Research Techniques Made Simple: Profiling the Skin Microbiota

Max D. Grogan[1], Casey Bartow-McKenney[1], Laurice Flowers[1], Simon A.B. Knight[1], Aayushi Uberoi[1] and Elizabeth A. Grice[1,2]

Skin is colonized by microbial communities (microbiota) that participate in immune homeostasis, development and maintenance of barrier function, and protection from pathogens. The past decade has been marked by an increased interest in the skin microbiota and its role in cutaneous health and disease, in part due to advances in next-generation sequencing platforms that enable high-throughput, culture-independent detection of bacteria, fungi, and viruses. Various approaches, including bacterial 16S ribosomal RNA gene sequencing and metagenomic shotgun sequencing, have been applied to profile microbial communities colonizing healthy skin and diseased skin including atopic dermatitis, psoriasis, and acne, among others. Here, we provide an overview of culture-dependent and -independent approaches to profiling the skin microbiota and the types of questions that may be answered by each approach. We additionally highlight important study design considerations, selection of controls, interpretation of results, and limitations and challenges.

Journal of Investigative Dermatology (2019) **139,** 747–752; doi:10.1016/j.jid.2019.01.024

CME Activity Dates: 20 March 2019
Expiration Date: 19 March 2020
Estimated Time to Complete: 1 hour

Planning Committee/Speaker Disclosure: All authors, planning committee members, CME committee members and staff involved with this activity as content validation reviewers have no financial relationships with commercial interests to disclose relative to the content of this CME activity.

Commercial Support Acknowledgment: This CME activity is supported by an educational grant from Lilly USA, LLC.

Description: This article, designed for dermatologists, residents, fellows, and related healthcare providers, seeks to reduce the growing divide between dermatology clinical practice and the basic science/current research methodologies on which many diagnostic and therapeutic advances are built.

Objectives: At the conclusion of this activity, learners should be better able to:
- Recognize the newest techniques in biomedical research.
- Describe how these techniques can be utilized and their limitations.
- Describe the potential impact of these techniques.

CME Accreditation and Credit Designation: This activity has been planned and implemented in accordance with the accreditation requirements and policies of the Accreditation Council for Continuing Medical Education through the joint providership of Beaumont Health and the Society for Investigative Dermatology. Beaumont Health is accredited by the ACCME to provide continuing medical education for physicians. Beaumont Health designates this enduring material for a maximum of 1.0 *AMA PRA Category 1 Credit(s)*™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Method of Physician Participation in Learning Process: The content can be read from the Journal of Investigative Dermatology website: http://www.jidonline.org/current. Tests for CME credits may only be submitted online at https://beaumont.cloud-cme.com/RTMS-Apr19 – click 'CME on Demand' and locate the article to complete the test. Fax or other copies will not be accepted. To receive credits, learners must review the CME accreditation information; view the entire article, complete the post-test with a minimum performance level of 60%; and complete the online evaluation form in order to claim CME credit. The CME credit code for this activity is: 21310. For questions about CME credit email cme@beaumont.edu.

INTRODUCTION

The skin is an ecosystem that supports the growth of a plethora of indigenous microbiota consisting of bacteria, fungi, mites, and viruses. Skin commensal microbes coexist with the host and contribute to tissue integrity and immune homeostasis. Perturbation of skin commensal microbial communities can influence normal skin health, predispose skin to pathogenic colonization, and contribute to inflammatory dermatological disorders. The goal of most skin microbiota surveys is to identify individual taxa (e.g., genera, species, strains) or community features (e.g., diversity, richness) that are associated with a phenotype or a perturbation. Profiling the skin microbiota is often a jumping-off point for studies that seek to establish causation and/or to dissect the molecular and biochemical mechanisms of host-microbe crosstalk through reductionist approaches. Furthermore, because microbes are exquisitely sensitive to their environment, they are a reservoir of potential biomarkers that could

