Microbial Genomics Research Review

An Overview of Publications Featuring Illumina® Technology
# Table of Contents

Contents .......................................................................................................................... 2  
Introduction ..................................................................................................................... 3  
Mosaicism .......................................................................................................................... 4  
Fitness ................................................................................................................................. 6  
The Core Genome .............................................................................................................. 7  
Evolution in the Laboratory ............................................................................................ 8  
Antibiotic Resistance and Virulence ................................................................................ 9  
Epidemics and Transmission ............................................................................................ 11  
Microbial Identification .................................................................................................. 13  
Directed Evolution and Bioengineering ......................................................................... 14  
Biofuels and Bioremediation ......................................................................................... 15  
Viruses ............................................................................................................................... 16  
  Virus Detection and Identification ............................................................................... 16  
  Vaccine Production ...................................................................................................... 18  
Yeast .................................................................................................................................. 19  
Well-studied Microorganisms .......................................................................................... 20  
  Escherichia Coli .......................................................................................................... 20  
  Staphylococcus Aureus ............................................................................................... 21  
  Streptococcus ................................................................................................................ 22  
  Mycobacterium ............................................................................................................. 22  
  Salmonella ..................................................................................................................... 24  
Bibliography ....................................................................................................................... 25
Introduction

Next-generation sequencing\(^1\) lends itself particularly well to the microbial laboratory, where the genomes are small. The appealing difference between sequencing and all other laboratory measurements is that the results can be directly related to a genomic locus and a potential explanation of the biological impact. This represents a quantum step forward in the interpretation of the experimental results and understanding of the biological system. The second advantage is the ability to measure single base changes anywhere in the genome without any prior knowledge. Single base resolution allows us to track microbial adaptation over short periods of time, both in the laboratory and in the environment. The advantages are so profound that in the foreseeable future it will be difficult to imagine a biological laboratory without a sequencer.

Recent studies have shown that the genomes of biological systems are remarkably active in adapting to the laboratory and clinical environments. Historically the spread of global epidemics was followed over a period of years. With the single base resolution of next-generation sequencing applied to bacterial genomes, it is possible to rapidly track epidemics within a local population, hospital, or even within a family over a period of weeks.

This review highlights recent examples where Illumina sequencing technology is used to track rapid genetic adaptation in nature, the laboratory, and the clinic.

---

\(^1\) Next-generation sequencing (NGS) and massively parallel sequencing (MPS) are often used interchangeably to refer to high throughput sequencing technologies. Sequencing by synthesis (SBS) refers specifically to Illumina sequencing technology.
Mosaicism

Horizontal gene transfer is the movement of genetic material between bacteria other than by descent.\(^2\) This can happen through several mechanisms but the end result is mosaicism, where the majority of the organism's genome is composed of sequences inherited from its predecessors, with some fraction consisting of DNA fragments derived from other organisms in its environment. Since the exchange of genetic material is not by descent, it is unpredictable. This makes the microbial genome particularly challenging to analyze. Sequencing is the only tool that can accurately map mosaic genomes.

Horizontal transfer of functional genes, or even significant genomic rearrangements, may not be reported through methods that use markers representing only a small fraction of the genome.

Reviews:


References:


This paper presents a detailed phylogeny based on whole-genome sequencing of representative strains of C. trachomatis from both trachoma and lymphogranuloma venereum (LGV) biovars. It shows that predicting phylogenetic structure using ompA, which is traditionally used to classify Chlamydia, is misleading because extensive recombination in this region masks the true relationships.

**Illumina Technology:** Genome AnalyzerII


Vibrio mimicus, the species most similar to V. cholerae, is a microbe present in the natural environment. It is naturally not particularly virulent and sometimes causes diarrhea and internal infections in humans. Horizontal transfer of virulence-related genes from an uncommon clone of V. cholera has resulted in the pathogenic V. mimicus strain carrying cholera toxin genes. This is an outstanding example of horizontal gene transfer (mosaicism) and the value of whole-genome sequencing.

