(a measure of the total branch length of a phylogenetic tree encompassing a sample). Diversity estimators are particularly sensitive to errors introduced during PCR and sequencing.

Beta Diversity *Beta diversity* refers to the differences between communities and can be defined with phylogenetic or nonphylogenetic distance measurements. UniFrac is a commonly used phylogenetic metric that compares the evolutionary history of different microbial communities, noting the degree to which any two communities share branch length on a tree of microbial life: the more similar communities are to each other, the more branch length they share (Fig. 86e-1). UniFrac-based measurements of distances between communities can be visually represented with principal coordinates analysis or other geometric techniques that project a high-dimensional dataset down onto a small number of dimensions for a more approachable analysis (Fig. 86e-1). Principal coordinates analysis can also be applied to nonphylogenetic methods for comparing communities, such as Euclidean distance, Jensen-Shannon divergence, or Bray-Curtis dissimilarity, which operate independent of evolutionary tree data but can make biologic patterns more difficult to identify. The taxonomic data or distance matrices can also be used as input into a range of machine-learning algorithms (such as Random Forests) that employ supervised classification to identify differences between labeled groups of samples. Supervised classification is useful for identifying differences between cases and controls but can obscure important patterns intrinsic to the data, including confounding variables such as different sequencing runs or patient populations.

As noted above, the greatest beta diversity is that among body sites. This fact underscores the need to specify body habitat in microbiota analyses of any type, including microbial surveillance studies examining the flow of normal and pathogenic organisms into and out of different body sites in patients and their health care providers. Several other key points have emerged from beta diversity studies of human-associated microbial communities—notably, that (1) there is a high level of *inter*personal variability in every body habitat studied to date, (2) *intra*personal variation in a given body habitat is less pronounced, and (3) family members have more similar communities than unrelated individuals living in separate households. Thus, a person is his/her own best control, and examination of an individual over time as a function of disease state or treatment intervention is desirable. Similarly, family members serve as logical reference controls, although age is a major covariate that affects microbiota structure.

Studies of fecal samples obtained from twins over time have shown that the overall degree of phylogenetic similarity of bacterial communities does not differ significantly between monozygotic and dizygotic twin pairs, although monozygotic twin pairs may be more similar in some populations at earlier ages. These results, together with intervention studies in mice and epidemiologic observations in humans, emphasize that early environmental exposures are a very important determinant of adult-gut microbial ecology. In humans, the initial exposures depend on delivery mode: babies sampled within 20 min of birth have relatively undifferentiated microbial communities in the mouth, the skin, and the gut. For vaginally delivered babies, these communities resemble the specific microbial communities found in the mother's vagina. For babies delivered by cesarean section, the communities resemble skin communities. Although studies of older children and of adults stratified by delivery mode are still rare in the literature, these differences have been shown to persist until at least 4 months of age and perhaps until age 7 years. The infant gut microbiota changes to resemble the adult gut community over the first 3 years of life; comparable studies have not been done in other body habitats to date.

Exposures to environmental microbial reservoirs can continue to influence community structure. For example, unrelated cohabiting adults have more similar microbiotas in all of their body habitats than do non-cohabiting adults, and humans resemble the dogs they live with, at least in terms of skin microbiota. Gender and sexual maturation may also affect the microbiota structure, although efforts to isolate these variables are complicated by many confounding factors; any gender effect must be small compared with the effects of other variables such as diet (except in the case of the female urinary tract, which is influenced by the vaginal microbiota).

The vaginal microbiota illustrates another intriguing aspect of the contributions made by various factors to interpersonal differences in microbial community structure within a given body habitat. Bacterial 16S rRNA–based studies of the midvaginal microbiota in sexually active women have documented significant differences in community configurations between four self-reported ethnic groups: Caucasian, black, Hispanic, and Asian. Unlike most other body habitats that have been surveyed, this ecosystem is dominated by a single genus, *Lactobacillus*. Four species of this genus together account for more than half of the bacteria in most vaginal communities. Five community categories have been defined: four are dominated by *L. iners*, *L. crispatus*, *L. gasseri*, and *L. jensenii*, respectively, and the fifth has proportionally fewer lactobacilli and more anaerobes. The representation of these community categories is distinct within each of the four ethnic groups and correlates with vaginal pH and Nugent score (the latter being a biomarker for bacterial vaginosis). Longitudinal studies of individuals are being conducted to identify factors that determine the assembly of these distinct communities—both within and among ethnic groups—as well as their resistance to or resilience after various physiologic and pathologic disturbances. For example, the menstrual cycle and pregnancy turn out to be surprisingly significant factors (cause larger changes) compared with sexual activity.

Yet another factor affecting beta diversity is spatial location within a habitat. Several surveys show that the skin harbors bacterial communities with predictable, albeit complex, biogeographic features. To determine whether these differences are due to differences in local environmental factors, to the history of a given site's exposure to microbes, or to a combination of the two, reciprocal microbiota transplantation has been performed. Microbial communities from one region of the skin were depleted by treatment with germicidal agents, and the region (*plot*) was inoculated with a "foreign" microbiota harvested from different regions of the skin or from different body habitats from the same or another individual. Community assembly at the site of transplantation was then tracked over time. Remarkably, assembly proceeded differently at different sites: forearm plots receiving a tongue microbiota remained more similar to tongue communities than to native forearm communities in terms of their composition and diversity, while forehead plots inoculated with tongue bacteria changed to become more similar to native forehead communities. Thus, in addition to the history of exposure to tongue bacteria, environmental factors operating at the forehead plot likely shape community assembly. Intriguingly, the factors that shape fungal skin communities appear to be entirely different from those that shape bacterial skin communities. The palm and forearm have high bacterial and low fungal diversity, whereas the feet have the opposite diversity pattern. Moreover, fungal communities are generally shaped by location (foot, torso, head), whereas bacterial communities are generally shaped by moisture phenotype (dry, moist, or sebaceous).

Co-Occurrence Analysis *Co-occurrence analysis* seeks to identify which phylotypes are co-distributed across individuals in a given body habitat and/or between habitats and to determine the factors that explain the observed patterns of co-distribution. Positive correlations tend to reflect shared preferences for certain environmental features, while negative correlations typically reflect divergent preferences or a competitive relationship. *Syntrophic* (cross-feeding) relationships reflect interdependent interactions based on nutrient-sharing strategies. For example, in *food webs*, the products of one organism's metabolism can be used by the other for its own unique metabolic capabilities (e.g., the interactions between fermentative organisms and methanogens).

Enterotype Analysis *Enterotype analysis* seeks to classify individuals into discrete groups based on the configuration of their microbiotas, essentially drawing boundaries on a map defined by principal coordinates analysis or other ordination techniques. The first enterotype analysis used supervised clustering to define three major types of human-gut microbial configurations across three distinct human studies and provided a view that presupposed the existence of three