

## Invited Review

Int J Oral Biol 45:77-83, 2020  
 pISSN: 1226-7155 • eISSN: 2287-6618  
<https://doi.org/10.11620/IJOB.2020.45.3.77>

# Practical considerations for the study of the oral microbiome

Yeuni Yu<sup>1</sup>, Seo-young Lee<sup>1</sup>, and Hee Sam Na<sup>2\*</sup>

<sup>1</sup>Interdisciplinary Program of Genomic Science, Pusan National University, Yangsan 50612, Republic of Korea

<sup>2</sup>Department of Oral Microbiology, School of Dentistry, Pusan National University, Yangsan 50612, Republic of Korea

In the oral cavity, complex microbial community is shaped by various host and environmental factors. Extensive literature describing the oral microbiome in the context of oral health and disease is available. Advances in DNA sequencing technologies and data analysis have drastically improved the analysis of the oral microbiome. For microbiome study, bacterial 16S ribosomal RNA gene amplification and sequencing is often employed owing to the cost-effective and fast nature of the method. In this review, practical considerations for performing a microbiome study, including experimental design, molecular analysis technology, and general data analysis, will be discussed.

**Keywords:** Oral microbiome, 16S rDNA, Bacterial associations, High-throughput nucleotide sequencing

## Introduction

구강 내에는 700여 종 이상의 세균이 존재하는 것으로 보고되어 있고 이들은 다양한 요인들에 의해 영향을 받으며 복잡하게 구성되어 있다[1]. 이 가운데 400에서 500여 종의 세균이 치은연하에서 보고되었고 혀, 치아 표면, 구강 점막, 연구개, 경구개 등 다양한 환경에 따라 각기 독특한 미생물 군집을 형성하고 있다[2,3]. 많은 미생물의 경우 구강 내 다양한 위치에서 동시에 발견되기도 하지만, *Streptococcus*, *Gemella*, *Granulicatella*, *Neisseria*, *Prevotella*와 같은 일부 미생물의 경우 특정한 곳을 보다 선호하기도 한다. 예를 들어 *Streptococcus salivarius*의 경우 혀에 주로 군락을 형성하는 반면 *Treponema*는 주로 치하연하 틈새를 선호한다[3]. 구강 내에 군락을 이루는 미생물 중 일부는 충치, 치주질환, 치수병변, 구취, 치원성 감염 등의 구강 질환과 깊은 관련성을 갖는다[4-8]. 뿐만 아니라 구강 미생물 가운데 일부는 당뇨[9], 크론병[10], 심장질환[11], 조기분만[12]과 같은 전신 질환도 깊은 관련성을 갖는다.

DNA 염기서열 분석법의 발달은 구강 미생물 군집 분석에 획기적인 변화를 가져왔다. 미생물은 16S rRNA의 염기서열을 분석함으로써 분

류할 수 있다[13]. 차세대염기서열 분석(next generation sequencing, NGS)의 도입은 수많은 염기서열을 동시에 분석함과 동시에 분석 비용을 획기적으로 낮춤으로써 질병[14], 항생제[15], 유산균[16], 식습관[17]의 영향 등 다양한 분야에서 미생물 군집의 분석을 가능하게 하였다. 그러나, 연구자들은 NGS 분석을 통해 얻어지는 엄청난 양의 데이터와 함께 데이터 분석에 사용되는 다양한 분석 도구와 기법에 노출되어 당황하게 된다. 다른 연구와 마찬가지로 미생물 군집 연구 또한 연구 방법, 환경 인자 및 분석 방법에 따라 결과에 영향을 끼칠 수 있다. 미생물 군집 연구에서 데이터 수집과 분석의 표준화는 아직 도입단계지만 현재 시행되는 분석으로도 상당히 유의한 결과를 얻을 수 있다. 비록 미생물 군집 분석에 대해 모두 고찰하는 것은 어렵지만, 여기에서는 미생물 군집 분석을 위한 실험과 결과 분석에 대해 소개하고자 한다.