[1]*Department of Dermatology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, USA; and* [2]*Department of Microbiology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, Pennsylvania, USA*

Correspondence: Elizabeth A. Grice, 421 Curie Boulevard, BRB 1015, Philadelphia, Pennsylvania 19104. E-mail: egrice@pennmedicine.upenn.edu

Abbreviations: OTU, operational taxonomic unit; rRNA, ribosomal RNA

SUMMARY POINTS

Benefits

- Sequencing-based approaches do not require growth and isolation of microorganisms in culture and therefore select for microorganisms that do not readily grow in isolation under artificial conditions.
- Skin microbiota surveys are a powerful hypothesis-generating tool that produce quantitative, community-wide data sets.
- Several user-friendly bioinformatic tools have been developed for the analysis and visualization of microbiome sequencing data.
- Metagenomic shotgun sequencing is increasingly being applied to skin microbiota, and it provides strain-level taxonomic resolution and insight into the genetic repertoire of the microbiota.

Limitations

- Skin specimens are typically low in bioburden and extremely susceptible to reagent and environmental contamination, which produces false positive results.
- Culture-independent approaches cannot distinguish live versus dead microorganisms.
- The sequencing data obtained are associative, and additional experiments are required to show causality.
- Many analytical approaches require reference data sets, which are limited for skin microbes, especially fungi.

Box 1. Definitions and misnomers

The authors recommend the adoption of definitions proposed by Marchesi and Ravel (2015).

- Microbiota: The assemblage of microorganisms existing in a defined environment
- Microbiome: Microorganisms, their genetic material, and the surrounding environmental conditions
- Metagenomics: The application of shotgun sequencing of DNA to reveal the genomes and genes of the microbiota

Common misnomers to avoid (Marchesi and Ravel, 2015):

- 16S/16S analysis/16S survey: The proper term to apply here is *16S rRNA genes* or *16S rRNA gene sequencing/analysis* because the methods that these misnomers refer to rely on sequencing DNA and the gene that encodes a structural subunit of the ribosomal RNA. RNA is not sequenced, and 16S is a metric that refers to particle size.
- Microflora: Although widely used in the literature, the original definition of *flora* refers to plants and not microbes. The correct term to use is *microbiota*.

inform the status of skin health, distinguish between variations of disease, or suggest optimal treatment approaches. Therefore, exhaustive analyses of skin microbiota can not only provide better understanding of cutaneous processes and diseases but can also suggest targets for developing therapies.

With a growing appreciation for the importance of the human microbiome has come a surge in the development of next-generation sequencing technology and analytical tools that serve as the workhorses for characterizing microbial communities. Historically, detection and characterization of skin microbiota has depended on culture-based methods. Next-generation sequencing–based methods eliminate the biases associated with isolating and culturing microbes in the laboratory to more precisely profile the composition of microbial communities.

Amplicon-based sequencing is the most common strategy used to construct community profiles of skin microbiota. This method has been extensively used to characterize bacterial communities by targeting the highly conserved 16S ribosomal RNA (rRNA) gene, which contains hypervariable regions that are widely divergent among different bacterial taxa (Lane et al., 1985). More recently, metagenomic shotgun sequencing has been used for both taxonomic and functional annotation of skin microbial communities. This approach captures multikingdom communities (including fungi, viruses, and archaea) at the strain-level resolution and enables reconstruction of the community-level microbial genetic repertoire. Here, we provide an overview of the current approaches used to profile skin microbiota, the metrics associated with each, and the bioinformatic tools that are commonly used to analyze and visualize data. Please refer to Box 1 for definitions of terms used commonly throughout, as well as misnomers to avoid.