**Illumina Technology:** Genome Analyzer with 130x coverage.


Fitness

There is a trade-off in microbial populations between the optimal adaption of a homogeneous population and the maintenance of less-optimal variants that will survive when conditions change.\textsuperscript{3,4} For example, in a microbial population under antibiotic pressure, microbes that are antibiotic resistant will survive even though they may grow slower and appear less fit.\textsuperscript{5,6} The fitness landscape describes the possible mutational trajectories by which lineages evolve in a stepwise manner from genotypes that lie in regions of low fitness to ones of higher fitness.\textsuperscript{7}

A more systematic approach uses site-directed mutagenesis to introduce mutations into a pool of microbes before culturing or passage through an animal, followed by high throughput sequencing to detect the mutations in the survivors. This approach uses the strengths of high-throughput sequencing and is highly efficient.\textsuperscript{8,9}

References:


This is a detailed description of insertion sequencing (INSeq), a method for determining the insertion site and relative abundance of large numbers of transposon mutants in a mixed population of isogenic mutants of a sequenced microbial species. The protocol is easy to scale up, amenable to automation, and useful for a variety of samples.

Illumina Technology: Protocol specific for GA/HiSeq

---


\textsuperscript{7} Kvitek, D. J. and Sherlock, G. (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genet 7: e1002056


This study demonstrates the tradeoff between fitness and diversity in viral genomes.

**Illumina Technology:** Genome AnalyzerIIx


Nair, D., Memmi G., Hernandez D., Bard J., Beaume M., et al. (2011) Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. J Bacteriol 193: 2332-2335

**The Core Genome**

Within a taxonomic group there is an evolutionarily conserved set of core genes that is shared by all members of the group. This core genome is critical for survival. Individual members of the group also have accessory genes that contribute niche-specific phenotypes such as virulence and drug-resistance. It is important to distinguish between the core and accessory genes if we want to classify or modify the microbes. The core genome of conserved genes can be identified by comparing members of a taxonomic group. The core genome can also be determined experimentally by introducing random mutations and tracking the ability of the mutants to survive. If a critical gene is disabled, the microorganism will fail to survive under the culture conditions. This approach is very flexible and can be highly informative.

**Reviews**


**References:**


In this study, the authors used PanSeq 2.0 to analyze 39 Salmonella enterica genomes (16 closed, 23 draft). Salmonella is a species that contains two human-specific serovars that cause typhoid fever, as well as a large number of zoonotic serovars that cause gastroenteritis in humans. The authors achieved high levels of discrimination, even amongst the most closely related strains of S. enterica Typhi. This information can then be used to determine the potential core genes of the species.
Evolution in the Laboratory

In the laboratory, the genomes of model organisms accumulate mutations over time.\textsuperscript{11,12} As a result, the genomes of the same model organism may diverge over time and experiments will become progressively more difficult to replicate. Whole-genome resequencing has become a powerful tool to routinely monitor genomes and treat them just like any other experimental variable that needs to be recorded and controlled.

Based on the model of neutral evolution we expect bacteria to accumulate neutral mutations at a steady rate over time in a stable laboratory environment. In reality, the interaction between adaptive and neutral evolution is much more complex. A long-range study over 40,000 generations from a laboratory population of Escherichia coli showed that almost all the mutations appear to be beneficial and that there were sharp changes in neutral mutation over time in the apparently stable laboratory population.\textsuperscript{13} It is also notable, and somewhat unsettling, that there appears to be no saturation in the number of mutations accumulated even after 40,000 generations.\textsuperscript{14}

References:


The authors show that in short-term adaptive laboratory evolution (up to 40-50 days), Escherichia coli, under growth rate selection pressure, was found to undergo approximately 1011.2 total cumulative cell divisions within the population, producing a new stable growth phenotype that results from two to eight mutations. Continuous exposure to a low level of the mutagen can accelerate this timescale.