## Main Body

### 1. 16S rRNA 유전자 분석을 통한 미생물 군집 분석

16S rRNA 분석을 통한 미생물 분류는 기존에 분리 배양에 의한 분류

Received June 1, 2020; Revised August 14, 2020; Accepted August 20, 2020

\*Correspondence to: Hee Sam Na, E-mail: heesamy@pusan.ac.kr  <https://orcid.org/0000-0002-3246-4681>

Copyright © The Korean Academy of Oral Biology

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

법에 비하여 다양한 환경에서 미생물 군집을 분석하는 데 획기적인 변화를 가져왔다[18]. 구강 미생물 연구에서도 16S rRNA 염기서열 비교는 기본적인 연구 방법으로 사용되고 있다. 순수 배양된 세균의 경우 16S rRNA 유전자(대략 1,500 bp)를 증폭하여 Sanger sequencing 방법으로 염기서열을 분석한다[19]. 치태와 같이 다양한 세균에서 분리된 경우 16S rRNA 유전자를 증폭한 후 증폭된 유전자를 대장균에 복제하였다. 그리고 복제한 16S rRNA 염기서열을 분석하여 미생물 군집을 분석하였다[20]. 일반적으로 98.5% 상동성을 보이는 경우 동일한 종으로 간주된다.

다양한 NGS의 도입은 수많은 16S rRNA 염기서열을 동시에 분석하는 것을 가능하게 하였다. 그러나, NGS는 전체 16S rRNA를 분석하는 대신 미생물 분류군 간에 변이가 큰 구간을 분석하여 미생물 분류군을 분류하고 있다. 16S rRNA에는 크게 9개의 초가변 구역(hypervariable region)이 존재하는데 미생물 군 분류에는 V3-V4 구간이 가장 효과가 좋은 것으로 보고되었다[21]. 그러나, V1-V3도 세균의 다양성을 연구하거나[22], *Bifidobacteria*의 분포가 높은 시료의 경우에도 유용하다는 보고가 있다[23].

## 2. 실험 계획

실험 계획은 의미 있는 데이터를 생성하는 분석의 중요한 첫 단계이다. 미생물 군집 연구는 동물실험, 단면 조사 연구, 환자 대조군 연구, 코호트 연구 등 다양한 모델에서 적용될 수 있다. 미생물 군집 분석에 대한 일반적인 접근 방식은 시료 출처와 상관없이 적용 가능하다. 그러나, 분석의 특정 세부 사항은 시료의 종류에 따라 다를 수 있다. 예를 들어, 16S rRNA의 증폭 구간 선정은 시료 채취 부위에 따라 다양한 결과가 보고되는 경우가 있다[24].

다른 고려 사항으로는 실험 계획과 시료 수집이다. 미생물 군집 분석에서 정확하고 의미 있는 결과를 얻기 위해 세심한 실험 설계는 매우 중요하다. 많은 요소들이 통제되지 않으면 미생물 군집 데이터의 패턴을 모호하게 할 수 있다. 메타 데이터의 세심한 자료 정리, 적절한 대조군과 관심 변수에 영향을 끼칠 수 있는 다양한 인자들에 대한 신중한 계획이 가장 중요하다.

먼저 실험 범위를 정의하고 관심 있는 문제에 적합한 실험을 설계 선택해야 한다. 예를 들어, 단면 연구는 건강한 사람과 질병을 가진 사람 또는 다른 지리적 지역에 사는 사람과 같은 다른 집단 사이의 미생물 군집의 차이를 찾는 데 유용하다. 그러나 개인 간의 미생물 군집의 큰 차이와 생활 양식[25], 식습관[26], 약물[27] 등의 영향으로 인해 관심 있는 질병 이외의 요인으로 인해 집단 간의 차이가 발생할 수 있다. 질병 발병 전에 기준선 샘플을 수집하는 전향적 코호트 연구는 연구 비용이 비록 많이 들지만 이러한 문제를 해결하는 데 도움이 될 수 있다.