APPROACHES AND METHODS

Collection and processing of skin microbiota specimens and controls

The first step in any study to profile the skin microbiota requires collection of a microbial specimen (Figure 1), and the collection technique can profoundly influence study results. Although a standardized protocol for skin microbiome studies remains to be established, many investigators use a noninvasive, easy-to-perform swabbing technique. Collection techniques, including preparation of the skin, were recently comprehensively reviewed by Kong et al. (2017). Whatever technique is chosen, its application should be consistent across all specimens that are collected and compared in the study. A well-designed study also controls for factors that might affect the existing skin microbial community or expose skin to foreign communities (Goodrich et al., 2014). For example, many studies exclude participants whose skin was exposed to systemic or topical

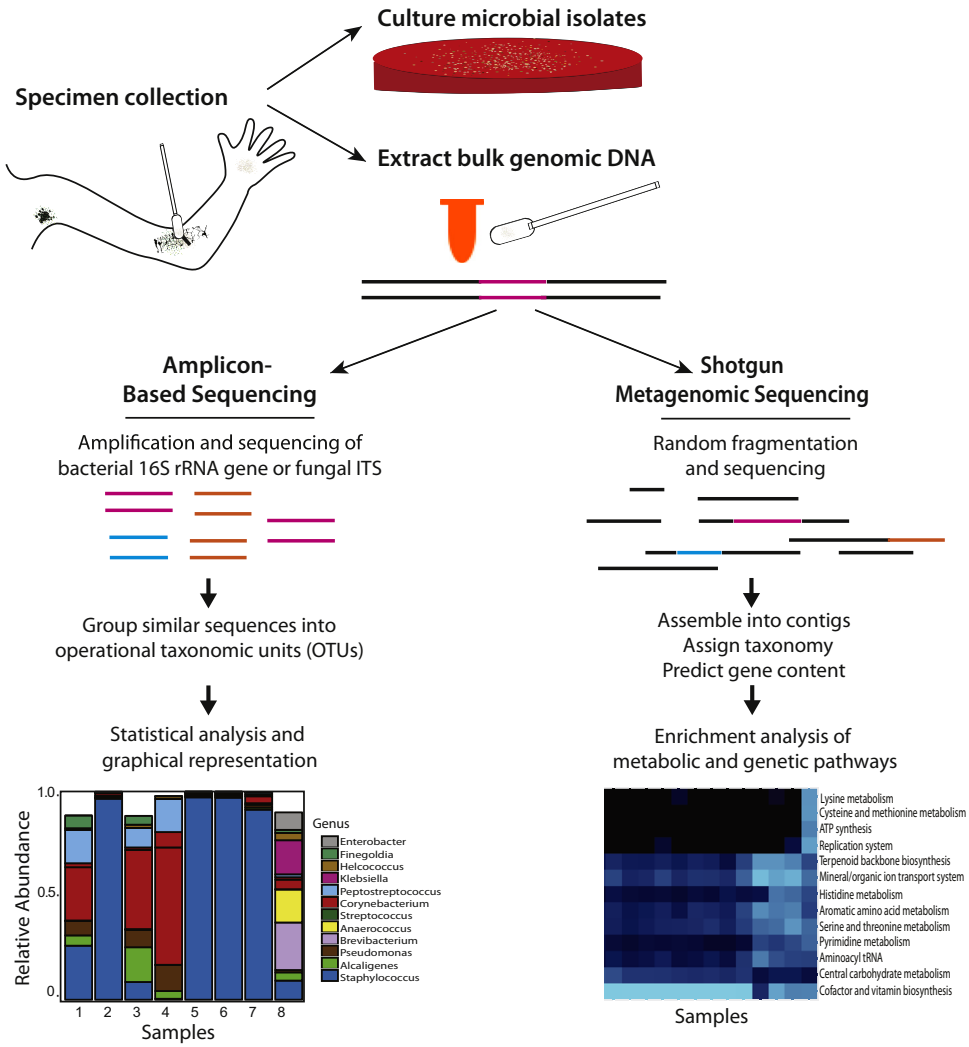

Figure 1. Approaches to profiling the skin microbiota. A specimen is collected and then subjected to culture-dependent and/or culture-independent techniques. Typical workflows for amplicon-based sequencing and shotgun metagenomic sequencing are compared. Examples of output for each are shown as a stacked bar plot depicting relative abundances of bacteria and a heatmap illustrating enrichment of different genetic and metabolic pathways among samples. OTU, operational taxonomic unit; rRNA, ribosomal RNA.

