statistical analysis of clinical data among groups was carried out in PASW Statistics for Windows software (VVs18.0, 2009; SPSS Inc., Chicago, IL, USA). For comparison between groups, continuous variables were analyzed using the ManneWhitney U test. Categorical variables were compared using Fisher's exact test ($p \leq 0.05$).

3. Results

Six unrelated HCU patients and six healthy individuals were included in the study (Tables 1 and 2). All patients had been diagnosed late and were unresponsive to pyridoxine. All were on a Metrestricted diet and presented with hyperhomocysteinemia. Oral supplementation of folate (n = 6) and pyridoxine (n = 5), oral Betaine intake (n = 4), and intramuscular vitamin B12 supplementation (n = 4) were reported only in the HCU group. Only two patients had good metabolic control (patients H1 and H3, Table 2).

Nutritional analysis (Table 1) revealed differences in the intake of some nutrients between patients and controls. Overall fat ($p = 0.025$), saturated fat ($p = 0.004$), monounsaturated fat ($p = 0.004$), cholesterol ($p = 0.004$), vitamin D ($p = 0.004$), and selenium ($p = 0.016$) intake was lower in HCU patients compared to controls.

3.1. Microbiota composition and the correlation between gut microbiota and HCU biochemical markers

Nine known bacterial phyla were detected within all samples (Fig.1). The most dominant was *Bacteroidetes* (HCU: 62.5%; controls: 55.2%), followed by *Firmicutes* (HCU: 32.7%; controls: 39.1%) and *Proteobacteria* (HCU: 2.1%; controls: 5.1%). Alpha diversity analysis (Fig. 2) showed no difference between HCU patients and controls, neither in richness nor in evenness. The lack of difference in diversity between HCU and controls suggests that HCU treatment does not have an effect on gut bacterial diversity. Differences in gut composition between treatments were verified by Principal Coordinates Analysis (PCoA) and permutational multivariate analysis of variance

(PERMANOVA). Beta diversity analysis is dependent upon distances of dissimilarities between samples; here, the Bray-Curtis and Binary distances were used. Irrespective of the distance applied, the beta diversity analysis showed no difference between HCU patients and controls, as observed by PCoA (Fig. 3) and PERMANOVA.

The association between HCU gut microbiota (number of OTUs), tHcy and Met is shown in Fig. 4. Microbial richness had no correlation with tHcy ($R = -0.43$, $p = 0.42$) and Met levels ($R = -0.77$, $p = 0.1$).

3.2. Biomarker analysis

Although no differences in overall alpha and beta diversity were detected between HCU patients and controls, significant associations between HCU patients and specific bacteria might still exist. Such associations were tested by modeling the data as a log-ratio transformed probability distribution rather than counts. Differential abundant microbes analysis indicated that one OTU closely related with the *Eubacterium coprostanoligenes* group was increased in HCU patients (Table 3). On the other hand, four OTUs closely related with *Alistipes* (2 OTUs), *Family XIII UCG-001*, and *Parabacteroidetes* were increased in controls (Table 3).

4. Discussion

To the best of our knowledge, this is the first study to characterize the gut microbiota of HCU patients. Surprisingly, our data suggest the microbiota of HCU patients and controls does not differ regarding diversity.

Dysbiosis can be defined as any perturbation to the structure of complex commensal communities [32]. Dysbiosis can contribute to the onset of chronic disease in one of three general ways: 1) pathogens and their functions can be acquired or opportunistically overgrow (gain of function dysbiosis); 2) health-protective bacteria and their functions may be lost or suppressed (loss of function dysbiosis), and 3) a combination of loss and gain of function dysbiosis [33]. Studies on the relationship between the gut microbiota and other IEM are still scarce and have shown that treated phenylketonuria (PKU) [34] and hepatic glycogen storage diseases (GSDs) patients have dysbiosis [18]. Even though the genetic and dietary aspects of these IEMs are different from HCU, both GSD and PKU patients had decreased alpha and beta diversity and distinct microbiota composition when compared to healthy subjects. We expected that the gut microbiota profile of patients with HCU would be similar to that of patients with PKU, as treatment for both diseases involves a restrictive diet (in PKU, there is a restricted intake of phenylalanine and in HCU, of Met) and use of a metabolic formula to supplement nutrients. Nevertheless, we found no differences in alpha nor beta diversity in HCU patients when compared to controls. We also found no association between gut microbiota and biochemical markers in HCU patients.

Our finding of lower intake of cholesterol, fat (saturated and monounsaturated), vitamin D, and selenium in HCU patients is mainly explained by their Met-restricted diet, which excludes or restricts many foods from animal origin as well as nuts and beans, which are sources of these nutrients [35]. It is important to point out, however, that the intake of these nutrients was not below dietary recommendations; that most patients, actually, were not fully compliant to the dietary treatment (and, so, showed a bad metabolic control); and that the metabolic formula for HCU includes supplementation of vitamin D and selenium (at the time of inclusion in the study, only 3 out of the 6 patients were taking the formula, but only one had a good metabolic

control). Regarding other nutrients, no differences were found between the groups. The most probably explanation for this finding is also the poor adherency to the Met-restricted diet.

A unique feature of HCU treatment is the supplementation of high doses of B vitamins [3], and the supplementation of dietary nutrients related to one-carbon metabolism has shown a role in modulating the gut microbiota[36e39]. Therefore, vitamin supplementation may be involved in the diversity profile found in patients with HCU. Furthermore, studies in mice have shown possible beneficial effects of a Met-restricted diet, such as decreased intestinal permeability, inflammation, and oxidative stress [40,41].

Gurwara et al. [36] reported that B vitamins, involved in onecarbon metabolism, were associated with variations in microbiota profile: high dietary intake of folate, pyridoxine, and vitamin B12 was associated with an increase in richness and evenness. Interestingly, our study found no difference in richness or in evenness. In addition, Gurwara et al. [36] found that high intake of pyridoxine, vitamin B12, and folate was associated with an increased abundance of Verrucomicrobia and *Alistipes*, while our study found a decreased abundance of *Alistipes* in patients with HCU. Furthermore, low abundance of *Alistipes* in the human gut microbiota is known to be associated with better dietary quality [42]. As HCU patients have a restricted diet, they are not expected to have a high dietary quality, and this was a surprising finding. However, some important differences must be noted: (a) the Gurwara et al. study [36] was performed only in men between the ages of 50 and 75; (b) bacterial samples were obtained by colonoscopy, not from stool samples; and (c) their criterion of high or low intake of the analyzed vitamins was exclusively dietary, while our patients with HCU were taking high supplemental doses of these vitamins.

The relationship between gut microbiota and the CNS has been widely described in the literature; however, its mechanisms are not fully understood. The gut-brain axis is bidirectional [43], and a number of studies have associated gut microbiota profile with neurodegenerative diseases [9] and neuropsychiatric disorders [8]. In this study, we were unable to evaluate CNS manifestations in light of microbiome profile, both due to the small sample size and because, in patients with HCU, CNS manifestations may be secondary to the toxic effects of high Hcy levels [44]. Nevertheless, Hcy itself is able to disrupt the blood-brain barrier [45], and it is plausible that the microbiota might be a contributor to CNS manifestations in this condition due to a myriad of immune-cellular mechanisms [46]. This relationship must be elucidated further to understand how the gut microbiota may be related to CNS manifestations.

Our study found an increase in the genus *Eubacterium coprostanoligenes* group in HCU patients. This bacterial genus has been associated with anxiety disorder [47], psychosocial stress [48], and cholesterol metabolism [49,50]. Although there is lack of data in the literature, much because it is a bacterial genus not yet cultivable. In addition, we found a decrease in the genera *Parabacteroides* and *Family XIII UCG-001*. The decrease in the genus *Parabacteroides* was related to lower intake of milk and dairy products [42], and, indeed, the diet of HCU patients restricts intake of these types of foods.

Despite little information in the literature, an increase in *Family XIII UCG-001* abundance has been described as a neuroprotective biomarker in chronic social defeat stress-induced depressive-like behavior in mice treated preventively with probiotics [51].

This was the first study designed to characterize the gut microbiota in HCU patients under treatment. As HCU is a rare disease, we were only able to enroll a small number of patients. This, and the cross-sectional design which precludes any causal inference were the main limitations of our study.

5. Conclusion

Our data suggest that the diversity of gut microbiota is similar in patients with HCU and healthy controls, despite differences in some genera. The gut microbiota profile found in HCU patients is probably a sum of several factors, such as diet and treatment; host genetics may be related to differences in microbial ecology and even to the presence of bacterial genera still little described in the literature. Future studies on the gut microbial composition of HCU patients are needed to confirm these findings and to investigate the association of gut microbiota with treatment regimens and biochemical features of HCU.

Author contributions

Gustavo M. Rizowy developed and designed the project, collected and analyzed data, and drafting and proofreading the manuscript.
Soraia Poloni helped collect and analyzed data, and drafting and proofreading of the manuscript.
Karina Colonetti developed and designed the project, collected and analyzed data, and drafted the manuscript.
Karina C. Donis helped collect data and draft the manuscript.
Priscila T. Dobbler helped collect and analyze data.
Sandra Leistner-Segal developed and designed the project, analyzed data, and drafted the manuscript.
Luiz Fernando W. Roesch took part in the development and review of the project, data collection and analysis, and drafting and proofreading of the manuscript.
Ida Vanessa D. Schwartz was in charge of development and review of the project, data analysis, and drafting and proofreading of the manuscript.

Declaration of competing interest

None.

Acknowledgments

We thank our subjects for their participation, and the teams at MGS-HCPA and Cip-Biotec for their assistance.
The present study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnologico (CNPq), Fundaç ao de~ Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS) e EDITAL PRONEX-FAPERGS/CNPq 12/2014, processo 16/25510000492-7, Fundo de Incentivo a Pesquisa e Eventos do Hospital de Clínicas de Porto Alegre (FIPE-HCPA), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior e Brasil (CAPES).

References

[1] K.J. Mudd SH, Levy HL, Disorders of transulfuration. In Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW et al, eds. The Metabolic and Molecular Bases of Inherited Disease. New York: McGraw Hill; 2007-2043., n.d, https:// ommbid.mhmedical.com/content.aspx?bookid=971& sectionid=62675596.
[2] S.H. Mudd, F. Skovby, H.L. Levy, K.D. Pettigrew, B. Wilcken, R.E. Pyeritz, G. Andria, G.H. Boers, I.L. Bromberg, R. Cerone, The natural history of homocystinuria due to cystathionine beta-synthase deficiency, Am. J. Hum. Genet. 37 (1985) 1e31. http://www.ncbi.nlm.nih.gov/pubmed/3872065. (Accessed 16 October 2019).
[3] A.A.M. Morris, V. Kozich, S. Santra, G. Andria,T.I.M. Ben-Omran, A.B. Chakrapani, E. Crushell, M.J. Henderson, M. Hochuli, M. Huemer, M.C.H. Janssen, F. Maillot, P.D. Mayne, J. McNulty, T.M. Morrison, H. Ogier, S. O'Sullivan, M. Pavlíkova, I.T. de Almeida, A. Terry, S. Yap, H.J. Blom, K.A. Chapman, Guidelines for the diagnosis and management of cystathionine beta-synthase deficiency, J. Inherit. Metab. Dis. 40 (2017) 49e74, https:// doi.org/10.1007/s10545-016-9979-0.
[4] L.A. Kluijtmans, G.H. Boers, J.P. Kraus, L.P. van den Heuvel, J.R. Cruysberg, F.J. Trijbels, H.J. Blom, The molecular basis of cystathionine beta-synthase deficiency in Dutch patients with homocystinuria: effect of CBS genotype on biochemical and clinical phenotype and on response to treatment, Am. J. Hum. Genet. 65 (1999) 59e67, https://doi.org/10.1086/302439.
[5] R. Sender, S. Fuchs, R. Milo, Revised estimates for the number of human and bacteria cells in the body, PLoS Biol. 14 (2016), https://doi.org/10.1371/ journal.pbio.1002533.
[6] P.C. Barko, M.A. McMichael, K.S. Swanson, D.A. Williams, The gastrointestinal microbiome: a review, J. Vet. Intern. Med. 32 (2018) 9e25, https://doi.org/ 10.1111/jvim.14875.
[7] Y.-J. Zhang, S. Li, R.-Y. Gan, T. Zhou, D.-P. Xu, H.-B. Li, Y.-J. Zhang, S. Li, R.Y. Gan, T. Zhou, D.-P. Xu, H.-B. Li, Impacts of gut bacteria on human health and diseases, Int. J. Mol. Sci. 16 (2015) 7493e7519, https://doi.org/10.3390/ ijms16047493.
[8] L.F. Iannone, A. Preda, H.M. Blottiere, G. Clarke, D. Albani, V. Belcastro, M. Carotenuto, A. Cattaneo, R. Citraro, C. Ferraris, F. Ronchi, G. Luongo, E. Santocchi, L. Guiducci, P. Baldelli, P. Iannetti, S. Pedersen, A. Petretto, S. Provasi, K. Selmer, A. Spalice, A. Tagliabue, A. Verrotti, N. Segata, J. Zimmermann, C. Minetti, P. Mainardi, C. Giordano, S. Sisodiya, F. Zara, E. Russo, P. Striano, Microbiota-gut brain axis involvement in neuropsychiatric disorders, Expert Rev. Neurother. 19 (2019) 1037e1050, https://doi.org/ 10.1080/14737175.2019.1638763.
[9] E.M.M. Quigley, Microbiota-brain-gut Axis and neurodegenerative diseases, Curr. Neurol. Neurosci. Rep. 17 (2017) 94, https://doi.org/10.1007/s11910017-0802-6.
[10] H. Formes, C. Reinhardt, The gut microbiota - a modulator of endothelial cell function and a contributing environmental factor to arterial thrombosis, Expet. Rev. Hematol. 12 (2019) 541e549, https://doi.org/10.1080/ 17474086.2019.1627191.
[11] G. Baffy, Potential mechanisms linking gut microbiota and portal hypertension, Liver Int. 39 (2019) 598e609, https://doi.org/10.1111/liv.13986.