분석 계획과 조사하고자 하는 주제로 연구를 설계하면 표본 크기를 결정하는 데 도움이 될 수 있다. 기술적 변동성과 실제 생물학적 결과를 식별하기 위해 통계적 검증력을 평가하는 적절한 방법을 사용해야 한다. 그러나, 통계적 검증력과 효과크기 분석은 미생물 군집 연구에서 아직도 도전 분야로 남아 있다. 현재 통계적 검증력과 효과크기 분석에 사

용되는 방법으로는 PERMANOVA [28], Dirichlet Multinomial [29] 또는 Random forest 분석[30] 기반 방법들이 사용되고 있다. 특정 실험 설계 고려 사항에 대해서는 유사한 샘플 유형과 원하는 결과로 다른 성공적인 연구의 설계를 검토하는 것이 가장 좋은 것으로 여겨진다.

대조군과 제외 기준을 명확히 하는 것이 좋다. 연구에 포함되거나 제외되는 명확한 기준을 정의하면 데이터 해석에 혼동을 줄 수 있는 변수들을 제한시킬 수 있다. 예를 들어, 개인 간 항생제 복용 후 미생물 군집 회복 시간의 변화는 지난 6개월 동안 항생제로 치료받은 개인은 대부분의 미생물 군집 연구에서 제외되어야 함을 시사한다[31]. 대조군 실험 연구의 경우 대조군은 적절히 선택하고 실험군과 일치시켜야 한다. 나이, 성별은 보편적인 기준이고 약물이나 식습관 등은 결과에 중요한 영향을 끼칠 수 있는 변수로 작용할 수 있다.

## 3. 샘플링 방법, 보관, DNA 추출

DNA 추출과 염기서열 분석 방법에 따라 기술적인 오차가 생기게 된다[32]. 따라서 동일한 연구 내에서는 동일한 시약을 사용하여 분석하는 것이 필수적이고 추적 연구의 경우 시점 간의 고유 변동성을 평가하기 위해 기준선 샘플을 여러 번 수집하는 것이 좋다. 시료 및 시약의 오염 여부를 확인하기 위하여 빈 시료를 시료 채취, DNA 추출, 중합효소 연쇄반응(polymerase chain reaction, PCR), 염기서열 분석 시에 추가하는 것이 좋다. 시료는 채취 즉시 가능하면  $-80^{\circ}\text{C}$ 에 보관하는 것이 좋다. 그러나, 야외 연구로 인해 얼리는 것이 불가능할 경우 95% 에탄올이나 RNAlater와 같은 시약을 사용하여 보관하는 것이 좋다[33]. 구성이 이미 알려진 인공 미생물 군집을 분석에 추가하는 경우 실험 간의 결과를 표준화하여 도움을 줄 수 있다[34].

## 4. 차세대 유전체 분석 플랫폼

미생물 군집 분석에는 16S rRNA 분석법을 이용한 연구가 활발히 진행되고 있는데, 신속하고 경제적인 면에서 장점을 갖는다. 이 분석법은 조직과 같이 숙주의 DNA에 오염된 시료나 소량의 샘플의 경우에도 사용할 수 있다. 그러나, PCR primer가 증폭하는 구역의 유전자 염기서열이 동일하지 않을 수 있기 때문에 표적 염기서열에 primer가 동일하게 결합하지 않을 수 있어 PCR 증폭 과정에서 편향된 결과가 유도될 수 있다. 다른 내재적 편향의 요인으로는 가변 구역의 선정, 증폭된 유전자 크기, PCR 사이클 등이 있다. 소량의 샘플은 특히 과증폭으로 인한 편향된 데이터에 취약한데 이는 PCR 사이클이 증가함에 따라 오염된 미생물이 과도하게 증폭되어 표현될 수 있기 때문이다. 이러한 편향은 primer를 적절하게 선택함으로써 최소화할 수 있으나 이를 위해서는 표적 미생물 군의 구성에 대한 사전 지식이 필요하다. 최근 널리 보급되어 사용되는 NGS는 대량의 염기서열을 분석함으로써 미생물 군집 연구에 획기적인 변화를 가져왔다. 다음에 소개할 방법들은 가장 많이 이용되고 있는 NGS 플랫폼이다.

