tissue specimens can be performed by specialty laboratories, though its sensitivity and clinical utility to date have remained somewhat limited because, for instance, of inhibitory molecules often found in clinical tissue samples that prevent reliable, sensitive PCR amplification. As such barriers are reduced through technological advances and as the causes of culture-negative infection are clarified (perhaps in part through sequencing efforts), these tests may become both more accessible and more helpful.

With the wealth of sequencing data now available, other regions beyond 16S rRNA can be targeted for bacterial species identification. These other genomic loci can provide additional information about a clinical isolate that is relevant to patient management. For instance, detection of the presence—or potentially even the expression—of toxin genes such as those for *C. difficile* toxins A and B or Shiga toxin may provide clinicians with additional information that will help distinguish commensals or colonizing bacteria from pathogens and thus aid in prognostication as well as diagnosis.

While amplification tests such as PCR exemplify one approach to nucleic acid detection, other approaches exist, including detection by hybridization. Although not currently used in the clinical realm, techniques for detection and identification of pathogens by hybridization to microarrays are being developed for other purposes. Of note, these different detection techniques require different degrees of conservation. Highly sensitive amplification methods require a high degree of sequence identity between PCR primer pairs and their short, specific target sequences; even a single base-pair mismatch (particularly near the 3′ end of the primer) may interfere with detection. In contrast, hybridization-based tests are more tolerant of mismatch and thus can be used to detect important regions that may be less precisely conserved within a species, thus potentially allowing detection of clinical isolates from a given species with greater diversity between isolates. Such assays take advantage of the predictable binding interactions of nucleic acids. The applicability of hybridization-based methods toward either DNA or RNA opens up the possibility of expression profiling, which can uncover phenotypic information from nucleic acid content.

Both PCR and hybridization methods target specific, known organisms. At the other extreme, as sequencing costs and turnaround times decrease, direct metagenomic sequencing from patient samples is becoming increasingly feasible. This shotgun sequencing approach is unbiased—i.e., is able to detect any microbial sequence, however divergent or unexpected. This new approach brings its own set of challenges, however, including the need to recognize pathogenic sequences against a background of expected host and commensal sequences and to distinguish true pathogens from either colonizers or laboratory contaminants. In a powerful example of this new frontier of sequencing-based clinical diagnosis, investigators diagnosed neuroleptospirosis in a child with an unexplained encephalitis syndrome by finding sequences corresponding to the *Leptospira* genus in cerebrospinal fluid from the patient. Rapid (<48-h) sequencing and analysis informed the patient's care in real time, leading to life-saving targeted antibiotic therapy for an unexpected diagnosis that was impossible to make through standard laboratory testing. The diagnosis was retrospectively confirmed through both convalescent serologies and PCR using primers designed on the basis of sequencing data.

PATHOGEN DISCOVERY

In addition to clinical diagnostic applications, novel genomic technologies, including whole-genome sequencing, are being applied to clinical research specimens with a goal of identifying new pathogens in a variety of circumstances. The tremendous sensitivity and unbiased nature of sequencing is also ideal in searching clinical samples for unknown or unsuspected pathogens.

Causal inference in infectious diseases has progressed since the time of Koch, whose historical postulates provided a rigorous framework for attributing a disease to a microorganism. According to an updated version of Koch's postulates, an organism, whether it can be cultured or not, should induce disease upon introduction into a healthy host if it is to be implicated as a causative pathogen. Current sequencing technologies are ideal for advancing this modern version of Koch's postulates because they can identify candidate causal pathogens with unprecedented sensitivity and in an unbiased way, unencumbered by limitations such as culturability. Yet, as direct sequencing on primary patient samples greatly expands our ability to recognize associations between microbes and disease states, critical thinking and experimentation will continue to be vital to establishing causality.

Virus discovery in particular has been greatly facilitated by new nucleic acid technology. These frontiers were first notably explored with high-density microarrays containing spatially arrayed sequences from a phylogenetically diverse collection of viruses. Although biased toward those with homology to known viruses, novel viruses in clinical samples were successfully identified on the basis of their ability to hybridize to these prespecified sequences. This methodology famously contributed to identification of the coronavirus causing severe acute respiratory syndrome (SARS). Once discovered, this SARS coronavirus was rapidly sequenced: the full genome was assembled in April 2003, less than 6 months after recognition of the first case. This accomplishment illustrated the advancing power and speed of new diagnostic technologies.

With the advent of next-generation sequencing, unbiased pathogen discovery is now possible through a process known as *metagenomic assembly* (Fig. 146-3). Sequences of random nucleotide fragments can be generated from clinical specimens with no a priori knowledge of pathogen identity through a process called *shotgun sequencing*. This collection of sequences can then be computationally aligned to host (i.e., human) sequences, with aligned sequences removed and remaining sequences compared with other known genomes to detect the presence of known microorganisms. Sequence fragments that remain unaligned suggest the presence of an additional organism that cannot be matched to a known, characterized genome; these reads can be assembled into contiguous nucleic acid stretches that can be compared to known sequences to construct the genome of a potentially novel organism. Assembled genomes (or parts of genomes) can then be compared to known genomes to infer the phylogeny of new organisms and identify related classes or traits. Thus, not only can this process identify unanticipated pathogens; it can even identify undiscovered organisms. Some early applications of sequencing on clinical samples have centered around the discovery of novel viruses, including such emerging pathogens as West Nile virus, SARS coronavirus, and the Middle East respiratory syndrome coronavirus (MERS-CoV) that has caused severe respiratory illnesses in healthy adults, as well as viral causes of myriad other conditions, from tropical hemorrhagic fevers to diarrhea in newborns.

More recently, metagenomic assembly has been successfully extended to bacterial pathogen discovery. Investigators identified a new bacterial species associated with "cord colitis"—a rare antibiotic-responsive, culture-negative colitis in recipients of umbilical cord-blood stem cells—by sequencing colon biopsy samples from affected patients and matched controls. A single dominant species emerged from metagenomic assembly in samples from patients that was absent from control samples. The presence of this species was confirmed by PCR and fluorescence in situ hybridization on primary tissue samples. On the basis of its similarity to other known species, the organism was named *Bradyrhizobium enterica*, a novel species from a genus that has proved difficult to culture and thus would have been hard to identify by other means. Correlation versus causation remains an open question; therefore, further efforts will be required to make such links.

As metagenomic sequencing and assembly techniques become more robust, this technology holds great promise for identifying microorganisms that are associated with clinical conditions of unknown etiology. Conventional methods already have unexpectedly linked numerous conditions with specific agents of infection—e.g., cervical and oropharyngeal cancers with human papillomavirus, Kaposi's sarcoma with human herpesvirus 8, and certain lymphomas with Epstein-Barr virus. Sequencing techniques offer unprecedented sensitivity and specificity for identifying foreign nucleic acid sequences that may suggest other conditions—from malignancies to inflammatory conditions to unexplained fevers or other clinical syndromes—associated with organisms