[12] A. Jovanovich, T. Isakova, J. Stubbs, Microbiome and cardiovascular disease in CKD, Clin. J. Am. Soc. Nephrol. 13 (2018) 1598e1604, https://doi.org/10.2215/ CJN.12691117.
[13] Human Microbiome Project Consortium, Structure, function and diversity of the healthy human microbiome, Nature 486 (2012) 207e214, https://doi.org/ 10.1038/nature11234.
[14] A.B. Hall, A.C. Tolonen, R.J. Xavier, Human genetic variation and the gut microbiome in disease, Nat. Rev. Genet. 18 (2017) 690e699, https://doi.org/ 10.1038/nrg.2017.63.
[15] E.R. Leeming, A.J. Johnson, T.D. Spector, C.I. Le Roy, Effect of diet on the gut microbiota: rethinking intervention duration, Nutrients 11 (2019) 2862, https://doi.org/10.3390/nu11122862.
[16] K. Colonetti, L.F. Roesch, I.V.D. Schwartz, The microbiome and inborn errors of metabolism: why we should look carefully at their interplay?, Genet. Mol. Biol. 41 (n.d.) 515e532. doi:10.1590/1678-4685-GMB-2017-0235.
[17] A.A.M. Morris, V. Kozich, S. Santra, G. Andria, T.I.M. Ben-Omran, A.B. Chakrapani, E. Crushell, M.J. Henderson, M. Hochuli, M. Huemer, M.C.H. Janssen, F. Maillot, P.D. Mayne, J. McNulty, T.M. Morrison, H. Ogier, S. O'Sullivan, M. Pavlíkova, I.T. de Almeida, A. Terry, S. Yap, H.J. Blom, K.A. Chapman, Guidelines for the diagnosis and management of cystathionine beta-synthase deficiency, J. Inherit. Metab. Dis. 40 (2017) 49e74, https:// doi.org/10.1007/s10545-016-9979-0.
[18] K. Colonetti, B. Bento dos Santos, T. Nalin, C.F. Moura de Souza, E.W. Triplett, P.T. Dobbler, I.V.D. Schwartz, L.F.W. Roesch, Hepatic glycogen storage diseases are associated to microbial dysbiosis, PloS One 14 (2019), e0214582, https:// doi.org/10.1371/journal.pone.0214582.
[19] A.D.M. Barboza, V.S. Pylro, R.J.S. Jacques, P.I. Gubiani, F.L.F. de Quadros, J.K. da Trindade, E.W. Triplett, L. Roesch, Seasonal dynamics alter taxonomical and functional microbial profiles in Pampa biome soils under natural grasslands, PeerJ 6 (2018), e4991, https://doi.org/10.7717/peerj.4991.
[20] J.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J.A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, ISME J. 6 (2012) 1621e1624, https:// doi.org/10.1038/ismej.2012.8.
[21] V.S. Pylro, L.F.W. Roesch, D.K. Morais, I.M. Clark, P.R. Hirsch, M.R. Totola, Data analysis for 16S microbial profiling from different benchtop sequencing platforms, J. Microbiol. Methods 107 (2014) 30e37, https://doi.org/10.1016/ J.MIMET.2014.08.018.
[22] R.C. Edgar, UPARSE: highly accurate OTU sequences from microbial amplicon reads, Nat. Methods 10 (2013) 996e998, https://doi.org/10.1038/nmeth.2604.
[23] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon, G.A. Huttley,~ S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335e336, https://doi.org/10.1038/nmeth.f.303.

[24] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glockner, The SILVA ribosomal RNA gene database project: improved data€ processing and web-based tools, Nucleic Acids Res. 41 (2012) D590eD596, https://doi.org/10.1093/nar/gks1219.
[25] L.N. Lemos, R.R. Fulthorpe, E.W. Triplett, L.F.W. Roesch, Rethinking microbial diversity analysis in the high throughput sequencing era, J. Microbiol. Methods 86 (2011) 42e51, https://doi.org/10.1016/j.mimet.2011.03.014.
[26] L.N. Lemos, R.R. Fulthorpe, L.F.W. Roesch, Low sequencing efforts bias analyses of shared taxa in microbial communities, Folia Microbiol. 57 (2012) 409e413, https://doi.org/10.1007/s12223-012-0155-0.
[27] R.C. Team, R development Core team. R: A Language and Environment for Statistical Computing 55 (2017) 275e286 (n.d.).
[28] P.J. McMurdie, S. Holmes, Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, PloS One 8 (2013), e61217, https://doi.org/10.1371/journal.pone.0061217.
[29] J. Oksanen, F.G. Blanchet, R. Kindt, P. Legendre, P. Minchin, R.B. O'Hara, G. Simpson, P. Solymos, M.H.H. Stevens, H. Wagner, Vegan: Community Ecology Package, 2013. R package version 2.0-7.[WWW document]. (Accessed 10 July 2013), http://Cran. Rproject. Org/.
[30] A.D. Fernandes, J.N. Reid, J.M. Macklaim, T.A. McMurrough, D.R. Edgell, G.B. Gloor, Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis, Microbiome 2 (2014) 15, https:// doi.org/10.1186/2049-2618-2-15.
[31] G.B. Gloor, J.M. Macklaim, A.D. Fernandes, Displaying variation in large datasets: plotting a visual summary of effect sizes, J. Comput. Graph Stat. 25 (2016) 971e979, https://doi.org/10.1080/10618600.2015.1131161.
[32] C. Petersen, J.L. Round, Defining dysbiosis and its influence on host immunity and disease, Cell Microbiol. 16 (2014) 1024e1033, https://doi.org/10.1111/ cmi.12308.
[33] L.J. Wilkins, M. Monga, A.W. Miller, Defining dysbiosis for a cluster of chronic diseases, Sci. Rep. 9 (2019), https://doi.org/10.1038/s41598-019-49452-y.
[34] F. Pinheiro de Oliveira, R.H. Mendes, P.T. Dobbler, V. Mai, V.S. Pylro, S.G. Waugh, F. Vairo, L.F. Refosco, L.F.W. Roesch, I.V.D. Schwartz, Phenylketonuria and gut microbiota: a controlled study based on next-generation sequencing, PloS One 11 (2016), e0157513, https://doi.org/10.1371/ journal.pone.0157513.
[35] P.J. Turnbaugh, R.E. Ley, M. Hamady, C.M. Fraser-Liggett, R. Knight, J.I. Gordon, The human microbiome project, Nature 449 (2007) 804e810, https://doi.org/ 10.1038/nature06244.
[36] S. Gurwara, N.J. Ajami, A. Jang, F.C. Hessel, L. Chen, S. Plew, Z. Wang, D.Y. Graham, C. Hair, D.L. White, J. Kramer, T. Kourkoumpetis, K. Hoffman, R. Cole, J. Hou, N. Husain, M. Jarbrink-Sehgal, R. Hernaez, F. Kanwal, G. Ketwaroo, R. Shah, M. Velez, Y. Natarajan, H.B. El-Serag, J.F. Petrosino, L. Jiao, Dietary nutrients involved in one-carbon metabolism and colonic mucosa-associated gut microbiome in individuals with an endoscopically normal colon, Nutrients 11 (2019), https://doi.org/10.3390/nu11030613.

[37] P.H. Degnan, M.E. Taga, A.L. Goodman, Vitamin B 12 as a modulator of gut microbial ecology, Cell Metabol. 20 (2014) 769e778, https://doi.org/10.1016/ j.cmet.2014.10.002.
[38] B.S. Ramakrishna, Role of the gut microbiota in human nutrition and metabolism, J. Gastroenterol. Hepatol. 28 (2013) 9e17, https://doi.org/10.1111/ jgh.12294.
[39] O.A. Masri, J.M. Chalhoub, A.I. Sharara, Role of vitamins in gastrointestinal diseases, World J. Gastroenterol. 21 (2015) 5191e5209, https://doi.org/ 10.3748/wjg.v21.i17.5191.
[40] Y. Yang, Y. Zhang, Y. Xu, T. Luo, Y. Ge, Y. Jiang, Y. Shi, J. Sun, G. Le, Dietary methionine restriction improves the gut microbiota and reduces intestinal permeability and inflammation in high-fat-fed mice, Food Funct. 10 (2019) 5952e5968, https://doi.org/10.1039/c9fo00766k.
[41] Y. Yang, Y. Wang, J. Sun, J. Zhang, H. Guo, Y. Shi, X. Cheng, X. Tang, G. Le, Dietary methionine restriction reduces hepatic steatosis and oxidative stress in high-fat-fed mice by promoting H2S production, Food Funct. 10 (2019) 61e77, https://doi.org/10.1039/c8fo01629a.
[42] Y. Liu, N.J. Ajami, H.B. El-Serag, C. Hair, D.Y. Graham, D.L. White, L. Chen, Z. Wang, S. Plew, J. Kramer, R. Cole, R. Hernaez, J. Hou, N. Husain, M.E. Jarbrink-Sehgal, F. Kanwal, G. Ketwaroo, Y. Natarajan, R. Shah, M. Velez, N. Mallepally, J.F. Petrosino, L. Jiao, Dietary quality and the colonic mucosaassociated gut microbiome in humans, Am.J. Clin. Nutr. 110 (2019) 701e712, https://doi.org/10.1093/ajcn/nqz139.
[43] H.-X. Wang, Y.-P. Wang, Gut Microbiota-brain Axis., Chin. Med. J. (Engl.) 129 (2016) 2373e2380, https://doi.org/10.4103/0366-6999.190667.
[44] H. Skovierov a, E. Vidomanova, S. Mahmood, J. Sopkova, A. Drgova, T. Cerve nov a, E. Hala sova, J. Lehotský, The molecular and cellular effect of homocysteine metabolism imbalance on human health, Int. J. Mol. Sci. 17 (2016), https://doi.org/10.3390/ijms17101733.
[45] A.F. Kamath, A.K. Chauhan, J. Kisucka, V.S. Dole, J. Loscalzo, D.E. Handy, D.D. Wagner, Elevated levels of homocysteine compromise blood-brain barrier integrity in mice, Blood 107 (2006) 591e593, https://doi.org/10.1182/ blood-2005-06-2506.
[46] F. Sommer, F. Backhed, The gut microbiota€ d masters of host development and physiology, Nat. Rev. Microbiol. 11 (2013) 227e238, https://doi.org/ 10.1038/nrmicro2974.
[47] Y.-H. Chen, J. Bai, D. Wu, S.-F. Yu, X.-L. Qiang, H. Bai, H.-N. Wang, Z.-W. Peng, Association between fecal microbiota and generalized anxiety disorder: severity and early treatment response, J. Affect. Disord. 259 (2019) 56e66, https://doi.org/10.1016/j.jad.2019.08.014.
[48] N. Michels, T. Van de Wiele, F. Fouhy, S. O'Mahony, G. Clarke, J. Keane, Gut microbiome patterns depending on children's psychosocial stress: reports versus biomarkers, Brain Behav. Immun. 80 (2019) 751e762, https://doi.org/ 10.1016/j.bbi.2019.05.024.
[49] L. Li, S.M. Batt, M. Wannemuehler, A. Dispirito, D.C. Beitz, Effect of feeding of a cholesterol-reducing bacterium, Eubacterium coprostanoligenes, to germ-free mice, Lab. Anim. Sci. 48 (1998) 253e255. http://www.ncbi.nlm.nih.gov/ pubmed/10090024. (Accessed 19 October 2019).

[50] P. Gerard, Metabolism of cholesterol and bile acids by the gut microbiota, Pathogens 3 (2013) 14e24, https://doi.org/10.3390/pathogens3010014.
[51] T. Tian, B. Xu, Y. Qin, L. Fan, J. Chen, P. Zheng, X. Gong, H. Wang, M. Bai, J. Pu, J. Lu, W. Zhou, L. Zhao, D. Yang, P. Xie, Clostridium butyricum miyairi 588 has preventive effects on chronic social defeat stress-induced depressive-like behaviour and modulates microglial activation in mice, Biochem. Biophys. Res. Commun. 516 (2019)430e436, https://doi.org/10.1016/j.bbrc.2019.06.053.

Figures

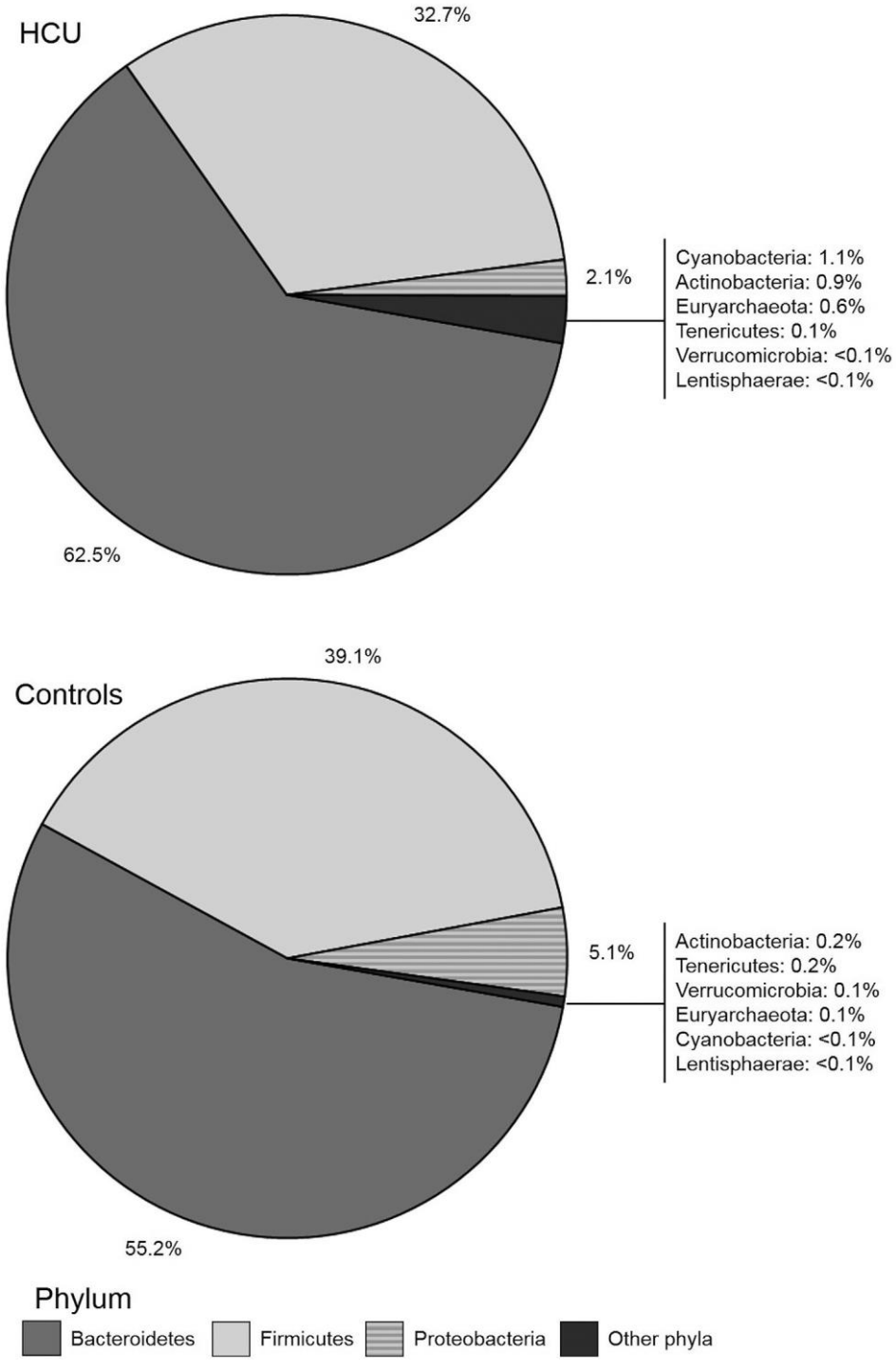


Fig. 1. Frequencies of phyla found in HCU patients (n=6) and healthy controls (n=6). The panel represents the average abundance per group. HCU: classical homocystinuria.