\textbf{Illumina Technology:} Genome Analyzer\textsubscript{II} 36 bp reads


This paper describes the “arms race” between a bacterium and a phage when they are copropagated. E. coli first adapted by developing partial resistance to infection and later by increasing specific growth rate. The phage counter-adapted by improving release efficiency with a change in host specificity and a decrease in virulence.

\textbf{Illumina Technology:} Genome AnalyzerIIx 51 bp single reads


Antibiotic Resistance and Virulence

The development of antibiotic resistance can be considered as a special case of directed evolution. It has been the subject of intense research due to the increasing prevalence of antibiotic resistant strains.

Due to the mosaic nature of the microbial genome, it can rapidly acquire antibiotic resistance genes or virulence-associated genes from the environment. The acquisition of these genes is often independent of serotype or gene markers. Next-generation sequencing of the whole genome has proved to be the ideal tool to comprehensively and unambiguously track antibiotic resistance- and virulence-associated genes.

The classic approach to studying the changes in antibiotic resistance and virulence is to compare isolates obtained from outbreaks and epidemics.\textsuperscript{15,16} A new approach follows the microbial adaptation to sustained antibiotic pressure in the laboratory. It takes advantage of the ability of next-generation sequencing to detect single nucleotide changes rapidly and cost-effectively. The advantages of this approach are that all variables can be carefully controlled in a laboratory environment and the results can be directly related to biological functions.\textsuperscript{17}

References:


The authors analyzed the evolution of resistance in Escherichia coli under selection with single drugs, including chloramphenicol, doxycycline and trimethoprim. Over a period of approximately 20 days, resistance levels increased dramatically with parallel populations showing similar phenotypic trajectories. Whole-genome sequencing of the evolved strains identified mutations both specific to resistance of a particular drug and shared in resistance to multiple drugs. This is an example of the power that routine sequencing can bring to a microbiology laboratory.

Illumina Technology: Genome Analyzer\textsubscript{IIx} 75 bp single-end reads


The comparisons of geographically and genetically diverse CA-MRSA genomes suggest that the apparent convergent evolution in CA-MRSA may be better explained by the rapid dissemination of a highly conserved accessory genome from a common source. This is a good example of how misleading clinical and epidemiological profiles can be and how important it is to sequence whole bacterial genomes when tracking epidemics.

**Illumina Technology:** Genome Analyzer IIx 36 bp paired-end reads


Harvey, R. M., Stroeher U. H., Ogunniyi A. D., Smith-Vaughan H. C., Leach A. J., et al. (2011) A variable region within the genome of Streptococcus pneumoniae contributes to strain-strain variation in virulence. PLoS ONE 6: e19650
Epidemics and Transmission

Epidemics and microbial transmission have traditionally been tracked with serological or other markers that monitor only a small, arbitrary part of the microbial genome. Due to the mosaic nature of the microbial genome any approach that monitors only a part of the genome will be relatively unreliable and insensitive. By contrast, next-generation sequencing can track every base in the genome, which has led to a revolution in our understanding of these processes. Transmission can now be tracked over relatively short periods of time, even within families or hospitals and the source of the outbreak can be determined. This allows a much more rapid and targeted response to outbreaks. An additional benefit is that the mutations that accumulate during an outbreak can provide information about the potential development of antibiotic resistance and changes in virulence.

Reviews:


References:


This paper presents a detailed phylogeny based on whole-genome sequencing of representative strains of Chlamydia trachomatis from both trachoma and lymphogranuloma venereum (LGV) biovars. It shows that predicting phylogenetic structure using ompA, which is traditionally used to classify Chlamydia, is misleading because extensive recombination in this region masks the presence of the true relationships.

Illumina Technology: Genome Analyzer

Reeves, P. R., Liu B., Zhou Z., Li D., Guo D., et al. (2011) Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS ONE 6: e26907

The authors report the genome sequences of 14 isolates of a uropathogenic E. coli clone that persisted for three years within a household, including a dog. The host data imply at least six host transfer events occurred over the three years, with two lineages present over much of that period. An earlier study using traditional typing techniques did not resolve the transmission.