antimicrobials. DNA extraction techniques also should be consistent across studies and ideally performed using purposely designed kits that use a combination of chemical and physical lysis methods (e.g., detergents and bead beatings), followed by an isolation protocol that minimizes DNA loss and contamination (Goodrich et al., 2014).

Negative controls are a critical component of any well-designed skin microbiome study because they allow empirical assessment of background contamination from reagents and the environment. This is of particular concern for skin samples that are relatively low in bioburden (Salter et al., 2014), and proper steps need to be taken to minimize and/or remove contaminants (de Goffau et al., 2018; Kim et al., 2017). A negative control, null-exposure specimen should be collected and processed through DNA extraction, library preparation, and sequencing exactly as the experimental specimens. It is also critical to include positive controls. Sequencing of a mock community, containing microbial DNA from known organisms in known quantities, allows one to benchmark experimental approaches and pipelines. These positive controls can be generated and validated in house or purchased from a repository.

Amplicon-based sequencing approaches

The 16S rRNA gene provides a highly suitable target for bacterial classification by DNA sequencing. A description of this method, as it applies to the skin microbiome, has recently been described in the "Research Techniques Made Simple" series (Jo et al., 2016). In brief, this region of the bacterial genome consists of conserved and hypervariable regions and, in particular for the skin microbiome, the V1—V3 region was found to yield accurate results for taxonomic classification (Meisel et al., 2016). The V4 primers that are commonly used in studying microbiota from the gastrointestinal tract (Caporaso et al., 2012) require some minor modifications to capture the highly prevalent and abundant skin microbe *Cutibacterium acnes* (Zeeuwen et al., 2017). For the analysis of fungal communities, regions of DNA between the 18S, 5.8S, and 28S rRNA genes, termed *internal transcribed spacers*, contain both hypervariable regions and conserved regions for taxonomic identification and primer annealing, respectively. The internal transcribed spacer sequence resides in a much broader phylogenetic population and is thought of as a more "universal barcode" for fungi, but the variation also comes with less accuracy and specificity in taxonomic identification (Schoch et al., 2012). Additionally, fungi are less studied, and thus, their phylogenetic placement through computational methods and expert-based curation of phylogenetic relationships are lacking, which can be a limiting factor.

Most investigators rely on institutional cores or commercial operations to perform the sequencing. Here, we will focus more on the computational pipeline of analysis, which starts with the input of raw sequence data and ends with statistical analysis and graphical

representation of the microbial communities (Figure 1). The current recommended pipeline tools are QIIME2, mothur, and HmmUFOtu (Caporaso et al., 2010; Kuczynski et al., 2011; Schloss et al., 2009; Zheng et al., 2018). The first step in the pipeline is preprocessing, in which sequencing errors are eliminated. The next step is grouping of DNA sequences into operational taxonomic units (OTUs). This grouping is based on similarity or sequences that are close based on a defined sequence distance metric (Rossello-Mora and Amann, 2001). Because it is highly likely that all microbes in an environment are not known, OTUs have become the standard for cataloging the microbiome (Rossello-Mora and Amann, 2001; Schloss and Handelsman, 2005).