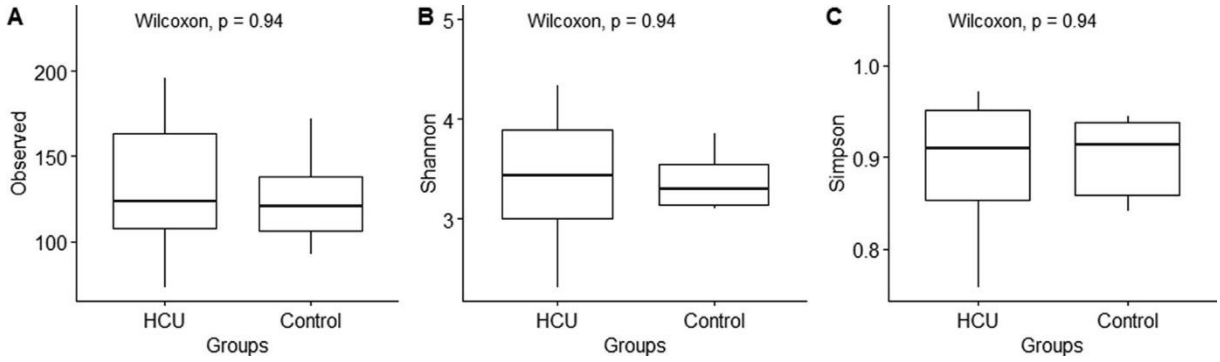


Fig. 2. Alpha diversity measurements of microbial communities in HCU patients (n=6) and controls (n=6). Each panel represent one alpha diversity measure: (A) Observed index estimates the amount of unique OTUs found (richness). (B) Shannon index accounts for richness (count) and evenness (distribution). (C) Simpson index accounts for both richness and evenness.

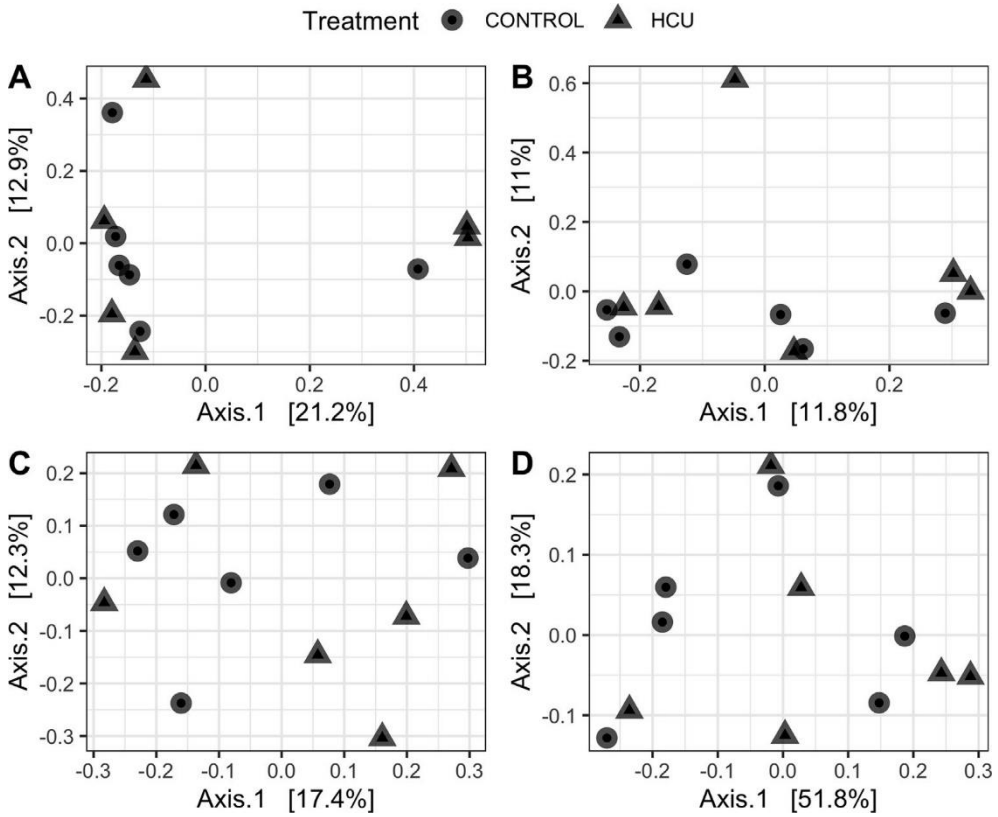


Fig. 3. Principal coordinates analysis (PCoA) representing the comparison of microbial communities in HCU patients and controls. PCoA base on (A) Bray-Curtis dissimilarity (F-value = 0.937; R^2 = 0.085; p-value = 0.453). (B) Binary distance (F-value = 0.966; R^2 = 0.088; p-value = 0.671). (C) Unweighted UniFrac (F-value = 0.212; R^2 = 0.895; pvalue = 0.72). (D) Weighted UniFrac (F-value = 0.824; R = 0.076; p-value = 0.453). HCU: classical homocystinuria

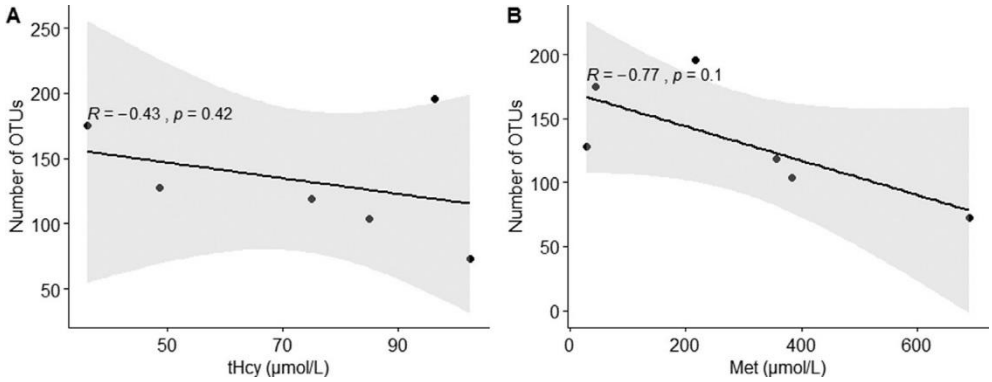


Fig. 4. Correlation between gut microbiota and biochemical markers. Spearman correlations between HCU gut microbiota richness (total number of OTUs), (A) tHcy and (B) Met levels.

Tables

Table 1. Comparison between the HCU and control groups.

Variable	HCU[a] (n = 6)	Control[a] (n = 6)	*p*-value
Sex (male:female)	5:1	5:1	1.000
Age (years)	25.5 (15.2e31.2)	24.5 (17.2e32.0)	0.810
Weight (kg)	63.0 (52.5e74.8)	68.9 (50.0e84.5)	0.631
Height (cm)	173.9 (154.6e182.3)	177.0 (153.2e183.5)	0.873
BMI (kg/m^2)	22.0 (19.7e24.0)	22.1 (19.7e24.8)	1.000
Fecal pH	7.2 (6.74e7.6)	7.3 (7.0e7.6)	0.873
Antibiotics[b] (yes:no)	2:4	1:5	1.000
Laxatives[b] (yes:no)	0:6	0:6	1.000
Probiotics[b] (yes:no)	0:6	1:5	1.000
Fibers supplementation[b] (yes:no)	0:6	0:6	1.000
Vitamin supplementation[b] (yes:no)	3:3	6:0	0.181
tHcy (mmol/L)	80.0 (45.5e97.8)	e	
Met (mmol/L)	287.1 (40.0e460.1)	e	
Daily intake -Calories (kcal)	1349.3 (1308.2e1863.3)	1863.0 (1220.0e2493.0)	0.631
-Calories (kcal/kg)	24.4 (19.1e37.2)	24.7 (18.5e35.5)	0.749
-Protein (g)	41.7 (28.1e102.6)	82.9 (56.1e140.7)	0.150
-Protein (g/kg)	0.8 (0.4e1.5)	1.0 (0.9e1.8)	0.418
-Carbohydrates (g)	262.5 (219.8e304.3)	225.5 (148.2e292.5)	0.423
-Dietary fiber (g)	20.7 (16.5e29.5)	20.1 (13.6e28.4)	0.749
-Sucrose (g)	1.8 (0.2e8.3)	1.5 (0.8e10.1)	0.631
-Fat (g)	30.5 (17.2e40.2)	59.0 (38.8e89.5)	0.025
–Saturated (g)	8.8 (3.1e10.3)	20.3 (14.1e29.7)	0.004
–Monounsaturated (g)	5.7 (1.7e10.7)	16.9 (14.0e25.2)	0.004
–Polyunsaturated (g)	4.8 (2.2e9.6)	8.8 (5.3e16.0)	0.150
–Omega-3 (g)	0.2 (0.2e0.4)	0.3 (0.1e0.4)	0.629
–Cholesterol (mg)	32.4 (7.8e48.8)	223.0 (186.2e301.9)	0.004
-Vitamins –A (mcg)	372.3 (64.0e1324.6)	180.3 (97.4e534.7)	0.522
–Pyridoxine (mg)	1.4 (0.6e2.2)	1.2 (0.5e1.9)	0.749
–Choline (mg)	102.9 (47.6e233.7)	197.8 (131.4e397.8)	0.150
–Total folate (mcg)	179.4 (130.7e276.2)	349.1 (190.0e405.0)	0.109
–B12 (mcg)	3.0 (0.1e5.7)	4.5 (2.7e7.9)	0.200

–C (mg)	139.0 (33.9e209.0)	82.6 (43.5e110.3)	0.262
–D (IU)	16.2 (2.9e22.7)	58.5 (36.4e128.5)	0.004
–E (IU)	2.4 (0.8e3.5)	4.4 (2.6e7.5)	0.109
–K1 (mcg)	43.8 (28.6e75.0)	23.3 (13.5e68.0)	0.262
-Minerals –Calcium (mg)	602.6 (300.9e2577.43)	671.4 (400.4e973.0)	0.749
–Iron (mg)	9.7 (8.2e42.8)	13.3 (9.3e21.5)	0.748
–Magnesium (mg)	164.1 (111.7e419.0)	220.3 (166.2e276.8)	0.522
–Phosphorus (mg)	517.1 (379.3e1483.1)	962.8 (799.0e1426.5)	0.200
–Potassium (mg)	1500.7 (1116.6e2621.1)	1800.1 (1539.4e2652.8)	0.337
–Selenium (mcg)	37.1 (23.2e56.5)	93.8 (57.4e146.5)	0.004
–Zinc	7.1 (4.4e17.8)	12.7 (7.7e19.4)	0.423

HCU: classical homocystinuria; BMI: body mass index; tHcy: total homocysteine; Met: methionine; -: not measured.
Numerical variables summarized as median (interquartile range) and compared using the Mann-Whitney U test. Categorical variables were compared using Fisher's exact test. Significant p-values (<0.05) highlighted in bold.
[a] Formula and diet, but not vitamin supplementation, were taken into account for analysis.
[b] In the previous 6 months.

Table 2. Characteristics of patients with HCU included in the study.

Patient	**H1[b]**	**H2**	**H3**	**H4**	**H5**	**H6**
Age (years)	4	19	23	28	28	41
Age at treatment onset (years)	4	13	2	8	6	14
Sex (M/F)	M	M	M	F	M	M
Genotype	p.Glu176Lys/p.Val533Gly	p.Thr191Met/c.209þ1delG	p.Ile95Thr/p.Ile95Thr	p.Asp376Asn/p.828ins104,737del9	p.Asn149fs/p.Asn149fs	p.Gly85Arg/p.Gly85Arg
Consanguinity	No	No	Yes	No	Yes	Yes
Pyridoxine responsiveness	No	No	No	No	No	No
tHcy (mmol/L)	48.6	102.6	36.2	96.3	75	85
1-year tHcy median (mmol/L)	-	110.1	40.5	117.1	146.9	168.6
Met (mmol/L)	29.5	690.2	43.5	217.1	357.2	383.5
Metabolic control[a] (good/ poor)	Good	Poor	Good	Poor	Poor	Poor
Clinical manifestations						
-Ocular	Yes	Yes	Yes	Yes	Yes	Yes
-Skeletal	Yes	Yes	Yes	No	Yes	Yes
-CNS	Yes	No	Yes	Yes	Yes	Yes
-Vascular	No	No	Yes	No	Yes	No

Current treatment						
-Met-restricted diet	Yes	Yes	Yes	Yes	Yes	Yes
-Metabolic formula	Yes	No	No	No	Yes	Yes
-Pyridoxine	Yes	Yes	Yes	Yes	No	Yes
-Betaine	No	Yes	Yes	No	Yes	Yes
-Folate	Yes	Yes	Yes	Yes	Yes	Yes
-Vitamin B12	Yes	No	No	Yes	Yes	Yes

HCU: classical homocystinuria. tHcy: total homocysteine. Met: methionine. CNS: central nervous system.
[a] Metabolic control was defined on the basis of 1-year tHcy medians before the stool sample collection. If the tHcy level was within target, metabolic control was considered good. As all the patients were nonresponsive to pyridoxine, the target tHcy level was <100 mmol/L for the whole sample [3].
[b] The patient had been under treatment for 4 months. The patient had only one measure of tHcy and Met under treatment.

Table 3. Microbial biomarkers differentiating HCU patients from healthy controls.

Closest microbial relative	**Median of the centered log-ratio of the number of sequences**		**Effect size**	**p-Value**
	HCU (n = 6)	**Controls (n = 6)**		
Eubacterium coprostanoligenes group	7.037	-1.271	1.026	0.061
Alistipes[a]	4.197	11.723	-1.107	**0.021**
Alistipes[a]	-1.262	3.263	-1.138	**0.044**
Family XIII UCG-001	0.622	5.299	-1.252	**0.020**
Parabacteroidetes	2.147	8.660	-1.308	**0.009**

HCU: classical HCU: classical homocystinuria.
Positive values of effect size indicate greater abundance in the HCU patients, whereas negative values indicate greater abundance in controls.
As the data were centered log-ratio transformed, OTUs with extremely low abundance appear to be negative.
Significant p-values (<0.05) highlighted in bold.
[a] Two different OTUs corresponding to the same genus were identified, but identification at the species level was not possible.