### 1) 454 파이로시퀀싱

이 방법은 유화액 내 비드에 DNA 조각을 증폭하면서 파이로시퀀싱을 통하여 염기서열을 분석하였다[35]. 한 번 운영하였을 때 최대 700 bp 길이로 400,000개의 염기서열을 분석할 수 있다. 초기 미생물 군집 연구에 가장 많이 사용되었으나 지금은 다른 플랫폼에 비하여 한 번에 생산할 수 있는 데이터 양이 제한적이기 때문에 많이 사용하고 있지는 않다.

### 2) 일루미나

DNA 조각을 칩에 부착하고 형광으로 표지된 dNTP를 이용해 염기서열을 합성하면서 동시에 서열을 분석한다[36]. 분석에 사용되는 기종에 따라 생산되는 데이터 양에 차이는 있지만 한 번 운영하였을 때 최소한 수천만 개의 염기서열을 분석할 수 있다. 일루미나 제품에는 MiSeq, HiSeq, NextSeq, NovaSeq, iSeq 등 다양하게 있으며 최근 가장 많이 사용되고 있다. 특히, MiSeq은 2×300 bp를 읽을 수 있어서 16S rRNA 분석에 많이 사용되고 있다[37].

### 3) Pacific Biosciences (PacBio)

Single molecule real-time (SMRT) technology라는 기술을 적용한 제품으로 이 기술의 가장 큰 특징은 PCR을 통한 DNA 증폭을 생략하였다는 것이다. DNA를 한 분자 상태에서 그대로 시퀀싱하여 염기서열을 분석함으로써 10,000 bp 이상을 읽을 수 있는 장점을 갖는다[38]. PacBio 플랫폼은 참고 유전체 자료 없이 전체 유전체를 분석할 때 많이 이용된다.

### 4) Ion Torrent

Ion Torrent는 반도체 소자 기반 제품으로 DNA 합성 중 발생하는 수소이온에 의한 Ph 변화를 전기 신호로 변환하여 염기서열을 분석한다[39]. 장점으로는 낮은 분석비용과 빠른 분석 속도에 있다. 그러나, 반복되는 염기서열의 길이가 길어지는 경우 정확도가 떨어지는 단점을 갖는다. 최근 400 bp까지 분석할 수 있는 제품을 개발하여 16S rRNA 분석에도 사용하려는 시도가 있다.

## 5. 데이터 분석 방법

모든 NGS 분석에는 광범위한 생물정보학적 기능과 관련된 데이터 품질 관리, 양질의 판독을 위한 필터링, 양호한 참조 게놈에 대한 정렬 및 매핑, 키메라 제거, 의미 있는 해석을 위한 표본 및 모집단 간의 정규화 등이 필요하다.

미생물 군집 분석에 가장 많이 사용되는 16S rRNA 분석도 정확한 분류를 위해 적절한 프로그램과 분석 기법을 사용해야 한다. 데이터 분석의 첫 번째 단계는 시퀀싱 오류를 제거하는 것이다. 시퀀싱 오류율이 매우 낮음에도 불구하고(예를 들어, Illumina 시퀀싱에서 그 오류율은 뉴클레오티드당 ~0.1%), 시퀀스 다양성의 대부분은 시퀀싱 오류에서 발생한다[40]. 최근까지 유사한 시퀀스를 operational taxonomic unit (OUT)로 클러스터링하여 이 문제를 해결하였다. OTU 선택이라고 불리

는 이 과정은 시퀀스들을 OUT 단위로 클러스터링하여 오류에 의해 발생하는 시퀀스 변형을 포함하여 유사한 시퀀스들을 임계값(보통 유사도 97%) 기준으로 묶음으로써 단일 특징으로 통합할 수 있다[41].