This process of OTU picking can be achieved in two ways: by matching sample sequences to a database of reference sequences (such as Greengenes [DeSantis et al., 2006]), or alternatively, the sequences can be clustered into de novo OTUs with no references. Once clustered, these OTUs are mapped to known sequences to determine the taxonomic composition of the sample (Caporaso et al., 2010; Zheng et al., 2018). At this stage in the pipeline, the output is an OTU table (formatted as a delimited text file or BIOM file). This table includes all OTUs identified; their abundance, or number of reads, found in each sample; and, usually, the taxonomy assigned to each OTU. The level of taxonomic classification varies in accuracy and is dependent on the region of the 16S rRNA gene sequenced and the identity of microbes in the sample. Results given at the species level should be interpreted cautiously unless customized (e.g., skin-specific) databases are being used (Conlan et al., 2012; Meisel et al., 2016).

Once taxonomic assignment is complete, the typical next step in the pipeline is to examine the diversity of the microbiome both within and between different samples, termed *alpha* and *beta diversity*, respectively. Most of the tools for this analysis were developed by the field of ecology and have been adapted to microbial community ecology. The pipeline tools QIIME2 or mothur have built-in plugins or programs to perform these analyses directly, but many alternative tools exist, particularly for users with statistical and bioinformatics backgrounds. A large collection of these downstream tools, such as those in the vegan package (Oksanen et al., 2018), can be installed into an R environment. R is an open-source computer language/environment designed for statistical analyses and graphical presentation of data (R Core Team, Vienna, Austria). A widely used R tool is phyloseq, which offers an intuitive suite of functions to aggregate data, perform statistical analysis, and graph the results (McMurdie and Holmes, 2013).

Statistical analysis and graphical presentation

Typically, microbiome data are nonparametric; the distribution of data (OTUs) is unknown, and thus, assumptions about the distribution should not be made when selecting statistical tests. Consequently, nonparametric statistics need to be used. For example, in place of *t* tests, an appropriate choice is the Mann-Whitney/Wilcoxon rank-sum test. Instead of applying an analysis of variance (i.e., ANOVA) test across more than 2 groups, the Kruskal-Wallis one-way analysis of variance test can be used, and the Spearman rank correlation coefficient should be used rather than the Pearson when examining co-occurrence of OTUs and/or taxa. Additionally, microbiome data are inherently multidimensional and thus require specialized tools. One of these is UniFrac (Lozupone and Knight, 2005), which uses a distance matrix that incorporates phylogenetic distances in comparing dissimilarity of microbial communities between two or more samples (beta diversity).

Multidimensional data can be challenging to display visually; three-dimensional graphs can be difficult to interpret, and four dimensions and greater cannot be drawn. To overcome this, a procedure termed *principal component analysis* (i.e., PCA) can be used. This is a statistical technique that transforms large sets of observations into a set of uncorrelated variables termed *principal components*, which emphasize the major differences in the data. The first principal component has the largest variation in the data, the second principal component has the next largest variation and is unrelated to the first. The first and second principal components are then plotted as a two-dimensional graph. Other methods that perform this task are nonmetric multidimensional scaling (i.e., nMDS), of which principal coordinates analysis (i.e., PCoA) is a subtype. The details of these methods are beyond the scope of this review, but they permit statistical analysis to be performed on the data in the form of a permutational multivariate ANOVA, or PERMANOVA. These statistical and graphical tools are available in the vegan R package. Many other statistical methods have been adopted for more specific analyses and graphing of microbiome data, including defining community types through Dirichlet multinomial clustering and identifying biomarkers by testing multiple decision tree models, in a process known as random forests, both of which can be performed in the R environment.

Shotgun metagenomic sequencing

Shotgun metagenomic sequencing, or the untargeted sequencing of all microbial genomes present in a specimen, is considerably richer in providing information than amplicon-based profiling approaches. Unlike amplicon-based sequencing, where specific primers are targeted to regions of rRNA genes, DNA is prepared for shotgun metagenomics by random fragmentation, addition of barcoded sequencing tags, and limited cycle amplification (Figure 1). Because shotgun metagenomics captures a greater variety of gene content in a sample, multikingdom compositions at strain-level resolution (an example can be found in Oh et al. [2014]), as well as functional profiles for communities, are captured. Shotgun metagenomics have provided key insights into the skin microbiome in atopic dermatitis, including the role of strain-level variation of *Staphylococcus aureus* (Byrd et al., 2017) and mechanistic understanding of how microbial metabolic pathways are altered to enhance ammonia production and increase skin pH (Chng et al., 2016).