APÊNDICE G – Microbima do mecônio e sua relação com crescimento neonatal e adequação da circunsferência craniana em prematuros

Meconium microbiome and its relation to neonatal growth and head circumference catch-up in preterm infants

Ana Carolina Terrazzan Nutricionist[1], Renato S. Procianoy[2], Luiz Fernando Wurdig Roesch[3], Andrea Lúcia Corso[2], Priscila Thiago Dobbler [3], Rita C. Silveira[1,2]*

1 Postgraduate Program in Child and Adolescent Health–PPGSCA, Federal University of Rio Grande do Sul–UFRGS, Porto Alegre, Brazil
2 Neonatal Intensive Care Unit Hospital de Clínicas de Porto Alegre, RS, Porto Alegre, Brazil
3 Interdisciplinary Center for Biotechnology Research–CIP-Biotec Federal University of Pampa, São Gabriel, Rio Grande do Sul, Brazil

* drarita.c.s@gmail.com

DOI: https://doi.org/10.1371/journal.pone.0238632
Received: December 11, 2019
Accepted: August 20, 2020
Published: September 21, 2020

Abstract

The purpose was identify an association between meconium microbiome, extra-uterine growth restriction, and head circumference catch-up. Materials and methods: Prospective study with preterm infants born <33 weeks gestational age (GA), admitted at Neonatal Unit and attending the Follow-Up Preterm Program of a tertiary hospital. Excluded out born infants; presence of congenital malformations or genetic syndromes; congenital infections; HIV-positive mothers; and newborns whose parents or legal guardians did not authorize participation. Approved by the institution's ethics committee. Conducted 16S rRNA sequencing using PGM Ion Torrent meconium samples for microbiota analysis. Results: Included 63 newborns, GA 30±2.3 weeks,

mean weight 1375.80±462.6 grams, 68.3% adequate weight for GA at birth. *Polynucleobacter* (p = 0.0163), Gp1 (p = 0.018), and *Prevotella* (p = 0.038) appeared in greater abundance in meconium of preterm infants with adequate birth weight for GA. Thirty (47.6%) children reached head circumference catch-up before 6 months CA and 33 (52.4%) after 6 months CA. *Salmonella* (p<0.001), *Flavobacterium* (p = 0.026), and *Burkholderia* (p = 0.026) were found to be more abundant in meconium in the group of newborns who achieved catch-up prior to 6th month CA. Conclusion: Meconium microbiome abundance was related to adequacy of weight for GA. Meconium microbiome differs between children who achieve head circumference catch-up by the 6th month of corrected age or after this period.

Introduction

The balance between the host and intestinal microbes is protective to health [1–3]. Gut microbiota is essential for suitable nutrient absorption, energy storage, and immune response, and it's also responsible for multiple metabolic tasks, including production of essential vitamins, fermentation and breakdown of oligosaccharides and production of short-chain fatty acids and gases [4, 5]. However, for the microbiota to perform such tasks, the host must maintain a favorable gut environment.

The mechanisms by which microbiota formation occurs via placenta and amniotic fluid are still not fully elucidated. Some studies support the hypothesis that fetal intestinal microbiome is derived from the swallowing of amniotic fluid containing bacteria [6, 7]. The mechanism related to this hypothesis is that maternal bacteria might translocate through maternal bloodstream, achieving other organs and systems, reaching amniotic fluid also [8] Yet, more studies are needed in order to better elucidate mechanisms involved in microbiota formation via placenta and amniotic fluid [9, 10].

There is evidence of a gut-brain axis, linking gut microbiota and the development of nervous system function. The maintenance of this bidirectional communication between central and enteric nervous system evolves endocrine, immune and neuronal pathways and it's essential for neurological development and brain growth [11, 12].

For many reasons preterm infants are also high-risk infants for impaired growth, nutrition and neurodevelopment; and the possible early dysbiosis might interfere on microbiota metabolic capacity, and consequently alter nutrient absorption, influencing growth and neurodevelopment [1, 13].

A better understanding of microbiome variation may allow the early detection of a subpopulation of preterm infants at higher risk for growth and developmental impairment during follow-up. Thus, we aimed to identify and describe the composition of the microbiota of the first meconium of preterm infants. We also aimed to verify if there was an association between microbiota composition with restricted extra-uterine growth and with head circumference catch-up after discharge, both important growth variables that may influence the neurodevelopmental outcomes.

Material and methods

The study was approved by the Institutional Ethics Committee of Hospital de Clinicas de Porto Alegre and Brazilian review board. All mother or legal guardian had provided written informed consent. This was a prospective cohort study including preterm infants gestational age <33 weeks, born and admitted at the Neonatal Unit and attending the Follow-Up Preterm Program of a tertiary hospital in Porto Alegre, RS. Infants born in another hospital, presence of congenital malformations or genetic

syndromes, congenital infections, and HIV+ mother were exclusion criteria. Data collection started following Institution Ethics Committee approval (140009 – n˚1.388.950). Clinical data and sample characterization were prospectively recorded and associated to meconium microbiome sequencing data bank. Maternal variables studied were: maternal age, mode of delivery, maternal antibiotics, presence of urinary tract infections (urine culture test positive and clinical signs), or clinical chorioamnionitis (maternal fever, uterine hypertonia, malodorous or purulent amniotic fluid, maternal leukocytosis or fetal tachycardia), preeclampsia, and gestational diabetes. Preeclampsia was defined as presence of hypertension (blood pressure > 140/90 mmHg after 20 weeks of gestation with significant proteinuria). For gestational diabetes, fasting was 92g/dL or glycemia of 153 g/dL following oral glucose tolerance test, with onset during pregnancy. Neonatal variables: gender, birth weight, gestational age (determined by the best obstetrical estimate, including first trimester ultrasound and/or last menstrual period date, confirmed by pediatric physical examination immediately after birth), being appropriate-for-gestational-age (AGA), small-for-gestational-age (SGA: below the 10th percentile according to reference curve), intrauterine growth restriction (below 3rd percentile). We also looked at hospitalization data to verify periintraventricular leukomalacia, necrotizing enterocolitis, early and late sepsis, hospitalization after discharge, and use of anticonvulsant.

Following NICU discharge, patients were referred to the Follow-Up Program. According to the routine of the institution, all children have monthly appointments up to 6 months of corrected age. Routine also includes anthropometric measurement (weight, length, head circumference). For this study, we evaluated head circumference at 2, 4, and 6 months corrected age in order to identify those patients for whom catch-up head circumference was achieved before or after 6 months corrected age. Catch-up was defined as a 0.67 z-score variation between two consecutive z-scores [14]. Fenton Growth Calculator for Preterm Infants (2013) [15] was used to generate birth data z-scores, as well as to determine adequacy of weight for gestational age; and WHO Anthro, 3.2.2 version (2011) was used for z-scores from follow-up period. Both software take into account gender and age, with age being corrected for preterm infants. Standardized equipment for measuring the infants was used by a trained researcher (ACT). Weight was measured using a digital scale, accurate to within 5g (ELP, 25BBA, Balmak*1*), with the infant wearing no clothes. Length was measured to the nearest centimeter in horizontal position using a length board accurate to 0.1 cm, with the infant lying down. Head circumference was measured using a non-stretch tape, accurate to 0.1 cm, placed on the broadest part of the forehead above eyebrows, above the ears, and around the most prominent part of the back of the head.

Feeding practices, regarding type of milk the infants were receiving (mother's milk, infant formula, or cow's milk) were evaluated, from hospital discharge up to six months corrected age.

Meconium collection samples

After the mother or legal guardian had provided written informed consent, the first meconium passed by the infant was collected from diaper in sterile conditions, immediately stored at -80˚C in a cryogenic storage Dewar, and transported to a laboratory where microbial DNA extraction and microbial community composition analysis was performed. This collection occurs mandatorily before the newborn

receives any enteral feeding, as some studies suggest differences in microbial colonization between breastfed infants and formula-fed infants [16].

Microbial DNA extraction, amplification, and sequencing

Microbial DNA was isolated from 180 mg of each meconium sample using the QIAamp Fast DNA Stool Mini Kit (Qiagen, Valencia, CA, USA), in accordance with manufacturer instructions. DNA quality was verified by spectrophotometry in a NanoVue™ system (GE Healthcare, Chicago, IL, USA). All DNA samples were stored at -80˚C until use. V4 region of 16S rRNA gene was amplified and sequenced using ION PGM™ Ion Torrent (Thermo Fisher Scientific, Waltham, MA, USA), with primers 515F and 806R. Multiple samples were amplified by polymerase chain reaction (PCR) using barcoded primers linked to adapter "A" sequence (5^0-CC ATCTCATCCCTGCGTGTCTCCGACTCAG-3^0) and "P1" sequence (5^0-CCTCTCTATGGGCAGTCGGTGAT-3^0) to obtain a primer sequence composed for the A-barcode-806R and P1515F adapter and primers. PCR reaction final volume was 25 µL. Each mix consisted of 2U Platinum*1* Taq DNA High Fidelity Polymerase (Invitrogen, Carlsbad, CA, USA), 4 µL 10X High Fidelity PCR Buffer, 2 mM MgSO4, 0.2 mM dNTPs, 0.1 µM of both primers described above, 25 µg UltraPure BSA (Invitrogen, Carlsbad, CA, USA), and approximately 50 ng of template DNA. PCR conditions used were: 95˚C for 5 min, 35 cycles at 94˚C for 45 s, 56˚C for 45 s, and 72˚C for 1 min, followed by 72˚C for 10 min. Resulting PCR products were purified with Agencourt*1* AMPure*1* XP Reagent (Beckman Coulter, La Brea, CA, USA) and quantified using the Qubit Fluorometer kit (Invitrogen, Carlsbad, CA, USA), following manufacturer recommendations.

Finally, reactions were combined in equimolar concentrations to create a mixture composed of amplified fragments of 16S gene from each sample. This composite sample was used for library preparation with OneTouch™ 2 Ion system using the ION™ PGM Template 400 OT2 kit (Thermo Fisher Scientific, Waltham, MA, USA). Sequencing was performed using commercially available ION PGM™ Sequencing 400 kit on an ION PGM™ System, using an Ion 318™ Chip v2, with a maximum of 40 samples per microchip.

Sequence processing for analysis

Fastq files exported from ION PGM™ system were analyzed following recommendations from Brazilian Microbiome Project (BMP) [17], using the BMP Operating System [18]. Briefly, an Operational Taxonomic Unit (OTU) table was compiled using UPARSE pipeline [19] wherein sequences were truncated at 200 base pairs and quality filtered using a maximum expected error cutoff of 0.5. Sequences were clustered into OTUs using a 97% similarity cutoff, and chimeric sequences were removed. Taxonomic classification was performed in QIIME software environment [20], based on UCLUST method, against Greengenes 13.5 database [21], with a confidence limit of 80%. Sampling effort was estimated using Good's coverage formula [22]. For downstream analysis, the data set was filtered by removing Chloroplast/Cyanobacteria sequences and only OTUs with more than 5 sequence reads were kept before rarefying all samples to 5379 sequences each [23].

Functional prediction for the gut microbiome was performed using PICRUSt 24]. For that, the raw 16S rRNA dataset was prepared following the instructions of Langille

et al. (2013) [24]. After quality filtering and trimming, OTUs were picked against the Greengenes [21] database.

Statistical analyses

Data obtained were stored in a database constructed for this specific purpose, using Excel software. Afterwards, data were processed and analyzed using PASW (SPSS) software, 18.0 version (Statistical Package for Social Sciences). Results are expressed as mean ± Standard Deviation (SD), minimum and maximum values, or median and interquartile (p25-p75). Differences between medians were analyzed with Mann-Whitney test. Between-groups differences were analyzed by T test, Qui Square, and ANOVA when more than two groups were analyzed.

Microbiome database was imported into R (R Development Core Team, 2008) to assess structural differences in the microbial community and detect possible confounders; a compositional dissimilarity matrix was generated based on the Bray-curtis distances between samples using the phyloseq package [25]. The matrix was used in a nonparametric Multivariate Analysis of Variance (PERMANOVA) with the Adonis function available in the vegan package [26]. To estimate alpha diversity, microbial dominance and Shannon diversity index were calculated and plotted using the "phyloseq" package [25]. Alpha diversity measurements were tested for normality with Shapiro-Wilk test and variables were compared by Kruskal-Wallis rank sum test. Differential abundance analysis was performed with DESEq2 [27]. The p-values were adjusted for multiple comparisons using the FDR method.

For the functional prediction of the gut microbiota, functions were categorized by the third KEGG Pathway Hierarchy Level and hypothesis testing was performed with two-sided White's non-parametric t-test. Hypothesis testing and plotting were done using STAMP [28] Only features with a difference in proportion of 0.1 (Effect size > 0.1) were considered as active.

Results

Eighty-seven samples were collected. Eleven were excluded for not being sterile, six did not have enough material for analysis, and in seven it was not possible to determine microbial DNA. In total, for this study we analyzed 63 meconium samples of preterm infants, of whom 30 (47.6%) were boys, with mean gestational age of 30±2.3 weeks. Mean weight, length, and head circumference at birth were 1375.80±462.6 grams, 38.0±4.0 centimeters, and 27±2.7 centimeters, respectively. Mean maternal age was 25.95±6.5 years, and 45 (71.4%) infants were delivered by C-section. Prevalence of preeclampsia, gestational diabetes, and urinary tract infection was 16(25.4%), 7 (11.1%), and 7 (11.1%), respectively. At discharge, mean gestational age was 38±3 weeks and mean weight was 2573.05±292.18 grams.

Forty-nine (68.3%) were AGA, and of these 57.14% (n = 36) were also discharged AGA. Thirteen (20.63%) were born AGA and were SGA at discharge. Twelve (19.4%) were born SGA and were discharged also SGA. Only two (3.17%) of those born SGA were LGA at discharge (this group was excluded from data analysis, because of its limited size). The growth pattern was significantly higher among the AGA neonates. Regarding use of breast milk or formula during the hospital stay, no difference was found according to adequacy of weight for gestational age at birth and discharge. (Table 1).

In total, we identified 5,309 different OTUs across all samples, of these, 16 OTUs had mean abundance higher than 1%. Microbial composition was similar when compared according to weight at birth and at discharge. Alpha diversity measurements between groups AGA-AGA vs. AGA-SGA vs. SGA-SGA were similar (Observed OTUs, p-value = 0.745) and Shannon Diversity Index, p-value = 0.127 (Fig 1).

The overall microbial composition at phylum level according to weight adeqacy at birth is presented in Fig 2A, and at discharge in Fig 2B. Four phyla were found to be dominant across the samples irrespective of weight adequacy at birth or delivery. They were *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria*. On average, infants in the SGA group at birth or discharge had higher *Firmicutes* while those in the AGA group had higher *Proteobacteria* then their couterparts.