그러나 이 방법은 단일 OTU로 통합되는 과정에서 단일염기다형성(single nucleotide polymorphism, SNP)과 같이 미묘하고 실제적인 생물학적 시퀀스 변동을 놓칠 수 있는 단점이 있다[41]. Oligotyping은 16S rRNA 시퀀싱의 위치별 정보를 포함시켜 미묘한 뉴클레오티드 변동을 식별하고 밀접하지만 구별되는 미생물 종을 구별함으로써 전통적인 OTU 선택을 개선하였다[42]. Deblur 및 DADA2와 같은 알고리즘은 오류 프로파일을 사용하여 시퀀스 데이터를 정확한 시퀀스 특징군으로 분류한다[43,44].

가장 중요한 분석 단계 중 하나는 데이터의 미생물 시퀀스에 분류학적 이름을 할당하는 것이다. 분류법은 일반적으로 naïve Bayesian classifier인 RDP classifier와 같은 기계 학습 접근법에 의해 할당된다[45]. QIIME, Mothur와 같은 인기 있는 미생물 군집 분석 패키지는 분류학 분류에 대한 지원을 제공한다[46]. 원칙적으로 참조 데이터베이스(가장 특징적이고 자주 사용되는 데이터베이스 중 세 가지는 GreenGenes, RDP, Silva)와 정확히 일치해야 보다 정확하게 분류할 수 있다.

분류된 데이터는 각 샘플별 미생물의 리딩 수를 보여주는 커다란 테이블을 생산한다. 이 결과는 매우 간단하다. 미생물 군집 데이터는 종종 수천 가지의 미생물 종을 포함하는 다차원 결과로 의미 있는 결과를 도출하기 위해 세심한 통계적 처리가 필요하다. 미생물 군집의 전반적인 패턴은 일반적으로 알파와 베타 다양성으로 평가된다.

알파 다양성은 개별 표본 내의 형상 다양성을 수량화하고 표본 그룹 간에 비교할 수 있다. 예를 들어, 질병이 있는 표본을 건강한 대조군과 비교할 때 알파 다양성을 사용하여 두 표본 사이의 평균 종의 다양성을 비교할 수 있다. 종의 다양성 측정(예를 들어, Chao1)은 표본당 샘플의 수에 민감하지만 다양성과 균등성을 평가하는 경우(Shannon index)는 비교적 영향을 덜 받는다.

베타 다양성은 각 샘플 쌍 간의 특성 차이를 비교하여 모든 샘플 쌍 간의 거리 매트릭스를 생성한다. 매트릭스 선택은 획득한 결과에 영향을 미칠 수 있으며 생물학적 데이터 해석을 염두에 두고 선택해야 한다[47]. 정량적 매트릭스(Bray-Curtis, Canberra, weighted UniFrac)는 미생물의 비율 데이터를 계산에 사용하는 반면, 정성적 매트릭스(Jaccard, unweighted UniFrac)는 미생물의 유무만을 고려한다. 알파 및 베타 다양성 계산을 수행하기 위한 소프트웨어에는 QIIME, Mabr 및 R 등이 있다. 그룹 간 유의한 베타 다양성 클러스터링을 평가하기 위해 PERMANOVA와 ANOSIM이 있지만, PERMANOVA가 그룹 내에서 다양한 분산을 갖는 데이터의 경우 더 잘 수행될 수 있다[48]. 베타 다양성 데이터를 시각화하기 위해 주좌표 분석(principal coordinates analysis, PCoA) 또는 주성분 분석(principal component analysis, PCA)과 같은 순서 지정 기법이 일반적으로 사용된다. 이러한 방법은 크고 복잡한 거리 행렬을 축소하여 시각적으로 관리할 수 있는 2차원 또는 3차원 표본 거리 표현으로 만든다.