Two different analytical approaches are used for shotgun metagenomic data sets: assembly-based and read-based profiling (for a comprehensive discussion, the authors recommend Quince et al. [2017]). Although read-based, assembly-free profiling is faster and mitigates issues with assembly, it relies on reference genomes at the expense of uncharacterized microbes that have no references available. A popular tool to generate taxonomic profiles without assembly is MetaPhlan, which maps shotgun reads to reference marker genes (Segata et al., 2012). These data may then be used to derive the alpha and beta diversity metrics previously described. Functional profiles can be produced using the HUMAnN tool (Abubucker et al., 2012) or similar, which takes the DNA reads and maps them against universal gene-protein databases. This allows identification of the proteins encoded by the DNA and functional pathway linkage of the proteins.

A case for integrating culture-dependent and -independent approaches

Sequencing technologies have illuminated the diversity of microbial species on the skin. However, when evaluating microbial data

observed from sequencing-based techniques, it is important to recognize the limitations. The approaches described measure only the presence of DNA in a sample. They are unable to show if the species was recently acquired or is a stable community member, a transient member, or deceased. The standard practice to identify microbial species in clinical settings relies on culture-based techniques. This traditional approach should not be disregarded in the design of research studies that take advantage of sequencing technologies, because culture-based techniques are able to identify viable organisms.

Ideally, samples for culturing should be processed immediately after collection. Depending on the study objective, various media and growth conditions can be used to quantify specific organisms. To successfully culture organisms that are traditionally difficult to grow, such as strict anaerobic bacteria, swabs should immediately be handled in anaerobic conditions. Therefore, if the study objective is to gain a comprehensive quantification of the most abundant microbial species, then separate specimens should be collected for aerobic and anaerobic growth. Once isolated, colonies can be identified either by 16S rRNA gene sequencing, matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (i.e., MALDI-TOF), or whole-genome sequencing, and comparisons can be made with culture-independent profiles obtained by sequencing-based approaches.

As described previously, reagent contamination is a major problem for low-bioburden microbial samples, including skin samples; examination of culture-based data and literature allows one to assess the plausibility of observing a given species in the ecosystem of the skin (de Goffau et al., 2018). For example, it is highly unlikely that extreme halophiles or thermophiles would be present on the skin, because the conditions and nutrients are not consistent with the biology of these microorganisms; nonetheless, results such as these continue to be reported in the literature in the absence of meaningful control data. Therefore, we urge readers to critically examine their data in the context of the biology of the microorganisms observed, which can be inferred from culture-based techniques.

CONCLUSIONS AND FUTURE DIRECTIONS

Culture-independent approaches for examining the skin microbiota have their own limitations that warrant consideration. These include but are not limited to (i) the inability to distinguish live versus dead organisms; (ii) reliance on reference databases that exclude uncharacterized microbes; (iv) reagent and environmental contamination, that when not properly controlled for, result in conclusions that are not consistent with cutaneous biology; and (iv) associative data sets that are unable to distinguish cause and effect. Additionally, the low bioburden of the skin has limited the application of techniques such as metatranscriptomics, which would allow the assessment of transcriptionally active microbiota. Because many of the analytical approaches require reference genomes, future efforts should focus on building comprehensive reference databases of skin-specific microbes, including yeasts, bacteria, viruses, and other microeukaryotes such as *Demodex* species. Even though a goal of the National Institutes of Health Human Microbiome Project was to create 3,000 microbial reference genomes for this purpose, at present count, only 124 of the 1,556 total

MULTIPLE CHOICE QUESTIONS

1. Which of the following is an advantage of culture-independent, sequencing-based approaches to analyzing skin microbiota?
   - A. It distinguishes microbes that are living from those that are dead.
   - B. It establishes causative links.
   - C. It does not require culturing microbes in artificial conditions.
   - D. It is difficult to contaminate reagents and samples.