When compared to the SGA at birth group, those born AGA had an increased abundance of OTUs belonging to genus *Polynucleobacter* (p = 0.0163), phylum *Proteobacteria*, *Gp1* (p = 0.018) phylum *Acidobacteria*, and *Prevotella* (p = 0.038) phylum *Bacteriodetes* (Fig 3A).

Between most abundant OTUs observed, when comparing preterm AGA or SGA at discharge, those OTUs belonging to *Escherichia fergusoni* (p = 0.014) and *Streptococcus dentisani* (p = 0.043) genus were more abundant in the AGA at discharge group, and this difference was statistically significant. By contrast, the SGA at discharge group presented increased abundance of *Prevotella copri* (p = 0.002), *Roseburia inulinivorans* (p = 0.003), *Staphylococcus sp*. (p = 0.003), *Staphylococcus capitis* subsp. *Capitis* (p = 0.004), *Sutterella stercoricanis* (p = 0.027), *Corynebacterium tuberculostearicum* (p = 0.033), and *Ruminococcaceae* (p = 0.043) (Fig 3B)

Regarding head circumference (HC) catch-up growth, 30 (47.6%) infants completed HC catch-up growth by the age of 6 months corrected age and 33 (52.4%) after 6 months of corrected age. Also, catch-up occurred independently of weight adequacy for gestational age at birth or at discharge. There were no statistically significant differences regarding clinic variables at birth, sepsis during NICU stay, use of anticonvulsant, and rehospitalizations after discharge. As expected, the group that completed HC catch-up growth by the age of 6 months corrected age had higher z-score and measures of weight and head circumference between 2 and 6 months of corrected age. There was a difference between groups only at 6 months of corrected age, with a higher number of infants receiving infant formula in those whose HC catchup growth was completed by the 6th month of corrected age (Table 2).

According to the PERMANOVA (Table 3) there was no statistically significant difference for microbial beta diversity between infants with early HC catch-up growth (up to 6 months) and late HC catch-up growth (after 6 months) (p = 0.093). However, after analyzing differences in microbial alpha diversity, Shannon Index was statistically significant (p = 0.045), indicating more microbial diversity in meconium from infants who had their HC catch-up growth later, after 6 months of corrected age (Fig 4). Pre-eclampsia was not associated to differences in the meconium microbiota (p-value = 0.64).

The overall microbial composition at phylum level within groups with the head circumference catch-up by 6 months and after 6 months is presented in Fig 5B. Four phyla were found to be dominant within the samples irrespective of the group. They were *Proteobacteria*, *Bacteroidetes*, *Firmicutes*, and *Actinobacteria*.

Differential abundance analysis showed increased abundance of *Bacterioidetes* and *Proteobacteria* phylum, with OTUs belonging to Salmonella (p<0.001),

Flavobacterium (p = 0.026), and *Burkholderia* (p = 0.026) genus being the most abundant in meconium from infants who achieved HC catch-up growth by the 6th month of corrected age. *Prevotella* (p = 0.005), *Enhydrobacter* (p = 0.036), *Brevundinomonas* (p = 0.043), *Bradyhizobium* (p = 0.018), and *Acinetobacter* (p = 0.007) genus were more abundant in meconium of those infants who achieved HC catch-up growth after 6 months of corrected age (Fig 5A and 5B).

In order to better understand the differences of the gut microbiota in relation with the time of HC catch up, we also explored the functional prediction of these communities, using PICRUSt [24]. Infants with HC catch up before the 6th month of corrected age presented a microbiota with higher predicted genes relateted with transportation (Transporters and ABC transporters), while those with HC cacth up after 6 months had more genes related with sugar and amino acid metabolism (Fig 6).

When analysing functional gene prediction according with weight adequacy at birth or discharge, there were no significant differences, considering the treshold of effect size > 0.1 (S1 Fig).

Discussion

Increased abundance of OTU belonging to *Prevotella*, *Polynucleobacter*, and *Gp1* genus in preterm infants born AGA was observed. Preterm AGA at discharge showed increased abundance of OTU belonging *to Escherichia fergusoni* and *Streptococcus dentisani* genus. We also found more abundance of OTUs *Salmonella*, *Flavobacterium*, and *Burkholderia* genus in the meconium of infants who achieved HC catch-up growth by the 6th month of corrected age. There are few studies with similar data; the great majority of studies consider the microbiome of fullterm infants, and those that assess prematurity take into account only gestational age, without relating it to adequacy of weight for gestational age [29, 30].

Ardissone et al. (2014) [31] found several taxonomic families within *Firmicutes* phylum correlated to gestational age, including *Staphylococcus* genus, which were most abundant among preterms born at <33 gestational weeks. Jacquot et al. [32] found an association between gestational age less than 28 weeks and lower microbial diversity score at first week of life, where *Staphylococcus spp* genus was found in 67% of the patients. The authors also enlight that although it is clear that preterm infants can also present an important Staphylococcus colonization, these infants are at higher risk of late onset sepsis related to coagulase negative Staphylococcus during the first weeks of life [32].

Itani et al. (2017) [33] also described increased *Staphylococcus* abundance in feces from preterm infants less than 33 weeks of gestational age. Our data represent meconium microbiome, and we observed significantly increased *Staphylococcus* genus abundance in preterm infants who were SGA at discharge, with hospital discharge being equivalent to the term of gestational age. We hypothesize that besides gestational age, adequacy of weight for gestational age at birth is also related to microbial community structure. Also, although *Staphylococcus* colonization is a normal characteristic of healthy gut microbiota [34], we understand that a microbiota more abundant in *Staphylococcus* might interfere for nutrient absorption e metabolism, leading to a worse weigh gain during NICU stay, despite the efforts of nutrition therapy.

Nataro and Guerrant (2017) [35] suggest that *Prevotella* genus is associated to better growth, while *Streptococcus lutetiensis* and *Escherichia coli* are associated to growth failure, but they do not distinguish preterm from full-term infants. In our study,

AGA at birth presented significant higher abundance of *Prevotella* genus, we believe this may reflect fetal period, once this microbe has been associated to improved glucose metabolism by promoting increased glycogen storage [36].

On the other hand, in contrast to Nataro and Guerrant (2017) [35] results, when we evaluate the adequacy of weight for gestational age at discharge, AGA preterms were the ones who presented increased *Escherichia fergusoni and Streptococcus dentisani* abundance in meconium, while SGA at discharge preterms presented increased *Prevotella copri* abundance in meconium. Through our results, we cannot infer about microbiota changes during the hospital stay, however, we have some hypothesis: a) Those infants with better growth (AGA at discharge) possibly had earlier contact with their parents and better evolution of dietary acceptance, both factors that can favor benefic changes in the microbiota. b) As we already mentioned, SGA infants at discharge also had abundant *Staphylococcus* in meconium and maybe during hospital stay this microbe was more resistant or had more impact host metabolism than *Prevotella copri*, influencing to the worse weigh gain. We understand that other external factors act together with the microbiome, being important influences in weight gain during hospital stay. Future studies, evaluating progressive changes in the microbiota, in association with dietary characteristics may answer this hypothesis.

Preterm infants miss an important phase of brain growth and maturation, which would occur during the last trimester of pregnancy [37]. During this phase the cortical gray matter is already matured, but some of the most important developing stages such as the increase in the complexity of connections, axons, glial cells, and oligodendrocytes in the withe matter, will be concluded as the 3rd trimester goes by [38, 39]. Therefore prematurity is associated with neurodevelopmental disability, with long term effects [3, 40, 41]. Catch down during hospital stay and during the first months of life are associated to increased risk of neurologic impairment in preterm infants, nevertheless the mechanisms that guarantee this association are not yet completely elucidated [6]. On the other hand, catch-up growth of head circumference in the first years of life is a protective factor for neurodevelopment, being associated to better cognitive and behavioral performance in early childhood [42, 43].

Taken together, neurological immaturity and a dysbiotic and immature gut, both associated with prematurity may disrupt the bidirectional communication between the nervous system and enteric cells, leading to altered signaling and neurological development, and also altered immune responses [3, 44, 45].

In the present study we were able to verify a higher microbial biodiversity in meconium from those children who had head circumference catch-up growth after 6 months of corrected age, with *Prevotella*, *Enhydrobacter*, *Brevundinomonas*, *Bradyhizobium*, and *Acinetobacter* being the most prevalent genus observed in the group. Moreover, in the group of infants whose head circumference catch-up growth was completed until 6 months of corrected age, *Salmonella*, *Flavobacterium*, and *Burkholderia* were most abundant. Community functional prediction suggests that the gut mictobiota of infants with head catch up until the 6th month presented higher presence of transporter genes, including ABC transporters, while infants with head catch up after the 6th month presented more genes predicted to be involved in the metabolism of complex carbohydrates, such as starch, and amino acids. This difference might influence energy intake from different sources and might influence growth.

Despite several studies aiming to explain the role of microbiome in the gut–brain axis, interactions between neurologic mechanisms and microbiome development in preterm infants are not well understood [38]. To our knowledge, this is the first study investigating meconium microbial composition and its association to head circumference catch-up growth in preterm infants. We suggest more studies should be conducted so that the pathways of this relationship may be better understood.

Guney Varal et al (2018) [46] conducted a study with preterm infants, using a prepared commercial symbiotic solution administered with enteral nutrition. Their results show a lower odd to lower head circumference growth in the study group. Wejryd et al (2018) [47] related supplementation with *L*. *reuteri* to better head circumference growth, also during hospital stay. Both studies corroborate the hypothesis that a favorable gut microbiota might enhance the chances of achieving better neurodevelopment/ growth via the beneficial effects on cytokines, nervous and immune system. However, a recent systematic review conducted by Hortensius et al (2019) [48] suggests that until the present, despite the positive results on head growth, there is no significant data regarding the effect of supplementation with probiotics on neurodevelopmental outcome was found. Therefore, it's indeed necessary more follow up studies.

Experimental studies with germ-free mice have observed systemic inflammation and neuroinflammation in the offspring as well as impaired myelination and blood–brain barrier formation. These studies suggest a relationship between microbial colonization, immune system, and brain activity, as well as an essential role for microbiota in neural, structural, and functional development [45, 49]. Although animal model studies have already clearly elucidated the role of gut microbiota in childhood development programming, and there is a window of opportunity in which microbiota can affect physiological function of several systems, with long-term consequences, there have been only a limited number of studies with humans, specifically preterm newborns, that would enable complete understanding of processes involving microbiome and neurologic development [50].

Several factors such as infection, neurologic impairment, diet, and antibiotic use are crucial in ensuring growth. In our study the groups were similar for sepsis. However, post-discharge hospitalizations, anticonvulsant treatment, and milk feeding were different at 6 months of corrected age, which may directly interfere with growth, neurodevelopment, and microbial colonization. Thus, we cannot infer if meconium microbiota was the only determinant factor for head circumference catch-up growth. Yet, considering the intimate relationship between brain and gut [51], we suggest identifying microbiome variations associated and predisposing to accelerated head circumference catch-up growth as a relevant tool for clinical practice in the context of improving care and future health of preterm infants.

It is worth mentioning that food directly influences bacterial flora establishment, and human milk is a greater promoter of *Bifidobacteria* and *Lactobacillus* colonization when compared to formula based on cow's milk [52]. Oligosaccharides (HMO) present in breast milk, which are complex glycans and not digestible by humans, are the main microbiome substrate, especially for *Bifidobacteria*, playing a fundamental role for beneficial bacterial community proliferation in children's gut, due to both probiotic and prebiotic effects, highlighting the importance of promoting breastfeeding in the NICU environment [52–54].

It was a challenge to analyze the relationship between microbiome, born SGA or AGA, and head circumference catch-up growth, since there are so few studies and many unanswered questions. This study encountered limitations, such as the lack of microbiome data at discharge and follow up, which could give us more information regarding changes that occurred during hospital stay. We also understand the sample size as a limitation of this study; on the other hand, we emphasize the follow-up of preterm infants as strength.

Conclusion

Meconium microbial abundance seems to be related to adequacy of weight for gestational age as well as to weight gain during neonatal period in low-birth-weight preterm infants. Also, abundance of meconium OTUs from infants who achieved early head circumference catch-up growth (defined in this study as up to the 6th month of corrected age) differs from those who had late head circumference catch-up growth (in this study, after 6 months of corrected age). Further studies following changes in microbial colonization, as well as its associations to diet patterns, in order to verify associations between microbiota and medium-term outcomes, may lead to new conduct definitions for clinical practice.

Author Contributions

Conceptualization: Renato S. Procianoy, Luiz Fernando Wurdig Roesch, Rita C. Silveira.
Data curation: Ana Carolina Terrazzan Nutricionist, Andrea Lúcia Corso, Rita C. Silveira.
Formal analysis: Ana Carolina Terrazzan Nutricionist, Luiz Fernando Wurdig Roesch, Priscila Thiago Dobbler.
Funding acquisition: Rita C. Silveira.
Investigation: Ana Carolina Terrazzan Nutricionist, Luiz Fernando Wurdig Roesch, Andrea Lúcia Corso, Rita C. Silveira.
Methodology: Ana Carolina Terrazzan Nutricionist, Renato S. Procianoy, Priscila Thiago Dobbler.
Project administration: Renato S. Procianoy.
Resources: Renato S. Procianoy, Rita C. Silveira.
Software: Priscila Thiago Dobbler.
Supervision: Rita C. Silveira.
Visualization: Andrea Lúcia Corso.
Writing – original draft: Ana Carolina Terrazzan Nutricionist.
Writing – review & editing: Renato S. Procianoy, Luiz Fernando Wurdig Roesch, Rita C. Silveira.