또 다른 일반적인 분석 접근법은 관심 있는 비교 그룹(즉, 치료 대 대조군)에서 미생물의 비율을 분석하는 것이다. 미생물 군집 데이터는 고

차원적(즉, 수천 개의 종을 포함)이고 희소성을 갖는 데이터가 많기 때문에 비교하고자 하는 그룹 간의 차이를 설명할 수 있는 미생물을 찾는 것은 특히 어려운 일이다. 기계 학습은 미생물 군집 데이터를 사용하여 현재 상태를 기준으로 표본을 분리하거나 미래 상태를 예측하는 방법을 결정하는 데 특히 유용한 기법으로 떠오르고 있다[49]. 예를 들어, 개인의 구강 미생물에 근거하여 치주염의 심각성과 민감성을 모델링하는 것이 가능하다[50]. 중요한 것은 기계 학습 분석은 상당한 표본 크기를 필요로 하며 항상 교차 검증, 독립 시험 세트 또는 다른 실험 및 생물학적 확인과 결합되어야 한다는 점이다.

## Conclusions

이 고찰에서는 실험 설계부터 시료 수집 및 저장, 시퀀스 데이터의 시각화에 이르기까지 마이크로바이옴 연구를 수행하는 모든 단계가 결과와 그들의 생물학적 해석에 실질적으로 영향을 미치는 것에 대해 논의하였다. 표본 크기가 충분하지 않거나 검증(validation) 과정과 같은 실

수를 피하고, 적절한 표준(standard) 시료, 시료 처리의 표준화 및 기타 실수를 줄이는 노력을 한다면 미생물 군집 연구의 진전을 촉진할 수 있을 것이다. 기술의 표준화를 높이고 편견과 오류가 낮은 실험 방법의 보급은 미생물 군집 영역의 능력을 크게 증가시켜 실험실 규모의 연구에서부터 임상, 자연환경에 대한 응용 가능성을 높일 것이다.

## Acknowledgements

This work was supported by a 2-year research grant of Pusan National University.

## Conflicts of Interest

No potential conflict of interest relevant to this article was reported.

## References

- Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, Lakshmanan A, Wade WG. The human oral microbiome. *J Bacteriol* 2010;192:5002–17. doi: 10.1128/JB.00542–10.
- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* 2005;43:5721–32. doi: 10.1128/JCM.43.11.5721–5732.2005.
- Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, Huttenhower C, Izard J. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* 2012;13:R42. doi: 10.1186/gb-2012-13-6-r42.
- Colombo AP, Boches SK, Cotton SL, Goodson JM, Kent R, Haffajee AD, Socransky SS, Hasturk H, Van Dyke TE, Dewhirst F, Paster BJ. Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J Periodontol* 2009;80:1421–32. doi: 10.1902/jop.2009.090185.
- Belda-Ferre P, Alcaraz LD, Cabrera-Rubio R, Romero H, Simón-Soro A, Pignatelli M, Mira A. The oral metagenome in health and disease. *ISME J* 2012;6:46–56. doi: 10.1038/ismej.2011.85.
- Loesche WJ. Role of *Streptococcus mutans* in human dental decay. *Microbiol Rev* 1986;50:353–80.
- Corby PM, Bretz WA, Hart TC, Schork NJ, Wessel J, Lyons-Weiler J, Paster BJ. Heritability of oral microbial species in caries-active and caries-free twins. *Twin Res Hum Genet* 2007;10:821–8. doi: 10.1375/twin.10.6.821.
- Torlakovic L, Paster BJ, Ogaard B, Olsen I. Changes in the supragingival microbiota surrounding brackets of upper central incisors during orthodontic treatment. *Acta Odontol Scand* 2013;71:1547–54. doi: 10.3109/00016357.2013.776107.
- Demmer RT, Jacobs DR Jr, Singh R, Zuk A, Rosenbaum M, Papapanou PN, Desvarieux M. Periodontal bacteria and prediabetes prevalence in ORIGINS: the oral infections, glucose intolerance, and insulin resistance study. *J Dent Res* 2015;94(9 Suppl):201S–11S. doi: 10.1177/0022034515590369.
- Ozmeric N, Bissada N, da Silva APB. The association between inflammatory bowel disease and periodontal conditions: is there a common bacterial etiology? *J Int Acad Periodontol* 2018;20:40–51.
- Pietiläinen M, Liljestrand JM, Kopra E, Pussinen PJ. Mediators between oral dysbiosis and cardiovascular diseases. *Eur J Oral Sci* 2018;126 Suppl 1:26–36. doi: 10.1111/eos.12423.
- Cobb CM, Kelly PJ, Williams KB, Babbar S, Angolkar M, Derman RJ. The oral microbiome and adverse pregnancy outcomes. *Int J Womens Health* 2017;9:551–9. doi: 10.2147/IJWH.S142730.
- Woese CR. Bacterial evolution. *Microbiol Rev* 1987;51:221–71.
- Kim HJ, Li H, Collins JJ, Ingber DE. Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proc Natl Acad Sci U S A* 2016;113:E7–15. doi: 10.1073/pnas.