2. What is the advantage of shotgun metagenomic sequencing compared with 16S rRNA gene sequencing?
   - A. Increased taxonomic resolution
   - B. Enhanced growth of microbes
   - C. Recovery of bacterial, fungal, and viral sequences
   - D. A and C

3. Which of the following is not a recommended practice when designing a study for culture-independent profiling of microbiota?
   - A. Including negative controls to assess background contamination
   - B. Using a variety of DNA extraction kits
   - C. Controlling for antibiotic exposures
   - D. Including a mock community as a positive control

4. Which of the following is a common and recommended practice when analyzing 16S rRNA gene sequencing data?
   - A. Testing associations/correlations with every single variable until something is significant
   - B. Using parametric statistical tests because microbiome data are always normally distributed
   - C. Assigning sequences to operational taxonomic units, or OTUs
   - D. Ignoring sequences in negative control samples

5. Which of the following is a bioinformatic tool/pipeline that is commonly used for the analysis of microbiome data sets?
   - A. QIIME2
   - B. R
   - C. mothur
   - D. All of the above

genomes sequenced were derived from skin (Joint Genome Institute, 2019). Finally, increased attention to robust study designs and inclusion of essential controls will enable the interpretation and translation of skin microbiome studies and their biological and/or clinical relevance.

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by awards from the National Institutes of Health (R01-NR-015639, R01-AR-006663 and R00-AR-060873 to EAG), Burroughs Wellcome Fund PATH Award (to EAG), and the Linda Pechenik Montague Investigator Award (to EAG). This research was also supported by the Penn Skin Biology and Disease Resource-Based Center (P30-AR-069589). CB-M and LF were supported by a Dermatology Research Training Grant (T32-AR-007465).

AUTHOR CONTRIBUTIONS

All authors participated in drafting and finalizing the manuscript and figures.

SUPPLEMENTARY MATERIAL

Supplementary material is linked to this paper. Teaching slides are available as supplementary material.

REFERENCES

Abubucker S, Segata N, Goll J, Schubert AM, Izard J, Cantarel BL, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput Biol 2012;8(6):e1002358.

Byrd AL, Deming C, Cassidy SKB, Harrison OJ, Ng WI, Conlan S, et al. *Staphylococcus aureus* and *Staphylococcus epidermidis* strain diversity underlying pediatric atopic dermatitis. Science Transl Med 2017;9(397).

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;7:335–6.

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;6:1621–4.

Chng KR, Tay AS, Li C, Ng AH, Wang J, Suri BK, et al. Whole metagenome profiling reveals skin microbiome-dependent susceptibility to atopic dermatitis flare. Nat Microbiol 2016;1(9):16106.

Conlan S, Kong HH, Segre JA. Species-level analysis of DNA sequence data from the NIH Human Microbiome Project. PLoS One 2012;7(10):e47075.

de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS, et al. Recognizing the reagent microbiome. Nat Microbiol 2018;3:851–3.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;72:5069–72.

Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a microbiome study. Cell 2014;158:250–62.

Jo J, Kennedy EA, Kong HH. Research Techniques Made Simple: Bacterial 16S ribosomal RNA gene sequencing in cutaneous research. J Invest Dermatol 2016;136:e23–7.

Joint Genome Institute. Integrated microbial genomes. https://img.jgi.doe.gov/cgi-bin/imgm_hmp/main.cgi; 2019 (accessed February 18, 2019).

Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. Microbiome 2017;5:52.

Kong HH, Andersson B, Clavel T, Common JE, Jackson SA, Olson ND, et al. Performing skin microbiome research: a method to the madness. J Invest Dermatol 2017;137:561–8.

Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R. Using QIIME to analyze 16s rRNA gene sequences from microbial communities. Curr Protoc Bioinformatics 2011;36:10.7.1–10.7.20.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci USA 1985;82:6955–9.

Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 2005;71:8228–35.

Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome 2015;3:31.

McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 2013;8(4):e61217.

Meisel JS, Hannigan GD, Tyldsley AS, San Miguel AJ, Hodkinson BP, Zheng Q, et al. skin microbiome surveys are strongly influenced by experimental design. J Invest Dermatol 2016;136:947–56.

Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, et al. Biogeography and individuality shape function in the human skin metagenome. Nature 2014;514(7520):59–64.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. Vegan: community ecology package. R package version 2.4-6. http://CRAN.R-project.org/package=vegan; 2018 (accessed February 18, 2019).

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017;35:833–44.

Rossello-Mora R, Amann R. The species concept for prokaryotes. FEMS Microbiol Rev 2001;25:39–67.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 2014;12:87.

Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 2005;71:1501–6.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 2009;75:7537–41.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 2012;109(16):6241–6.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 2012;9(8):811–4.

Zeeuwen P, Boekhorst J, Ederveen THA, Kleerebezem M, Schalkwijk J, van Hijum S, et al. Reply to Meisel et al. J Invest Dermatol 2017;137:961–2.

Zheng Q, Bartow-McKenney C, Meisel JS, Grice EA. HmmUFOtu: an HMM and phylogenetic placement based ultra-fast taxonomic assignment and OTU picking tool for microbiome amplicon sequencing studies. Genome Biol 2018;19(1):82.

DETAILED ANSWERS

1. **Which of the following is an advantage of culture-independent, sequencing-based approaches to analyzing skin microbiota?**

Answer: C. It does not require culturing microbes in artificial conditions.

Culture-independent approaches do not require isolation and growth of microbiota. However, limitations of these approaches include (i) the inability to distinguish live versus dead microorganisms; (ii) the inability to distinguish cause and effect in changes in the microbiota, which are observed between two conditions; and (iii) the fact that they it is very sensitive to reagent and environmental contamination.

2. **What is the advantage of shotgun metagenomic sequencing compared with 16S rRNA gene sequencing?**

Answer: D. A and C

Shotgun metagenomic sequencing allows the investigator to recover not only bacterial but also fungal and viral sequences, which are also important constituents of the skin microbiota. Shotgun metagenomics also facilitates strain-level resolution of skin microbiota, whereas amplicon-based approaches are usually limited to genus-level taxonomic identification. Neither approach requires the growth of microbes in culture. Thus, A and C are the correct answers.

3. **Which of the following is not a recommended practice when designing a study for culture-independent profiling of microbiota?**

Answer: B. Using a variety of DNA extraction kits

The same brand and type of DNA extraction kit (and the same lot number, if possible) should be consistently used throughout the study. Data obtained with different DNA extraction methods are not comparable in the same study, because of biases from lysis method, efficiency, and DNA recovery.

4. **Which of the following is a common and recommended practice when analyzing 16S rRNA gene sequencing data?**

Answer: C. Assigning sequences to operational taxonomic units, or OTUs

A common step in most 16S rRNA gene sequencing pipelines is the assignment of sequences to OTUs. Parametric tests are not recommended for compositional, nonnormally distributed data sets. Negative control samples should always be processed and analyzed alongside experimental samples to assess and control for empirical background contamination. The practice of exhaustively performing statistical tests on a data set to search for correlations and associations is known as *P*-hacking. A scientifically rigorous experimental design requires stating a hypothesis before performing the statistical test, and all outcomes, whether significant or not, are reported.

5. **Which of the following is a bioinformatic tool/pipeline that is commonly used for the analysis of microbiome data sets?**

Answer: D. All of the above

All answers are correct. QIIME2 and mothur are platforms specifically designed for microbial community analysis. R is often used for custom or more advanced statistical testing and visualization.