References

1. Krajmalnik-Brown R, Ilhan ZE, Kang DW, DiBaise JK. Effects of Gut Microbes on Nutrient Absorption and Energy Regulation. 2012; 27:201–214
2. Robertson RC, Manges AR, Finlay BB, Prendergast AJ. The Human Microbiome and Child Growth– First 1000 Days and Beyond. Trends in Microbiology 2019; 27: 131–147 https://doi.org/10.1016/j.tim. 2018.09.008 PMID: 30529020

3. Lu J, Claud E. Connection between gut microbiome and brain development in preterm infants. Dev Psychobiol. 2019; 61: 739–751 https://doi.org/10.1002/dev.21806 PMID: 30460694
4. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, Tuohy K. Gut microbiota functions: metabolism of nutrients and other food componentes. Eur J Nutr. 2018; 57:1–24
5. Turroni F, Milani C, Durant S, Lugli GA, Bernasconi S, Margolles A, et al. The infant gut microbiome as a microbial organ influencing host well-being. Ital J Pediatr. 2020. 46:16 https://doi.org/10.1186/ s13052-020-0781-0 PMID: 32024556
6. Jime´nez E, Marı´n ML, Martı´n R, Odriozola JM, Olivares M, Xaus J, Ferna´ndez L, e tal. Is meconium from healthy newborns actually sterile? Res. Microbiol. 2008; 159: 187–193. https://doi.org/10.1016/j. resmic.2007.12.007 PMID: 18281199
7. Mshvildadze M, Neu J, Shuster J, Theriaque D, Li N, Mai V. Intestinal Microbial Ecology in Premature Infants Assessed Using Non-Culture Based Techniques. J Pediatr. 2010; 156: 20–25. https://doi.org/ 10.1016/j.jpeds.2009.06.063 PMID: 19783002
8. Collado Maria Carmen & Rautava Samuli & Aakko, Juhani & Isolauri, Erika & Salminen, Seppo. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. Scientific Reports. 2016; 6: 23129 https://doi.org/10.1038/srep23129 PMID: 27001291
9. Jimenez E, Fernandez L, Marin ML, Martı´n R, Odriozola JM, Nueno-Palop C, et al. Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. Curr Microbiol. 2005; 51:270–274 https://doi.org/10.1007/s00284-005-0020-3 PMID: 16187156
10. Perez-Muñoz ME, Arrieta M-C, Ramer-Tait AE, Walter J. A critical assessment of the “sterile womb” and “in utero colonization” hypotheses: implications for research on the pioneer infant microbiome. Microbiome. 2017; 5:48 https://doi.org/10.1186/s40168-017-0268-4 PMID: 28454555
11. Rogers GB, Keating DJ, Young RL, Wong M-L, Licinio J, Wesselingh S. From gut dysbiosis to altered brain function and mental illness: mechanisms and pathways. Molecular Psychiatry. 2016; 21:738–748 https://doi.org/10.1038/mp.2016.50 PMID: 27090305
12. Niccolai E, Boem F, Russo E, Amedei A. The Gut–Brain Axis in the Neuropsychological Disease Model of Obesity: A Classical Movie Revised by the Emerging Director “Microbiome”. Nutrients. 2019; 11:156.
13. Henderickx JGE, Zwittink RD, van Lingen RA, Knol J and Belzer C. The Preterm Gut Microbiota: An Inconspicuous Challenge in Nutritional Neonatal Care. Front. Cell. Infect. Microbiol. 2019; 9:85. https:// doi.org/10.3389/fcimb.2019.00085 PMID: 31001489
14. Ong KK, Ahmed ML, Emmett PM, Preece MA, Dunger DB. Association be-tween postnatal catch-up growth and obesity in childhood: prospective cohort study. BMJ. 2000; 320:967–71. https://doi.org/10. 1136/bmj.320.7240.967 PMID: 10753147
15. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fen-ton growth chart for preterm infants. BMC Pediatr. 2013; 13:59. https://doi.org/10.1186/1471-2431-13-59 PMID: 23601190
16. Fanaro S, Chierici R, Guerrini P, Vigi V. Intestinal microflora in early infancy: composition and development. Acta Paediatr Suppl 2003; 91:48–55. https://doi.org/10.1111/j.1651-2227.2003.tb00646.x PMID: 14599042

17. Pylro VS, Roesch LF, Ortega JM, do Amaral AM, To´tola MR, Hirsch PR, et al. Brazilian Microbiome Project Organization Committee. Brazilian Microbiome Project: revealing the unexplored microbial diversity—challenges and prospects.Microb Ecol. 2014; 67: 237. https://doi.org/10.1007/s00248-0130302-4 PMID: 24173537
18. Pylro VS, Morais DK, de Oliveira FS, Dos Santos FG, Lemos LN, Oliveira G, et al. BMPOS: a Flexible and User-Friendly Tool Sets for Microbiome Studies. Microbial Ecology.2016; 72: 443–447. https://doi.org/10.1007/s00248-016-0785-x PMID: 27220974
19. Edgar R.C. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nature Methods. 2013; 10:996–8 https://doi.org/10.1038/nmeth.2604 PMID: 23955772
20. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. Nature Methods. 2010;
21. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012; 6: 610–618. https://doi.org/10.1038/ismej.2011.139 PMID: 22134646
22. Good IJ. The population frequencies of species and the estimation of popula-tion parameters. Biometrika.1953; 40: 237±264.
23. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW. Rethinking microbial diversity analysis in the high throughput sequencing era. Journal of Microbiological Methods. 2011; 86:42–51 https://doi.org/10.1016/j.mimet.2011.03.014 PMID: 21457733
24. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes J. et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology. 2013;1–10 https://doi.org/10.1038/nbt.2482 PMID: 23302909
25. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE. 2013; 8: e61217. https://doi.org/10.1371/journal.pone.0061217 PMID: 23630581
26. Oksanen J, Blanchet G, Kindt R, Legendre P, O'Hara R, Simpson G, et al. Vegan: Community Ecology Package. 2011.
27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014; 15:550 https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281
28. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014; 30: 3123–3124. https://doi.org/10.1093/bioinformatics/btu494 PMID: 25061070
29. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM et al. Patterned progression of bacterial populations in the premature infant gut. Proc Natl Acad Sci. 2014; 111:12522–7. https://doi.org/10.1073/pnas.1409497111 PMID: 25114261
30. Korpela K, Blakstad EW, Moltu SJ, Strømmen K, Nakstad B, Rønnestad AE, et al. Intestinal microbiota development and gestational age in preterm neonates. Scientific Reports. 2018; 8: 2453 https://doi.org/10.1038/s41598-018-20827-x PMID: 29410448

31. Ardissone AN, de la Cruz DM, Davis-Richardson AG, Rechcigl KT, Li N, Drew JC et al. Meconium Microbiome Analysis Identifies Bacteria Correlated with Premature Birth. PLoS ONE. 2014; 9: e90784. https://doi.org/10.1371/journal.pone.0090784 PMID: 24614698
32. Jacquot A, Neveu D, Aujoulat F, Mercier G, Marchandin H, Jumas-Bilak E, et al. Dynamics and Clinical Evolution of Bacterial Gut Microflora in Extremely Premature Patients. J Pediatr. 2011; 158:390–6 https://doi.org/10.1016/j.jpeds.2010.09.007 PMID: 20961563
33. Itani T, Ayoub Moubareck C, Melki I, Rousseau C, Mangin I, Butel MJ, et al. Establishment and development of the intestinal microbiota of preterm infants in a Lebanese ter-tiary hospital. Anaerobe.
2017; 43:4–1 https://doi.org/10.1016/j.anaerobe.2016.11.001 PMID: 27833033
34. Rinninella E, Raoul P, Cintoni M, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019; 7:14
35. Nataro J; Guerrant R. Chronic consequences on human health induced by microbialpathogens: Growth faltering among children in developing countries. Vaccine. 2017. 35: 6807–6812 https://doi.org/10.
1016/j.vaccine.2017.05.035 PMID: 28549806
36. Kovatcheva-Datchary P, Nilsson A, Akrami R, Lee YS, De Vadder F, Arora T. Dietary Fiber-Induced Improvement in Glucose Metabolism Is Associated with Increased Abundance of Prevotella. Clinical And Translational Report. 2015; 22: 971–982.
37. Cai S., Zhang G., Zhang H. et al. Normative linear and volumetric biometric measurements of fetal brain development in magnetic resonance imaging. Childs Nerv Syst. 2020. https://doi.org/10.1007/s00381020-04633-3
38. Lu L; Claud EC. Intrauterine Inflammation, Epigenetics, and Microbiome In-fluences on Preterm Infant Health. Current Pathobiology Reports. 2018. 6:15–21 https://doi.org/10.1007/s40139-018-0159-9 PMID: 29938128
39. Volpe JJ. The encephalopathy of prematurity—brain injury and impaired brain development inextricably intertwined. Semin Pediatr Neurol. 2009; 16: 167–78 https://doi.org/10.1016/j.spen.2009.09.005 PMID: 19945651
40. Volpe JJ. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. Lancet Neurol. 2009; 8: 110–124 https://doi.org/10.1016/S1474-4422(08)70294-1 PMID: 19081519
41. Cheong J.L.Y., Burnett A.C., Treyvaud K. et al. Early environment and long-term outcomes of preterm infants. J Neural Transm. 2020; 127: 1–8 https://doi.org/10.1007/s00702-019-02121-w PMID: 31863172
42. Belfort MB, Rifas-Shiman SL, Sullivan T, Collins CT, McPhee AJ, Ryan P, et al. Infant growth before and after term: effects on neurodevelopment in pre-term infants. Pediatrics 2011; 128:e899–906. https://doi.org/10.1542/peds.2011-0282 PMID: 21949135
43. Ramel SE, Demerath EW, Gray HL, Younge N, Boys C, Georgieff MK. The relationship of poor linear growth velocity with neonatal illness and two year neurodevelopment in preterm infants. Neonatology
2012; 102:19–24 https://doi.org/10.1159/000336127 PMID: 22441508
44. Ba¨ckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Resource. 2015; 17:690–703

45. Lu J, Lu L, Yu Y, Cluette-Brown J, Martin CR, Claud EC. Effects of Intestinal Microbiota on Brain Development in Humanized Gnotobiotic Mice. Scientific Reports. 2018.:1–16 https://doi.org/10.1038/s41598017-17765-5 PMID: 29311619
46. Guney Varal I, Koksal N, Ozkan H, Bagci O, Dogan P. Potential use of multi-strain synbiotics for improving postnatal head circumference. Pak J Med Sci. 2018; 34:1502–1506 https://doi.org/10.12669/pjms. 346.16107 PMID: 30559812
47. Wejryd E., Marchini G., Frimmel V., Jonsson B. and Abrahamsson T. Probiotics promoted head growth in extremely low birthweight infants in a double-blind placebo-controlled trial. Acta Paediatr. 2019; 108:
62–69 https://doi.org/10.1111/apa.14497 PMID: 29999201
48. Hortensius LM, van Elburg RM, Nijboer CH, Benders MJNL and de Theije CGM. Postnatal Nutrition to Improve Brain Development in the Preterm Infant: A Systematic Review From Bench to Bedside. Front.
Physiol. 2019; 10:961 https://doi.org/10.3389/fphys.2019.00961 PMID: 31404162
49. Hoban AE, Stilling RM, Ryan FJ, Shanahan F, Dinan TG, Claesson MJ et al. Regulation of prefrontal cortex myelination by the microbiota. Transl. Psychiatry. 2016; 6:e774 https://doi.org/10.1038/tp.2016. 42 PMID: 27045844
50. Ruiz L, Moles L, Gueimonde M, Rodriguez JM. Perinatal Microbiomes' Influence on Preterm Birth and Preterms' Health: Influencing Factors and Modulation Strategies. J Pediatr Gastroenterol Nutr. 2016;
63:e193–e203 https://doi.org/10.1097/MPG.0000000000001196 PMID: 27019409
51. DiBartolomeo ME, Claud EC. The Developing Microbiome of the Preterm Infant. Clin Ther. 2016.
38:733–739 https://doi.org/10.1016/j.clinthera.2016.02.003 PMID: 26947798
52. Guaraldi F, Salvatori G. Effect of breast and formula feeding on gut microbiota shaping in newborns. Front Cell Infect Microbiol. 2012; 2:94 https://doi.org/10.3389/fcimb.2012.00094 PMID: 23087909
53. Petherick A. Development: Mother's milk: A rich opportunity. Nature. 2010; 468:S5–S7 https://doi.org/
10.1038/468S5a PMID: 21179083
54. Victora CG, Bahl R, Barros AJ, Franc¸a GV, Horton S, Krasevec J, et al. Lancet Breastfeeding Series Group. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. Lancet. 2016; 387: 475–490 https://doi.org/10.1016/S0140-6736(15)01024-7 PMID: 26869575

Figures

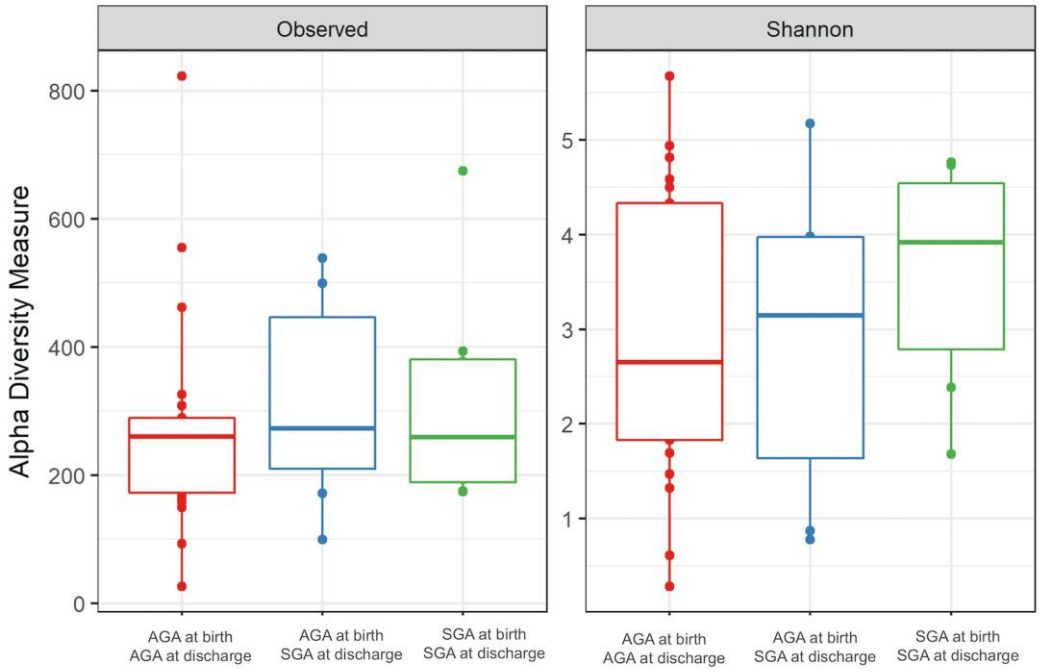


Fig 1. Alpha diversity measurements of meconium microbial communities from preterm infants comparing weight adequacy at birth and at discharge. The left panel presents the number of observed Operational Taxonomic Unities (OTUs) (p-value = 0.745). The right panel presents the Shannon microbial index of diversity (p-value = 0.127). Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles, and single circles indicate outliers. AGA: adequate for gestational age; SGA: small for gestational age. (The group SGA-LGA was excluded from data analysis, because of its limited size).