Illumina Technology: Genome Analyzer

19 Reeves, P. R., Liu, B., Zhou, Z., Li, D., Guo, D., et al. (2011) Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS ONE 6: e26907
Reeves et al. Next-generation sequencing was able to resolve multiple host transfer events within a family and their dog over several years where traditional techniques had previously failed.


Comparison of the whole-genome sequences of Vibrio cholerae isolates from Haiti and Nepal showed that 24 Vibrio cholerae isolates from Nepal belonged to a single monophyletic group that also contained isolates from Bangladesh and Haiti. One cluster contained three Nepalese isolates and three Haitian isolates that were almost identical, with only 1 bp or 2 bp differences. Results in this study are consistent with Nepal as the origin of the Haitian outbreak.

Illumina Technology: Genome Analyzer\textsuperscript{II}x multiplexed 76 bp paired-end reads


This is a very interesting study of an accidental infection that occurred in the lab while working with Z5463, a Neisseria meningitidis serogroup A strain. The authors estimate that 25 bacterial divisions occurred in the infected human body. The in vivo passage, despite the small number of divisions, permitted the selection of numerous genomic modifications, which may account for the strain’s high capacity to spread.

Illumina Technology: Genome Analyzer giving a 154-, 84- and 78-fold coverage.


\textsuperscript{26} Reeves, P. R., Liu, B., Zhou, Z., Li, D., Guo, D., et al. (2011) Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS ONE 6: e26907
Microbial Identification

Traditional microbial identification relies on clinical symptoms and some prior knowledge to identify microorganisms. Some cases are atypical and defy identification. The agnostic nature of microbial detection and sequencing with next-generation sequencing makes it a very useful tool in those cases.

Sequencing does not require prior knowledge to identify microorganisms.

References:


A recent case of melioidosis in non-endemic Arizona was determined to be the result of locally acquired infection, as the patient had no travel history to endemic regions and no previous history of disease. Diagnosis of the case was confirmed through multiple microbiologic and molecular techniques. This is a nice example of using sequencing to identify a pathogen.

Illumina Technology: Genome Analyzer; 50 bp, paired-end reads

Directed Evolution and Bioengineering

Directed evolution is emerging as a promising new supplement to standard bioengineering techniques. Microbes adapt remarkably quickly to changes in the environment. By systematically changing the environment, researchers can track the changes in gene expression and the mutations incorporated by the organisms to achieve desirable characteristics, for example, adaptive evolution to grow on galactose. In replicated experiments, the organism may achieve the same desirable characteristics by modifying different pathways. This provides a spectrum of possible bioengineering solutions.

References:


This paper demonstrates the adaptation of a Lactococcus lactis strain isolated from a plant to a dairy niche by propagating it for 1000 generations in milk. Two out of three independently evolved strains displayed significantly increased acidification rates and biomass yields in milk. Reproducing the transition from the plant to the dairy niche through experimental evolution revealed several genome, transcriptome, and phenotype signatures that resemble those seen in strains isolated from either niche. This is an interesting type of experiment that uses next-generation sequencing effectively to understand the pathways involved in adaptive evolution and genetic engineering.

**Illumina Technology:** Genome Analyzer


This paper tracks metabolic changes occurring in the yeast Saccharomyces cerevisiae as a result of its adaptive evolution to grow on galactose. The study demonstrates that adaptive evolution represents a valuable alternative to rational design in the bioengineering of improved strains, and that it is possible to identify mutations in evolved strains that can serve as unforeseen metabolic engineering targets for improving microbial strains for production of biofuels and chemicals.

**Illumina Technology:** Genome Analyzer IIx with 38 bp paired-end reads


Biofuels and Bioremediation

The search for new microbes for the creation of biofuels and bioremediation is usually carried out with metagenomic approaches. However, evolutionary pressure and manipulation in the laboratory can be very effective in improving newly-discovered candidates.

