produce pathogenic effects in certain host and environmental contexts (a pathologic community).

The ability to characterize microbial communities without culturing their component members has spawned the field of *metagenomics* (Table 86e-1). Metagenomics reflects a confluence of experimental and computational advances in the genome sciences as well as a more ecologic understanding of medical microbiology, according to which the functions of a given microbe and its impact on human biology depend on the context of other microbes in the same community. Traditional microbiology relies on culturing individual microbes, but metagenomics skips this step, instead sequencing DNA isolated directly from a given microbial community. The resulting datasets facilitate follow-up functional studies, such as the profiling of RNA and protein products expressed from the microbiome or the characterization of a microbial community's metabolic activities.

Metagenomics provides insight into how microbial communities vary in several situations critical to human health. One such situation is how microbial communities are assembled following birth and how they operate over time, including responses of established communities to various perturbations. Another is how microbial communities normally vary between different anatomic sites within an individual and between different groups of people representing different ages, physiologic states, lifestyles, geographies, and gender. Yet another is how microbial communities vary in disease; whether such variations are consistent among individuals grouped according to current criteria for a disease or its subtypes; whether the microbiota or microbiome provides new ways of classifying disease states; and, importantly, whether the structural and functional configurations of microbial communities are a cause or a consequence of disease.

Analysis of our microbiomes also addresses one of the most fundamental questions in genetics: How does environment select our genes and directly influence their function? Each human encounters a unique environment during the course of his or her lifetime. Part of this personally experienced environment is incorporated into the genes and capabilities of our microbial communities. The microbiome therefore expands our conceptualization of "human" genetic potential from a single set of genes "fixed" at birth to a microbiome with additional genes and capabilities acquired via a process influenced by our family and life experiences, including modifiable lifestyle choices such as diet. This view recognizes a previously underappreciated dimension of human evolution that occurs at the level of our microbiomes and inspires us to determine how—and how fast—this microbial evolution effects changes in our human biology. For example, Westernization is associated with loss of bacterial species diversity (richness) in the microbiota, and this loss may be associated with the suite of Western diseases. The study of our microbiomes also raises important questions about personal identity, how we define the origins of health disparities, and privacy. Further, it offers the possibility of entirely new approaches to disease prevention and treatment, including *regenerative* medicine, which involves administration of microbial species (*probiotics*) to individuals harboring communities that have not developed into a mature, fully functional state or that have been perturbed in ways that can be restored by the addition of species that fill unoccupied "jobs" (niches).

This chapter provides a general overview of how human microbial communities are analyzed; reviews ecologic principles that guide our understanding of microbial communities in health and disease; summarizes recent studies that establish correlations and, in some cases, causal relationships between our microbiota/microbiomes and various diseases; and discusses challenges faced in the translation of these findings to new therapeutic interventions.

A TOOLBOX FOR METAGENOMIC ANALYSES OF HUMAN MICROBIAL COMMUNITIES

Life on Earth has been classified into three domains: Bacteria, Archaea, and Eukarya. The habitats of the surface-exposed human body harbor members of each domain plus their viruses. In large part, microbial diversity has not been characterized by culture-based approaches, partly because we do not know how to re-create the metabolic milieu fashioned by these communities in their native habitats and partly because a few organisms tend to outgrow the others. Culture-independent methods readily identify which organisms are present in a microbiota and their relative abundance. The gene widely used to identify microbes and their evolutionary relationships encodes the major RNA component of the small subunit (SSU) of ribosomes. Within each domain of life, the SSU gene is highly conserved, allowing the SSU gene sequences present in different organisms in that domain to be accurately aligned and regions of nucleotide sequence variation to be identified. Pairwise comparisons of SSU ribosomal RNA (rRNA) genes from different microbes allow construction of a phylogenetic tree that represents an evolutionary map on which previously unknown organisms can be assigned a position. This approach, known as *molecular phylogenetics*, permits characterization of each organism on the basis of its evolutionary distance from other organisms. Different phylogenetic types (*phylotypes*) can be viewed as comprising branches on an evolutionary tree.

Characterization of Bacteria Because members of the Bacteria dominate our microbiota, most studies defining our various body habitat–associated microbial communities have sequenced the bacterial SSU gene that encodes 16S rRNA. This gene has a mosaic structure, with highly conserved domains flanking more variable regions. The most straightforward way to identify bacterial taxonomic groups (*taxa*) in a given community is to sequence polymerase chain reaction (PCR) products (*amplicons*) generated from the 16S rRNA genes present in that community. PCR primers directed at the conserved regions of the gene yield PCR amplicons encompassing one or more of that gene's nine variable regions. PCR primer design is critical: differential annealing with primer pairs designed to amplify different variable regions can lead to over- or underrepresentation of specific taxa, and different regions within the 16S rRNA gene can have different patterns of evolution. Therefore, caution must be exercised in comparisons of the relative abundance of taxa in samples characterized in different studies, as methodologic differences can lead to larger perceived differences in the inferred taxonomy than actually exist.

A key innovation is *multiplex sequencing*. Amplicons from each microbial-community DNA sample are tagged by incorporation of a unique oligonucleotide barcode into the PCR primer. Amplicons harboring these sample-specific barcodes can then be pooled together so that multiple samples representing multiple communities can be sequenced simultaneously **(Fig. 86e-1)**. One important choice is the tradeoff between the number of samples that can be processed simultaneously and the number of sequences generated per sample. Interpersonal differences in the bacterial components of the microbiota are typically large, as are differences between communities occupying different body habitats in the same individual (see below); thus fewer than 1000 16S rRNA reads are characteristically required to discriminate community type. However, the identification of systematic differences in microbiota composition that correlate with physiologic status or disease state is confounded by the substantial interpersonal variation that occurs normally.

Sequencing of bacterial 16S rRNA genes creates a challenge for medical microbiology: how to define the taxonomic groups present in a community in a systematic and informative manner, so that one community can be compared with and contrasted to another. Within each domain of life, microbes are classified in a hierarchy beginning with phylum (the broadest group) followed by class, order, family, genus, and species. To determine taxonomy, 16S rRNA sequences are aligned on the basis of their sequence similarity—a process known as *picking operational taxonomic units* (OTUs). Grouping of 16S rRNA sequences from a given variable region into "bins" that share ≥97% nucleotide sequence identity (97%ID OTUs) is a commonly accepted, albeit arbitrary, way to define a species.

Looking beyond the 16S rRNA gene, we find that different isolates (strains) of a given bacterial species have overlapping but not identical sets of genes in their genomes. The aggregate set of genes identified in all isolates (strains) of a given species-level phylotype represents its *pan-genome*. Most species are represented by multiple strains, sometimes with markedly different functions (for example,