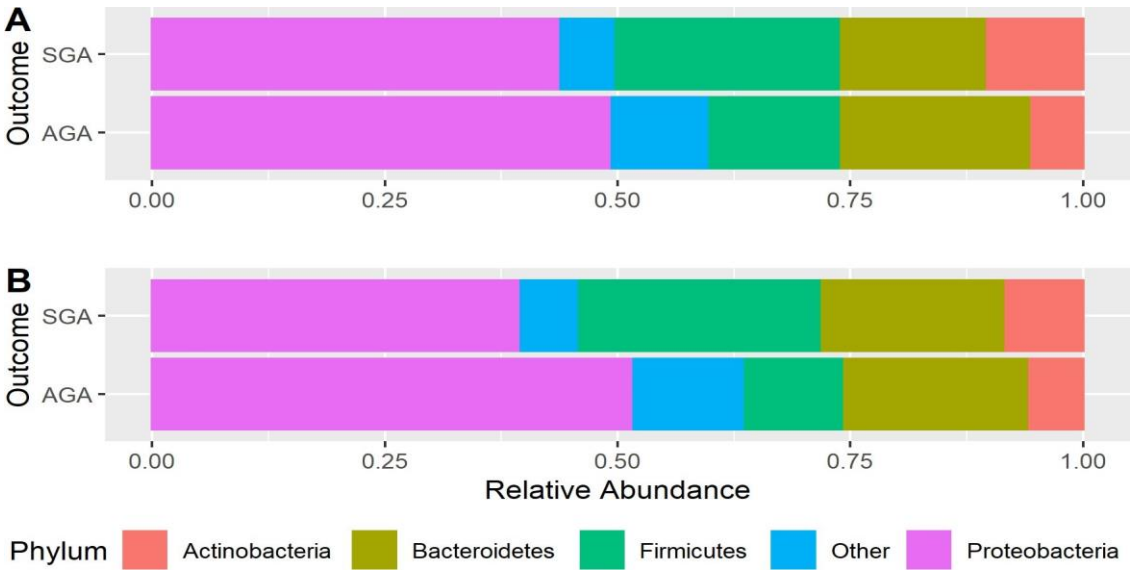


Fig 2. Relative phyla abundance of the gut microbiota according with weigh adequacy. Each stacked bar represents the mean relative abundance of weight adequacy group at birth (A) and at moment of discharge (B).

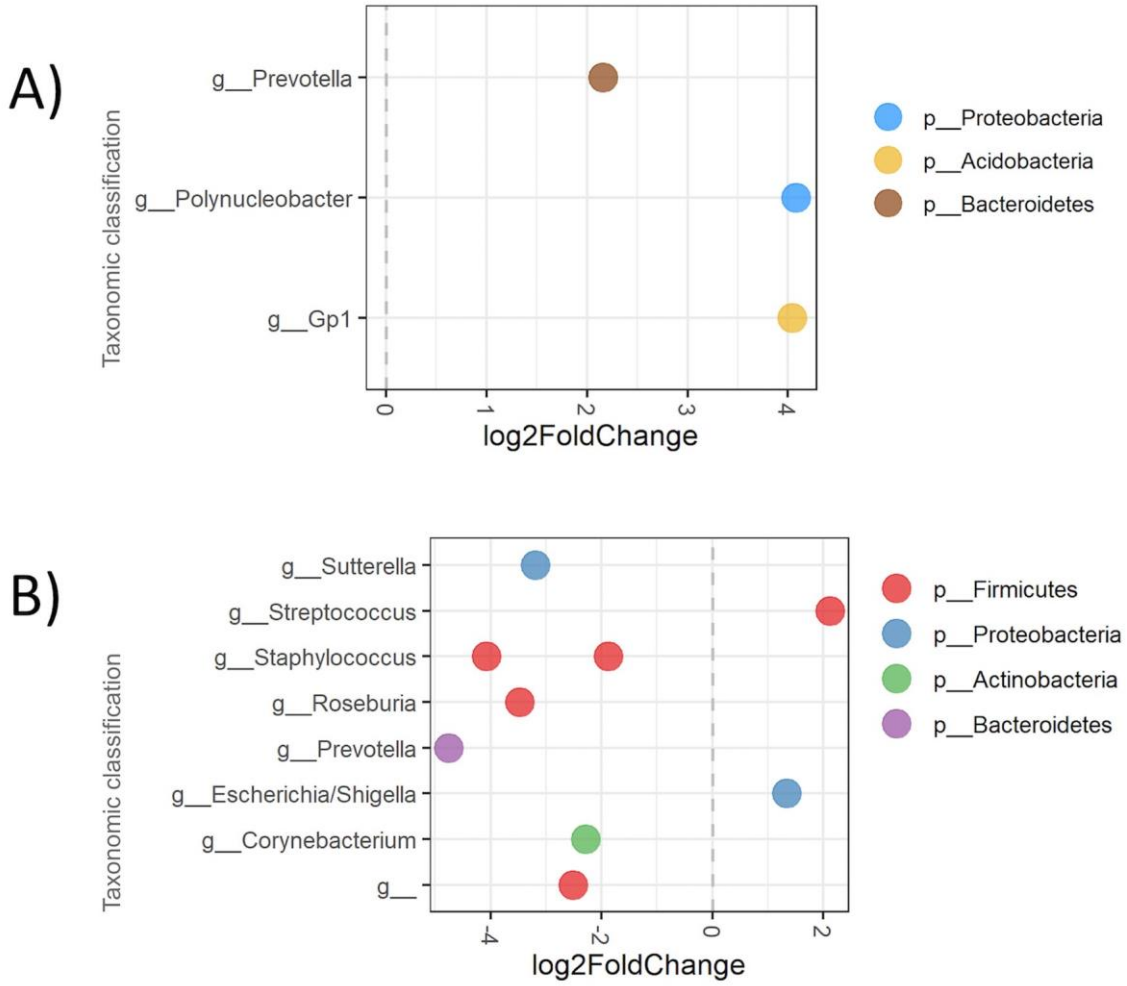


Fig 3. Differential abundance analysis according to weight adequacy. Each dot represents an individual OTU, organized by their Genus. **(A)** Differential abundance analysis according to weight adequacy at birth: *Polynucleobacter (p* = 0.0163), *Gp1 (p = 0.018) and Prevotella (p = 0.038)* were more abundant in meconium of preterm born AGA. **(B)** Differential abundance analysis according to weight adequacy for gestational age at discharge: *Escherichia fergusoni (p = 0.014)* and *Streptococcus dentisani (p = 0.43)* were more abundant in meconium of preterm AGA at discharge; *Prevotella copri (p = 0.002)*, *Roseburia inulinivorans (p = 0.003)*, *Staphylococcus* sp. (p = 0,003), *Staphylococcus capitis* subsp. *Capitis (p = 0.004)*, *Sutterella stercoricanis (p = 0.014)*, *Corynebacterium tuberculostearicum (p = 0.033) and Ruminococcaceae (0.043)* were more abundant in meconium of preterm SGA at discharge.

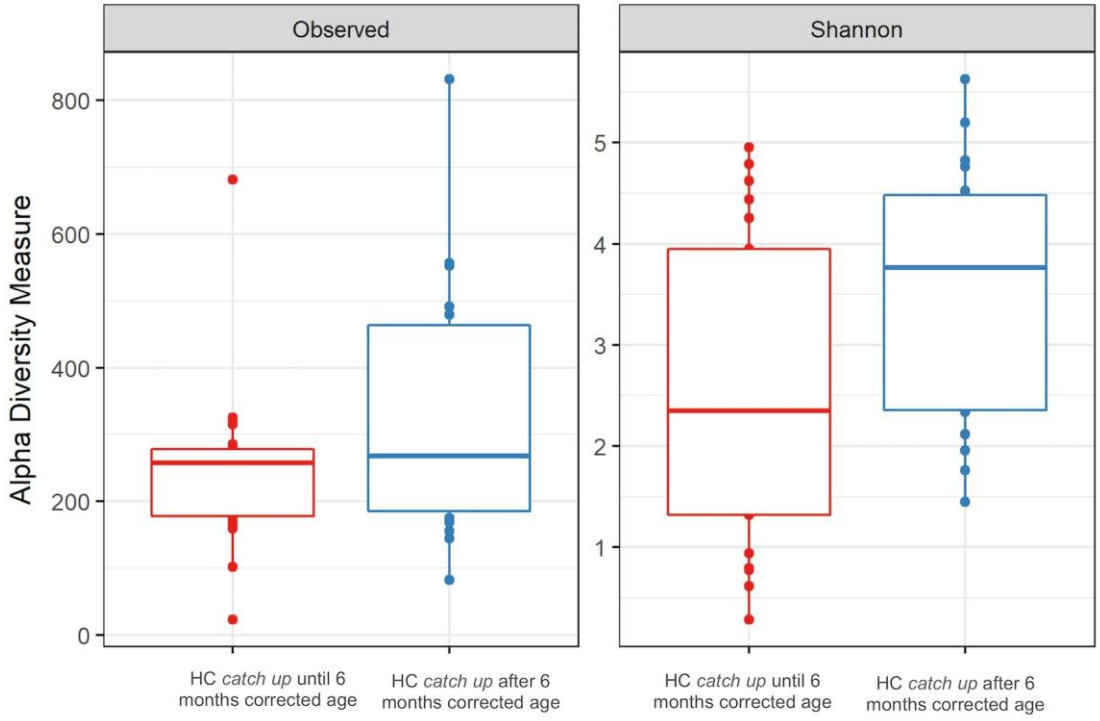


Fig 4. Alpha diversity measurements of meconium microbial communities from preterm infants comparing head circumference (HC) catch up until or after 6 months of corrected age. The left panel presents the number of observed Operational Taxonomic Unities (OTUs) (p-value = 0.225). The right panel presents the Shannon microbial index of diversity (p-value = 0.045). Boxes span the first to third quartiles; the horizontal line inside the boxes represents the median. Whiskers extending vertically from the boxes indicate variability outside the upper and lower quartiles, and single circles indicate outliers.

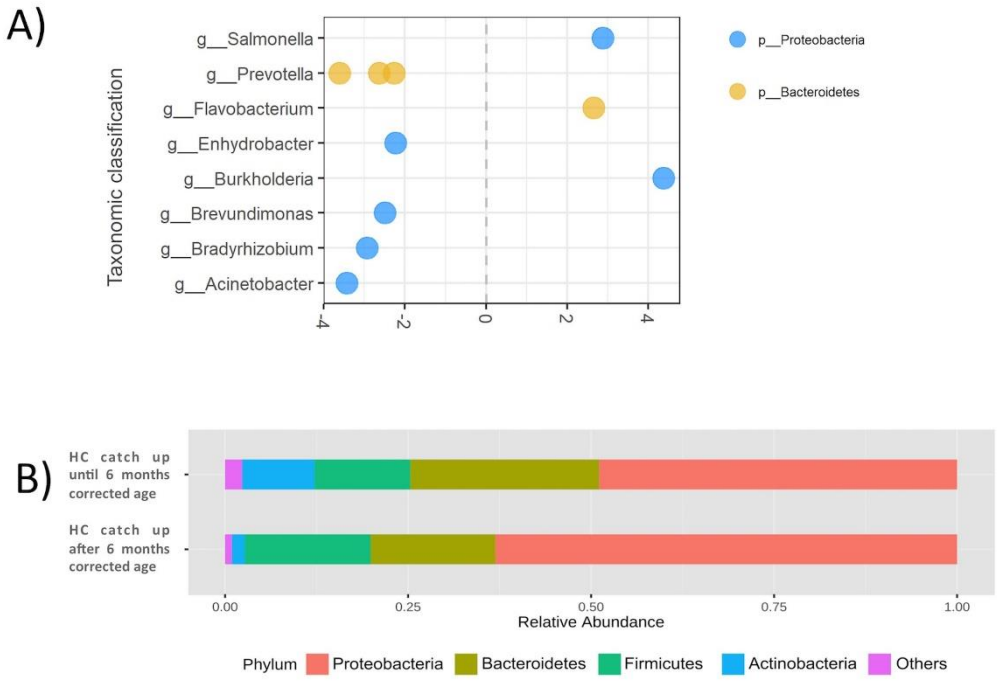


Fig 5. Differential abundance analysis according to head circumference catch up. Each dot represents an individual OTU, organized by their Genus. **(A)** Differential abundance analysis according

to early or late HC catch up. Data plotted as log2 fold change; OTUs to the right of the zero line were more abundant in HC catch up until 6 months corrected age group, and OTUs to the left of the zero line were more abundant in HC catch up after 6 months corrected age group. **(B)** Difference for microbial composition between infants with early HC catch up growth (up to 6 months) and late HC catch up growth (after 6 months); HC: head circumference.

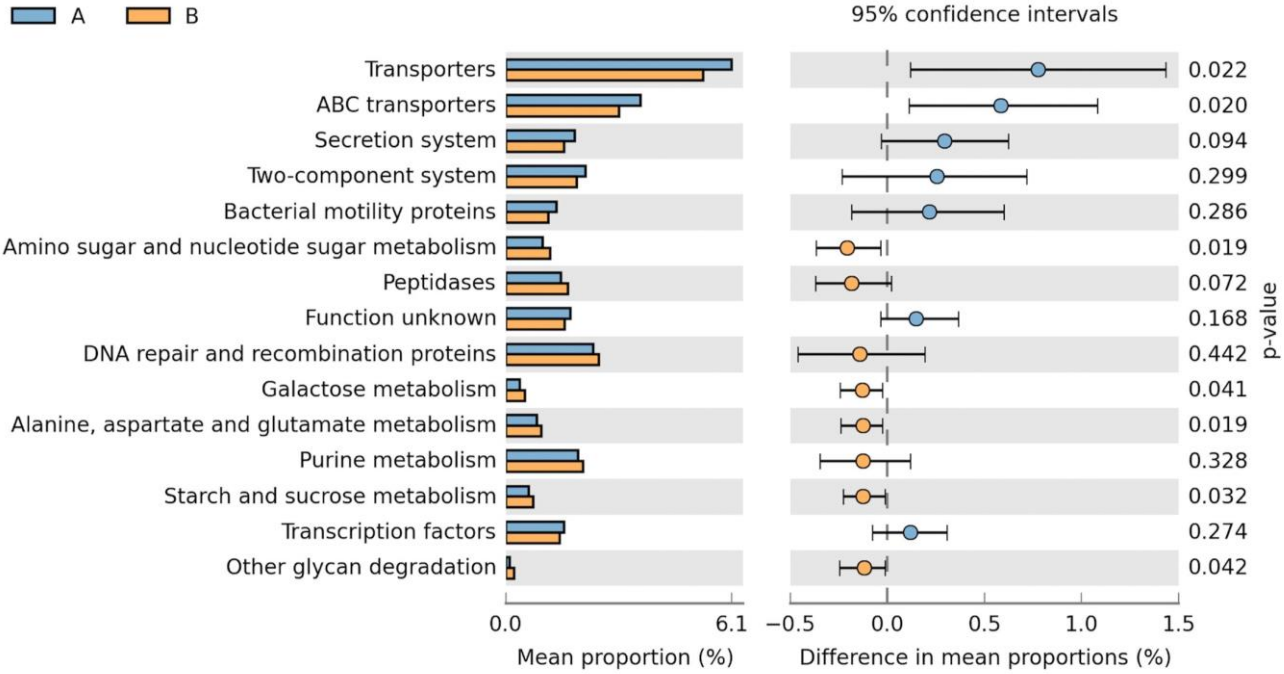


Fig 6. Microbial community functional prediction. Gut microbiota functional predictoin, using PICRUSt, of infants with early or late HC catch up. The bar plot respresents function mean proportion, and error bars represents the difference between the two groups. Coloring of the error bar is according with the group with the higher proportion of the respective function. Blue color (A) represents infants with HC catch up until 6 months, and Orange (B) represents those with HC catch up after 6 months of age.