References:


The authors demonstrated that a single base-pair mutation in a transcriptional regulator can have a significant impact on the capacity for substrate utilization and suggest that adaptive evolution should be considered as a potential response of microorganisms to environmental change(s) imposed during bioremediation.

Illumina Technology: Genome Analyzer II


The authors apply experimental evolution followed by genome resequencing and a gene expression study to elucidate genetic bases of adaptation to exogenous isobutanol stress. The evolved lineages exhibit adaptation to isobutanol stress based on remodeling the cell envelope and, surprisingly, stress response attenuation.

Illumina Technology: Genome Analyzer 36 bp single-end and paired-end sequencing with 125x and 500x read depth, respectively.


Viruses

The unique replication strategies of viruses make it particularly challenging to track their evolution in the laboratory. In most living cells misincorporations occur once every billion bases, while in some viruses errors can occur as often as once every thousand bases copied. This results from the use of enzymes without proofreading activity (RdRp or RT) or limited repair activity. Not only do mutations occur more frequently, many copies are made very quickly. A virus may be copied hundreds or even thousands of times in a single life cycle compared to the two progeny cells that result from a single cell cycle.

Virus Detection and Identification

The ambiguous symptoms of viral infections have made them a diagnostic challenge. In addition, their genetic nimbleness has made them a challenge to identify. A recent study by Yozwiak et al. demonstrates the ability of next-generation sequencing to detect and identify viruses that have escaped detection by standard techniques. This capability has generated new interest in searching for viral infections in tumors and chronic diseases. The combination of high sensitivity without the need for prior information will make next-generation sequencing the primary tool for viral detection and identification.

References:


The authors used deep sequencing to detect viral sequences in 37% (45/123) of previously negative cases. These included 13 cases with human herpesvirus 6 sequences. Other samples contained sequences similar to viral sequences found in the Herpesviridae, Flaviviridae, Circoviridae, Anelloviridae, Asfarviridae, and Parvoviridae families. In some cases, the putative viral sequences were virtually identical to known viruses, and in others they diverged, suggesting that they may be derived from novel viruses. By contrast, the Virochip analysis produced putative viral hits in 10/123 (8%) of the previously negative samples. These results demonstrate the utility of unbiased approaches in the detection of known and divergent viruses in the study of tropical febrile illness.

Illumina Technology: Genome Analyzer® and HiSeq® 2000. Total nucleic acid from 140 ml of serum was extracted using the QIAamp Viral RNA Isolation Kit (Qiagen), which co-purifies RNA and DNA.


This study uses next-generation sequencing to investigate viral infection in a variety of different tumor types stored as FFPE samples. The authors are able to detect human papillomavirus subtypes that would not have been detected by traditional methods and show that this approach could be applied to any tumor and any virus.

**Illumina Technology:** Genome Analyzer


To characterize the hepatitis B virus (HBV) genetic heterogeneity in association with anti-viral therapy, the authors perform ultra-deep sequencing of full-genome HBV in the liver and serum of 19 patients with chronic viral infection. They found that clones resistant to anti-viral therapy were common in both the liver and serum of treatment-naive patients, which indicates the putative risk of developing drug resistance.

**Illumina Technology:** Genome Analyzer


Vaccine Production

The manufacture of live viral vaccines requires rigorous quality control to ensure vaccine safety. One of the major risks is that the intrinsic genetic instability of RNA viruses may lead to the accumulation of virulent revertants during manufacture.