- 1522193112.
15. Candon S, Perez-Arroyo A, Marquet C, Valette F, Foray AP, Pelletier B, Milani C, Ventura M, Bach JF, Chatenoud L. Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of auto-immune insulin-dependent diabetes. *PLoS One* 2015;10:e0125448. doi: 10.1371/journal.pone.0125448.
  16. Gao Z, Guo B, Gao R, Zhu Q, Wu W, Qin H. Probiotics modify human intestinal mucosa-associated microbiota in patients with colorectal cancer. *Mol Med Rep* 2015;12:6119-27. doi: 10.3892/mmr.2015.4124.
  17. Bonder MJ, Tigchelaar EF, Cai X, Trynka G, Cenit MC, Hrdlickova B, Zhong H, Vatanen T, Gevers D, Wijmenga C, Wang Y, Zhernakova A. The influence of a short-term gluten-free diet on the human gut microbiome. *Genome Med* 2016; 8:45. doi: 10.1186/s13073-016-0295-y.
  18. Rajendhran J, Gunasekaran P. Microbial phylogeny and diversity: small subunit ribosomal RNA sequence analysis and beyond. *Microbiol Res* 2011;166:99-110. doi: 10.1016/j.micres.2010.02.003.
  19. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977; 74:5463-7. doi: 10.1073/pnas.74.12.5463.
  20. Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, Sahasrabudhe A, Dewhirst FE. Bacterial diversity in human subgingival plaque. *J Bacteriol* 2001;183:3770-83. doi: 10.1128/JB.183.12.3770-3783.2001.
  21. Yang B, Wang Y, Qian PY. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics* 2016;17:135. doi: 10.1186/s12859-016-0992-y.
  22. Kim M, Morrison M, Yu Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J Microbiol Methods* 2011;84:81-7. doi: 10.1016/j.mimet.2010.10.020.
  23. Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome* 2015;3:26. doi: 10.1186/s40168-015-0087-4.
  24. Meisel JS, Hannigan GD, Tyldsley AS, SanMiguel AJ, Hodkinson BP, Zheng Q, Grice EA. Skin microbiome surveys are strongly influenced by experimental design. *J Invest Dermatol* 2016;136:947-56. doi: 10.1016/j.jid.2016.01.016.
  25. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, Kurilshikov A, Bonder MJ, Valles-Colomer M, Vandeputte D, Tito RY, Chaffron S, Rymenans L, Verspecht C, De Sutter L, Lima-Mendez G, D'hoek K, Jonckheere K, Homola D, Garcia R, Tigchelaar EF, Eeckhaut L, Fu J, Henckaerts L, Zhernakova A, Wijmenga C, Raes J. Population-level analysis of gut microbiome variation. *Science* 2016;352:560-4. doi: 10.1126/science.aad3503.
  26. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334:105-8. doi: 10.1126/science.1208344.
  27. Jackson MA, Goodrich JK, Maxan ME, Freedberg DE, Abrams JA, Poole AC, Sutter JL, Welter D, Ley RE, Bell JT, Spector TD, Steves CJ. Proton pump inhibitors alter the composition of the gut microbiota. *Gut* 2016;65:749-56. doi: 10.1136/gutjnl-2015-310861.
  28. Kelly BJ, Gross R, Bittinger K, Sherrill-Mix S, Lewis JD, Colman RG, Bushman FD, Li H. Power and sample-size estimation for microbiome studies using pairwise distances and PERMANOVA. *Bioinformatics* 2015;31:2461-8. doi: 10.1093/bioinformatics/btv183.
  29. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biol* 2016;17:217. doi: 10.1186/s13059-016-1086-x.
  30. La Rosa PS, Brooks JP, Deych E, Boone EL, Edwards DJ, Wang Q, Sodergren E, Weinstock G, Shannon WD. Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PLoS One* 2012;7:e52078. doi: 10.1371/journal.pone.0052078.
  31. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* 2011;108 Suppl 1(Suppl 1):4554-61. doi: 10.1073/pnas.1000087107.
  32. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S; Microbiome Quality Control Project Consortium, Abnet CC, Knight R, White O, Huttenhower C. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol* 2017;35:1077-86. doi: 10.1038/nbt.3981.
  33. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, Knight R. Preservation methods differ in fecal microbiome stability, affecting suitability for field studies. *mSystems* 2016; 1:e00021-16. doi: 10.1128/mSystems.00021-16.

34. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS One* 2012;7:e39315. doi: 10.1371/journal.pone.0039315.
35. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–80. doi: 10.1038/nature03959.
36. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara E, Catenazzi M, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczky C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohna Sohna J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovskiy Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Rogers J, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Klenerman D, Durbin R, Smith AJ. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008;456:53–9. doi: 10.1038/nature07517.
37. Wei Y, Shi M, Zhen M, Wang C, Hu W, Nie Y, Wu X. Comparison of subgingival and buccal mucosa microbiome in chronic and aggressive periodontitis: a pilot study. *Front Cell Infect Microbiol* 2019;9:53. doi: 10.3389/fcimb.2019.00053.
38. Korlach J, Marks PJ, Cicero RL, Gray JJ, Murphy DL, Roitman DB, Pham TT, Otto GA, Foquet M, Turner SW. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proc Natl Acad Sci U S A* 2008;105:1176–81. doi: 10.1073/pnas.0710982105.
39. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011;475:348–52. doi: 10.1038/nature10242.
40. Reeder J, Knight R. The ‘rare biosphere’: a reality check. *Nat Methods* 2009;6:636–7. doi: 10.1038/nmeth0909–636.
41. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* 2017;11:2639–43. doi: 10.1038/ismej.2017.119.
42. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J* 2015;9:968–79. doi: 10.1038/ismej.2014.195.
43. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, Kightley EP, Thompson LR, Hyde ER, Gonzalez

- A, Knight R. Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2017;2:e00191-16. doi: 10.1128/mSystems.00191-16.
44. 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. doi: 10.1038/nmeth.3869.
45. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;73:5261-7. doi: 10.1128/AEM.00062-07.
46. Almeida A, Mitchell AL, Tarkowska A, Finn RD. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. *Gigascience* 2018;7:giy054. doi: 10.1093/gigascience/giy054.
47. Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nat Methods* 2010;7:813-9. doi: 10.1038/nmeth.1499.
48. Anderson MJ, Walsh DCI. PERMANOVA, ANOSIM, and the mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol Monogr* 2013;83:557-74. doi: 10.1890/12-2010.1.
49. Huang S, Li R, Zeng X, He T, Zhao H, Chang A, Bo C, Chen J, Yang F, Knight R, Liu J, Davis C, Xu J. Predictive modeling of gingivitis severity and susceptibility via oral microbiota. *ISME J* 2014;8:1768-80. doi: 10.1038/ismej.2014.32.
50. Ai D, Huang R, Wen J, Li C, Zhu J, Xia LC. Integrated metagenomic data analysis demonstrates that a loss of diversity in oral microbiota is associated with periodontitis. *BMC Genomics* 2017;18(Suppl 1):1041. doi: 10.1186/s12864-016-3254-5.