Tables

Table 1. Clinical characteristics of preterm infants according to adequacy of weight for gestational age at birth and discharge.

Variables	AGAbirth-AGAdischarge (n = 36)	AGAbirth-SGAdischarge (n = 13)	SGAbirth-AGAdischarge (n = 12)	SGAbirth-LGAdischarge (n = 2)	*p* value
Male	16 (44.4%)	7 (53.85%)	6 (50.%)	1 (50%)	0.944
Maternal Age (years)	25.92±6.69	25.62±6.13	26.58±6.62	27.5±10	0.973
C-section	15 (41.7%)	2 (15.4%)	1 (8.3%)	0	0.062
Preeclampsia	4 (11.1%) [a]	4 (30.8%)[a.b]	6 (50%)[b]	2 (100%)[b]	0.003
GDM	5 (13.9%)	1 (7.7%)	1 (8.3%)	0	0.855
UTI	5 (13.9%)	1 (7.75)	1 (8.3%)	0	0.855
GA at birth (weeks)	30.11±2.35	29.85±2.44	29.58±2.74	31.5±0.7	0.744
BW (kg)	1.500±0.507 [a]	1.3800±0.506 [a.b]	1.000±0 [b]	1.000±0 [a.b]	0.010
BW z-score	0.16 (-1.42–2.46) [a]	-0.28 (-1.11–1.51)[a.c]	-1.65 (-2.08–-1.35) [b]	-1.44 (-1.55–-1.34)[b.c]	<0.001

L at birth (cm)	40.18±3.28 [a]	38±3.69	34.5±5.1 [b]	38±1.41	0.001
BL z-score	0.20 (-2.0–1.69) [a]	-0.53 (-1.60–-0.67) [b]	-1.83 (-3.42–-0.12) [c]	-1.40 (-1.45–-1.35) [a.b.c]	<0.001
CP at birth (cm)	27.94±2.54 [a]	27.38±2.3 [a.b]	24.92±2.9 [b]	25.3±0.49 [a.b]	0.008
CP at birth z-score	0.13 (-1.66–2.05) [a]	-0.14 (-1.48–1.35) [a]	-1.67 (-2.40–-0.53) [b]	-1.24 (-1.57–-0.92)[a.b]	<0.001
Length of hospitalization (days)	47.4(14–114)	63.3 (29–122)	72.8 (25–137)	48 (25–71)	0.104
GA discharge (weeks)	36.8±2.24 [a]	38.9±2.95[a.b]	39.9±3.86 [b]	38.3±3.9 [a.b]	0.008
Weight at discharge (kg)	2.63±0.572	2.49±0.335	2.440±0.489	2.777±0.682	0.625
Type of milk at discharge					
EBM	5 (13.9%)	2 (15.4%)	1 (8.3%)	1 (50%)	0.176
BM+formula	19 (52.8%)	6 (46.2%)	10 (83.3%)	1(50%)	
Formula	12 (33.3%)	5 (38.5%)	1(8,3%)	0	

*Mean ± SD;
**Absolut frequency (%);
**Mean (Min-Max); AGA: Appropriate-for-Gestational-Age; SGA: Small-for-gestational-age BW: Birth weight; L: Length; CP: Head circumference; GA: Gestational Age; GDM: gestational diabetes mellitus; UTI: Urinary Tract Infection; EBM: Exclusive Breast Milk; BM; Breast Milk

Table 2. Clinical characteristics, growth and type of milk received according to catch-up before or after 6 months of corrected age.

Variables	***Catch up* <6m (n = 30)**	***Catch up* >6m (n = 33)**	**P value**
Male	16 (53.3%)	14 (42.4%)	0.454
Maternal age(years)	25.33±6.26	27±6.77	0.299
C-section	12 (40%)	18 (60%)	0.093
Preeclampsia	4 (13.3%)	12 (36.4%)	0.046
Gestational Diabetes	4 (13.3%)	3 (9.1%)	0.700
Urinary tract infection	4 (13.3%)	3 (9.1%)	0.700
Maternal antibiotics	20(66.7%)	21(63.6%)	1.000
GA at birth (weeks)	30.4±2.29	29.6±2.4	0.209
AGA at birth	22(73.3%)	26(78.8%)	0.612
Weight at birth (kg)	1.434 ±0.443	1.323±0.479	0.345
Z-score Weight at birth	-0.33 (-2.08–1.25)	-0.31 (-1.87–2.46)	0.933
Length at birth(cm)	38.7±3.84	38.2±4.32	0.654
z-score Length at birth	-0.34 (-3.4–1.5)	-0.42 (-3.04–1.69)	0.788
Head circunference at birth (cm)	27.52 ±2.66	26.55±2.92	0.247

Z-score Head circunference at birth	-0.27 (-2.36–1.79)	-0.34 (-2.4–2)	0.785
NICU stay (days)	49 (14–114)	61 (29–122)	0.137
Periventricular leukomalacia	2 (6.7%)	3 (9.1.%)	0.546
Necrotizing enterocolitis	4 (13.3%)	6(18.2%)	0.430
Early sepsis	0	1 (3%)	0.625
Late sepsis	2 (6.6%)	3 (9%)	0.423
Gestational age at discharge (weeks)	37.4±2.3	38.4±3.4	0.174
Weight at discharge (kg)	2.63±0.572	2.49±0.335	0.625
Weight z-score at discharge	-0.94 (-3.2–1.38)	-1.35 (-3.33–0.27)	0.104
AGA at discharge	18(60%)	19 (57.6%)	0.845
Hospitalization after discharge	4 (13.3%)	10 (30.3%)	0.106
Use of anticonvulsant	5 (16.7%)	10 (30.3%)	0.204
Weight at 2 months CA (kg)	5.450±0.970	4.98±0.810	0.055
Weight Z-score at 2 months CA	0 (-3.82–2.30)	-0.55 (-2.64–2.12)	0.134
Head circumference at 2 months CA (cm)	39.44±1.78	38.43±1.70	0.084
Head circumference Z-score at 2 months CA	0.75 (-2.69–2.87)	0 (-3.51–2.44)	0.040
Weight at 4 months CA (kg)	7.130±1.00	6.240v1.13	0.008
Weight Z-score at 4 months CA	0.37 (-1.81–2.66)	-0.68 (-4.31–2.44)	0.012
Head circumference at 4 months CA (cm)	42.57±1.14	40±2.0	<0.001
Head circumference Z-score at 4 months CA	1.13(-0.54–3.23)	-0.27 (-3.63–2.81)	0.001
Weight at 6 months CA (kg)	7.80±1.21	7.0±1.15	0.021
Variables	***Catch up* <6m (n = 30)**	***Catch up* >6m (n = 33)**	**P value**
Weight Z-score at 6 months CA	0.05(-4.75–2.55)	-0.72(-4.38–2)	0.050
Head circumference at 6 months CA (cm)	44.1±1.25	41.71±1.96	<0.001
Head circumference Z-score at 6 months CA	0.94 (-1.92–2.75)	-0.39 (-3.55–2.14)	<0.001
Type of milk			
Milk at discharge			
EBM	4 (13.3%)	5 (15.2%)	0.090
BM+Formula	18 (60%)	18 (54.4%)	
Formula	8 (26.7%)	10 (30.3%)	
Milk at 2 months CA			
EBM	4 (14.3%)	6 (17.9%)	0.0752
BM+Formula	9 (28.6%)	11 (33.3%)	
Formula	15(53.6%)	16 (48.4%)	
Cow’s milk	1 (3%)	0	
Milk at 4 months CA			
EBM	2 (6.6%)	5(14.8%)	0.404
BM+Formula	8 (27.3%)	11(33.3%)	
Formula	19(63.6%)	17(51.9%)	

Cow's milk	1 (3%)	0	
Milk at 6 months CA			
EBM	2 (6.9)[a.b]	3 (7.4%)[a.b]	0.038
BM+Formula	3 (10.3%)	12 (37%)	
Formula	21 (69.9%)	18(55.6%)[b]	
Cow's milk	4 (13.8%)	0b	

*Mean ± SD;
**Absolut frequency (%);
***Mean (Min-Max); CA: corrected age; AGA: Appropriate-for-Gestational-Age; SGA: Small-for-gestational-age BW: Birth weight; L: Length; CP: Head circumference; GA: Gestational Age; GDM: gestational diabetes mellitus; UTI: Urinary Tract Infection; EBM: Exclusive Breast Milk; BM; Breast Milk

Table 3. Nonparametric Multivariate Analysis of Variance of bacterial community structure used for controlling confounding variables.

Variables	**F Model**	**R2**	**p-value**
Weight Adequacy	0.961	0.101	0.536
HC Catch-up	1.255	0.033	0.201
Preeclampsia	0.836	0.022	0.640

ANEXO A – Parecer Comitê de Ética

FUNDAÇÃO UNIVERSIDADE FEDERAL DO PAMPA - UNIPAMPA

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Associação entre microbiota fetal, prematuridade e morbidades do recém-nascido pré-termo

Pesquisador: Renato Soibelmann Procianoy
Área Temática:
Versão: 2
CAAE: 26436013.0.3001.5323
Instituição Proponente: Hospital de Clínicas de Porto Alegre
Patrocinador Principal: MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 605.489-0
Data da Relatoria: 12/02/2014

Apresentação do Projeto:
De acordo com o pesquisador:
Projeto acadêmico que visa identificar associações entre o padrão da microbiota intestinal e vaginal com o risco de parto prematuro e com a morbimortalidade em recém-nascidos pré-termos. A hipótese conceitual é a de que possa existir um mecanismo causal potencial de parto prematuro com base na colonização fetal com microbiota vaginal ou intestinal, ao invés de infecção com patógenos estabelecidos. Os autores especulam que, a partir do conhecimento gerado com a presente pesquisa, possam ser desenvolvidas novas técnicas e intervenções baseadas na detecção de microbiota com o objetivo de diminuir a ocorrência da prematuridade e da mortalidade e morbidade recém-nascidos prematuros no Brasil.
A base da pesquisa consiste em identificação dos potenciais patógenos utilizando técnicas de

Endereço: Campus Uruguaiana BR 472, Km592
Bairro: Prédio Administrativo Central - Caixa **CEP:** 97.500-970
UF: RS **Município:** URUGUAIANA
Telefone: (55)3413-4321 **E-mail:** cep@unipampa.edu.br

biologia
molecular com extração de DNA e sequenciamento genético. O projeto não apresenta um maior detalhamento dos aspectos clínicos. Na etapa 2 por exemplo, não foi explicado como valorizar os achados da microbiota dos neonatos em uso de antibioticoterapia (muito frequente nesta população).
A pesquisa está dividida em três etapas:
Etapa 1- consistirá da coleta em pelo menos 50 mães e 50 crianças saudáveis e a termo de uma amostra fecal e de um swab vaginal obtidos a partir de mães no momento da admissão no hospital para o parto e do primeiro mecônio. A previsão é de avaliar 150 amostras nesta etapa.
Etapa 2- 200 mães e recém-nascidos prematuros com idade gestacional abaixo de 32 semanas serão recrutados durante um período de 24 meses. Swab vaginal e amostras de fezes serão coletadas de cada mãe e amostras do primeiro meconio dos recém-nascidos.
Etapa 3- consistirá de coletas semanais de amostras obtidas do mecônio enquanto o paciente estiver internado visando acompanhar as mudanças no microbioma (diversidade microbiana, abundância e estrutura).

Objetivo da Pesquisa:
Objetivo Primário:
Identificar associações entre o padrão da microbiota intestinal e vaginal com risco de parto prematuro e com a morbimortalidade em recém-nascidos pré-termos para que possam ser desenvolvidas novas técnicas e intervenções baseadas na detecção de microbiota com o objetivo de diminuir a ocorrência da prematuridade e da mortalidade e morbidade recém-nascidos prematuros no Brasil.
Objetivos Secundários:
Objetivo 1: Construir um atlas microbiano da microbiota vaginal e intestinal das mães no momento da admissão para o parto, e microbiota do primeiro mecônio de recém-nascidos saudáveis a termo.
Objetivo 2: Determinar a relação entre partos prematuros: microbiomas de mães e recém-nascidos.

Objetivo 3: Acompanhar as mudanças no microbioma (diversidade microbiana, abundância e estrutura) a partir do recém-nascido com as primeiras fezes (mecônio) e continuando até a alta hospitalar.
Objetivo 4: Estabelecer um repositório da microbiota que irá facilitar mais estudos

Avaliação dos Riscos e Benefícios:
De acordo com o pesquisador:
Riscos: não são conhecidos riscos imediatos e/ou diretos pela participação no estudo.
Benefícios: A participação no estudo não trará benefício direto ao participante, mas contribuirá para melhor entendimento sobre as causas biológicas de nascimentos de recém-nascidos prematuros.

Comentários e Considerações sobre a Pesquisa:
Projeto de relevância clínica

Considerações sobre os Termos de apresentação obrigatória:
Não foram avaliados os termos- Projeto já aprovado pelo CEP HCPS-UFRGS.

Recomendações:
Sem recomendações

Conclusões ou Pendências e Lista de Inadequações:
O CEP Unipampa recebeu este projeto para ciência por encaminhamento do CEP HCPA- UFRGS para esta instituição como co - partícipe, que possui comitê de ética. Tendo em vista que o presente projeto já foi aprovado pelo CEP do HCPA- UFRGS não foi analisado o mérito do mesmo, no entanto registramos as seguintes observações:
- ausência do termo de anuência da instituição co- partícipe (Unipampa).
- sugere-se realizar os registros pertinentes nos trâmites no âmbito interno da Universidade.

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

Considerações Finais a critério do CEP:

FUNDAÇÃO UNIVERSIDADE FEDERAL DO PAMPA - UNIPAMPA

Continuação do Parecer: 605.489-0

URUGUAIANA, 12 de Abril de 2014

Assinador por:
GIULIA ALESSANDRA WIGGERS PEÇANHA
(Coordenador)

Este parecer reemitido substitui o parecer número 605489 gerado na data 12/02/2014 18:20:57, onde o número CAAE foi alterado de 26436013.0.0000.5327 para 26436013.0.3001.5323.

Endereço: Campus Uruguaiana BR 472, Km592
Bairro: Prédio Administrativo Central - Caixa **CEP:** 97.500-970
UF: RS **Município:** URUGUAIANA
Telefone: (55)3413-4321 **E-mail:** cep@unipampa.edu.br