The human cytomegalovirus is a good example of the challenges researchers face in the laboratory. Commonly used variants of human cytomegalovirus (HCMV) strains – Towne and AD169 – have been distributed widely and developed as vaccine candidates. Over the years their detailed histories have become obscure and it has become clear that the biological properties of these strains are not conserved between stocks. These genetic differences may affect the interpretation of experimental studies and will obviously significantly impact their use in developing vaccines.³²

References:


In a recent evaluation of MPS platforms at the Center for Biologics Evaluation and Research at the Food and Drug Administration, the authors found that MPS offered significant advantages over standard quality control tool in vaccine production. MPS “… may represent the ultimate tool for monitoring genetic consistency of live viral vaccines.” The currently used mutant analysis by PCR and restriction enzyme cleavage (MAPREC) method measures the frequency of neuroviral mutations at the 5’ untranslated region (UTR) of the viral genome. This region correlates with the level of neurovirulence determined by the monkey neurovirulence test. However, MAPREC can only monitor mutations at a few genomic loci and misses mutations at other sites that could adversely affect vaccine quality. A critical advantage of MPS over MAPREC is that it allows all nucleotide positions in complete viral genomes to be screened in one assay and therefore addresses concerns that mutations at unknown genomic loci could emerge but remain undetected and thus compromise vaccine quality. In this evaluation the MPS results were in perfect agreement with MAPREC results. An unexpected benefit was that the authors were able to distinguish patterns of mutations in the MPS data that were characteristic for the specific seed virus. This would allow the tracking of vaccine lots based on the mutation pattern in the seed.

Illumina Technology: Genome Analyzer


The authors present full genome sequence comparisons of the veterinary herpesvirus, pseudorabies vaccine strain Bartha, and two virulent veterinary herpesvirus, pseudorabies isolates, Kaplan and Becker. These data add to growing evidence that even plaque-purified stocks of stable DNA viruses exhibit limited sequence heterogeneity, which likely seeds future strain evolution.

Illumina Technology: Genome Analyzer 38 bp reads


Yeast

The yeast genome is capable of remarkably complex genetic changes when under environmental pressure. A strain isolated after approximately 188 generations of a sulfate-limited continuous culture of Saccharomyces cerevisiae strain DBY10147 was sequenced by MPS. The authors found both single nucleotide polymorphisms and copy number amplifications that were not found by previous array-based studies.\(^\text{33}\)

Sequencing does not rely on the reference genome, which is an advantage in the many cases where the reference genome may be unavailable or incorrect. For example, when two strains of yeast S288C (12x coverage) and RM11 (15x coverage) were sequenced the authors found 803 and 1104 errors, respectively, in the public sequences.\(^\text{34}\) Any errors and omissions in the reference genomes will be designed into microarrays or tiling arrays that are based on those genomes.

References:


The authors use high-coverage whole-genome sequencing of a conditional mismatch repair mutant line of diploid yeast to identify mutations that accumulated after 160 generations of growth. The vast majority of the mutations accumulated as insertion/deletions (indels) in homopolymeric [poly(dA:dT)] and repetitive DNA tracts. Surprisingly, the likelihood of an indel mutation in a given poly(dA:dT) tract is increased by the presence of nearby poly(dA:dT) tracts in up to a 1,000 bp region centered on the given tract.

**Illumina Technology:** Genome Analyzer 101 bp reads


This paper tracks metabolic changes occurring in the yeast Saccharomyces cerevisiae to increase its specific growth rate on galactose. The study demonstrates that adaptive evolution represents a valuable alternative to rational design in bioengineering and that it is possible to identify mutations in evolved strains that can serve as unforeseen metabolic engineering targets for improving microbial strains for production of biofuels and chemicals.

**Illumina Technology:** Genome Analyzer\(^\text{IIx}\) 38 bp paired-end reads


Well-Studied Microorganisms

Escherichia Coli

It has been observed that isolates from some bacterial infections exhibit within-species diversity. Several lines of evidence suggest that this micro-heterogeneity is due to diversification during the infection process rather than an infection by multiple isolates. The observed diversity resembles results obtained in experimental evolution studies. Whatever the mechanisms leading to diversity, the results emphasize the need for more extensive isolate testing before deciding on antibiotic therapies.

