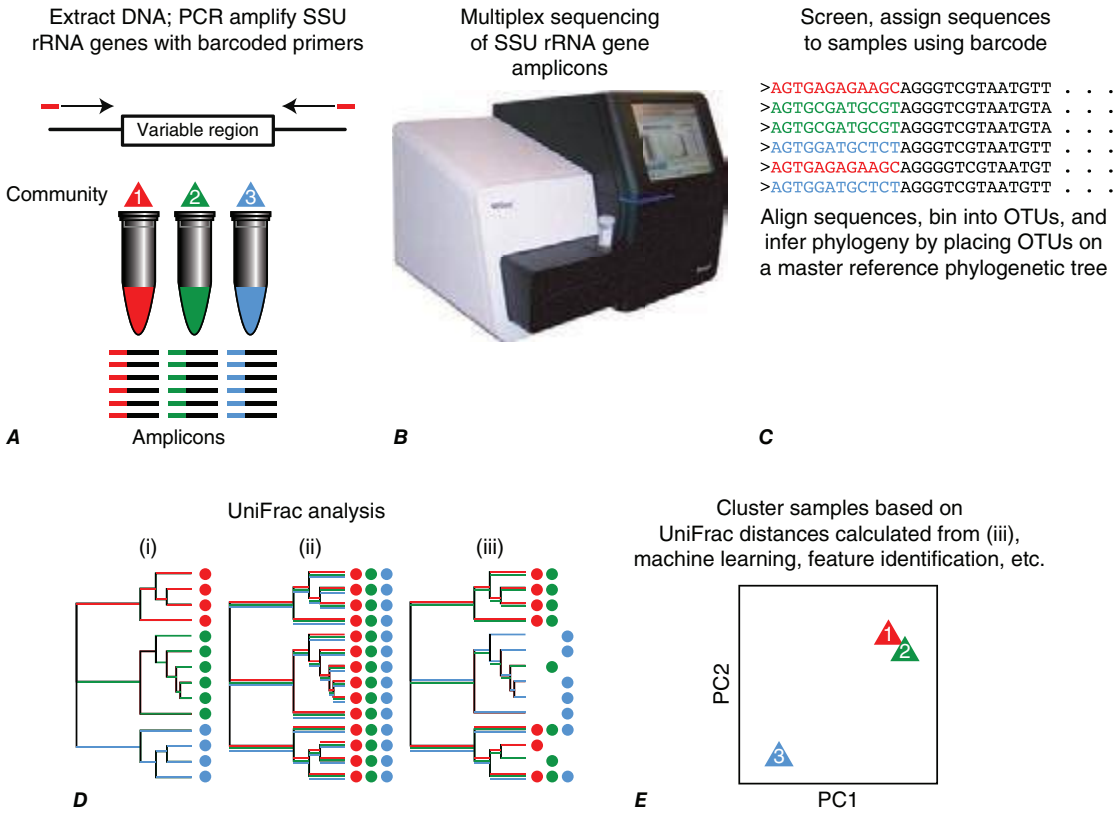

FIGURE 86e-1 Pipeline for culture-independent studies of a microbiota. (***A***) DNA is extracted directly from a sampled human body habitat–associated microbial community. The precise location of the community and relevant patient clinical data are collected. Polymerase chain reaction (PCR) is used to amplify portions of small-subunit (SSU) rRNA genes (e.g., the genes encoding bacterial 16S rRNA) containing one or more variable regions. Primers with sample-specific, error-correcting barcodes are designed to recognize the more conserved regions of the 16S rRNA gene that flank the targeted variable region(s). (***B***) Barcoded amplicons from multiple samples (communities 1–3) are pooled and sequenced in batch in a highly parallel next-generation DNA sequencer. (***C***) The resulting reads are then processed, with barcodes denoting which sample the sequence came from. After barcode sequences are removed in silico, reads are aligned and grouped according to a specified level of shared identity; e.g., sequences that share ≥97% nucleotide sequence identity are regarded as representing a species. Once reads are binned into operational taxonomic units (OTUs) in this fashion, they are placed on a phylogenetic tree of all known bacteria and their phylogeny is inferred. (***D***) Communities can be compared to one another by either taxon-based methods, in which phylogeny is not considered and the number of shared taxa are simply scored, or phylogenetic methods, in which community similarity is considered in light of the evolutionary relationships of community members. The UniFrac metric is commonly used for phylogeny-based comparisons. In stylized examples (i), (ii), and (iii), communities with varying degrees of similarity are shown. Each circle represents an OTU colored on the basis of its community of origin and placed on a master phylogenetic tree that includes all lineages from all communities. Branches (horizontal lines) are colored with each community that contains members from that branch. The three examples vary in the amount of branch length shared between the OTUs from each community. In (i), there is no shared branch length, and thus the three communities have a similarity score of 0. In (ii), the communities are identical, and a similarity score of 1 is assigned. In (iii), there is an intermediate level of similarity: communities represented in red and green share more branch length and thus have a higher similarity score than red vs. blue or green vs. blue. The amount of shared branch length in each pairwise community comparison provides a distance matrix. (***E***) The results of taxon- or phylogeny-based distance matrices can be displayed by principal coordinates analysis (PCoA), which plots each community spatially such that the largest component of variance is captured on the x-axis (PC1) and the second largest component of variance is displayed on the y-axis (PC2). In the example shown, the three communities in example (iii) from panel D are compared. Note that for shotgun sequencing of whole-community DNA (microbiome analysis), reads are compared with genes that are present in the genomes of sequenced cultured microbes and/or with genes that have been annotated by hierarchical functional classification schemes in various databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG). Communities can then be compared on the basis of the distribution of functional groups in their microbiomes—an approach analogous to taxon-based methods for 16S rRNA–based comparisons—and the results plotted with PCoA.

enteropathogenic versus commensal *Escherichia coli*). Differences in genome content among strains of a given species reflect differences in community membership as well as differences in the selective pressures these strains experience within and between habitats. Horizontal gene transfer among members of a microbiota—mediated by phage, plasmids, and other mechanisms—is a major contributor to this strain-level variation.

Strain-level diversity can be important in any consideration of how microbial communities differ between individuals and how these communities accommodate perturbations. For example, the great bacterial strain-level diversity that exists in the gut is thought to be one of the features that allows this microbiota, which occupies a constantly perfused ecosystem exposed to the complex and varying set of substances we ingest, to adapt to changing circumstances rather than depending on one strain to occupy a given niche important for proper community functioning. In ecologic studies of different environments, such as grasslands, forests, and reefs, increased diversity within a community increases its capacity to respond to disturbances and to restore itself (i.e., its *resilience*); the same is likely true of microbial ecosystems. When characterizing the mechanisms by which a given species produces an effect or effects on humans, it is important to consider the strain being tested; strain-level diversity has an impact on discovery and development efforts aimed at identifying next-generation probiotics that can be used therapeutically to promote health or treat disease.