References:


Staphylococcus Aureus

The spread of antibiotic resistance is usually tracked over long periods. For example, DNA sequences of methicillin-resistant Staphylococcus aureus (MRSA) (ST225) serially sampled through time led the authors to estimate that ST225 had diverged since approximately 1990 (1987 to 1994), and that expansion of the European clade began in 1995 (1991 to 1999), several years before the new clone was recognized.35

The spread of antibiotic resistance is also complicated by the potential of horizontal gene transfer. Some clinical MRSA strains are deficient in type III-like restriction endonuclease systems and are therefore hypersusceptible to the horizontal transfer of DNA from other species, such as Escherichia coli. For example, susceptible Staphylococcus aureus strains could easily acquire a vancomycin-resistance gene from enterococci.36

A host jump represents one of the most dramatic examples of genetic adaptation. The majority of S. aureus isolates from broiler chickens are the descendants of a single human-to-poultry host jump that occurred approximately 38 years ago (range, 30 to 63 years ago) by a subtype of the worldwide human ST5 clonal lineage unique to Poland. This represents the evolutionary history of a major new animal pathogen that has undergone rapid avian host adaptation and intercontinental dissemination and is a new paradigm for the study of the impact of human activities on the emergence and spread of animal pathogens.37

References:


**Streptococcus**

Streptococcus is a Gram-positive bacteria belonging to the phylum Firmicutes with a genome size of approximately two million bases.\(^8\)

**References:**


**Mycobacterium**

The fungus-like mycobacteria are responsible for a wide range of diseases, from tuberculosis to leprosy. With a genome size of four million base pairs with 3959 genes it is easily sequenced by next-generation sequencers.

**References:**

  
  An outbreak of Mycobacterium tuberculosis in British Columbia in 2006 demonstrates how sequencing can be used to unravel a complex epidemic. When the results of mycobacterial interspersed repetitive unit–variable-number tandem-repeat (MIRU-VNTR) genotyping and traditional contact tracing failed to identify a source of the M. tuberculosis epidemic, the authors used whole-genome sequencing and social-network analysis to describe the outbreak dynamics at a higher resolution. To do this, they sequenced a total of 36 Mycobacterium tuberculosis isolates (32 of the 37 outbreak isolates and four historical isolates with identical MIRU-VNTR patterns). This yielded an average of 99.21% of the reference genome being covered by at least one 50 bp read. The higher-resolution SNP patterns afforded by whole-genome sequencing revealed that the outbreak was the coalescence of two outbreaks, each with its own causative lineage of Mycobacterium tuberculosis. The simultaneous reappearance of two extant lineages suggests that a social or environmental factor, not a genetic change in the organism, most likely triggered the outbreak. A rise in crack cocaine use within the community, which peaked at the outbreak of the epidemic, may have been this trigger.

- Illumina Technology: Genome Analyzer II


---


The sequencing of Salmonella enterica serovars has demonstrated many of the benefits of genome-wide sequencing\textsuperscript{39,40,41}. It also serves as a cautionary tale of how deceptive current phenotypic characterization, such as serovars and phage types, can be. Based on mutational changes, phage type DT104 is heterogeneous and represented in multiple sequence types. This observation should not be surprising. The serotype is coded by a small part of the genome that is under very different selection pressure than the genes that determine drug resistance and virulence phenotypes. Using serotypes as a surrogate marker for virulence or other characteristic is severely limited. This limitation is significant because the multidrug-resistant variant of DT104 is the cause of epidemics in many parts of the world. Sequencing is a direct observation of the genome and the only definitive method to classify these types.

References:


Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., et al. (2011) Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. J Bacteriol 193: 2332-2335


Reeves, P. R., Liu, B., Zhou, Z., Li, D., Guo, D., et al. (2011) Rates of mutation and host transmission for an Escherichia coli clone over 3 years. PLoS ONE 6: e26907


Microbial Genomics Research Review
An Overview of Publications Featuring Illumina® Technology