A Centre for Infectious Diseases and Microbiology - Public Health (CIDM-PH), and Sydney Emerging Infections and Biosecurity Institute (SEIB) publication

April 2013
Issue 32

The Broad Street Pump

Next Generation Sequencing in the Clinical Microbiology Laboratory

Adrian Ong
Centre for Infectious Diseases and Microbiology - Public Health

Frederick Sanger was awarded his second Nobel Prize for Chemistry in 1980 for the dideoxy method of determining the base sequence of nucleic acids (his first, in 1958, was for determining the amino acid composition of insulin). Sanger sequencing uses the termination of dideoxynucleotides in DNA elongation to determine genomic sequence (Figure 1)(1) and remains the gold standard for determining genomic sequence. In "next-generation sequencing" (NGS), or second generation sequencing, thousands of DNA sequences are generated and read simultaneously. The commonly used platforms are Illumina, Ion Torrent and Roche 454. A third generation of sequencers aims to use single molecules as direct templates for sequencing, removing the bias that occurs from amplification steps in library preparation (2,3). *Continued next page…..*

centre for infectious diseases & microbiology
Public Health

Inside this issue

News & Events

Whole Genome Sequencing in Diagnostic Microbiology Workshop (p11)
Westmead Hospital
Friday, 31 May 2013

Neglected Zoonotic Diseases Workshop (p10)
University of Sydney
Friday, 2 August 2013

SEIB/ SIBRN Colloquium (p10)
University of Sydney
Thursday, 5 December 2013

Next Generation Sequencing

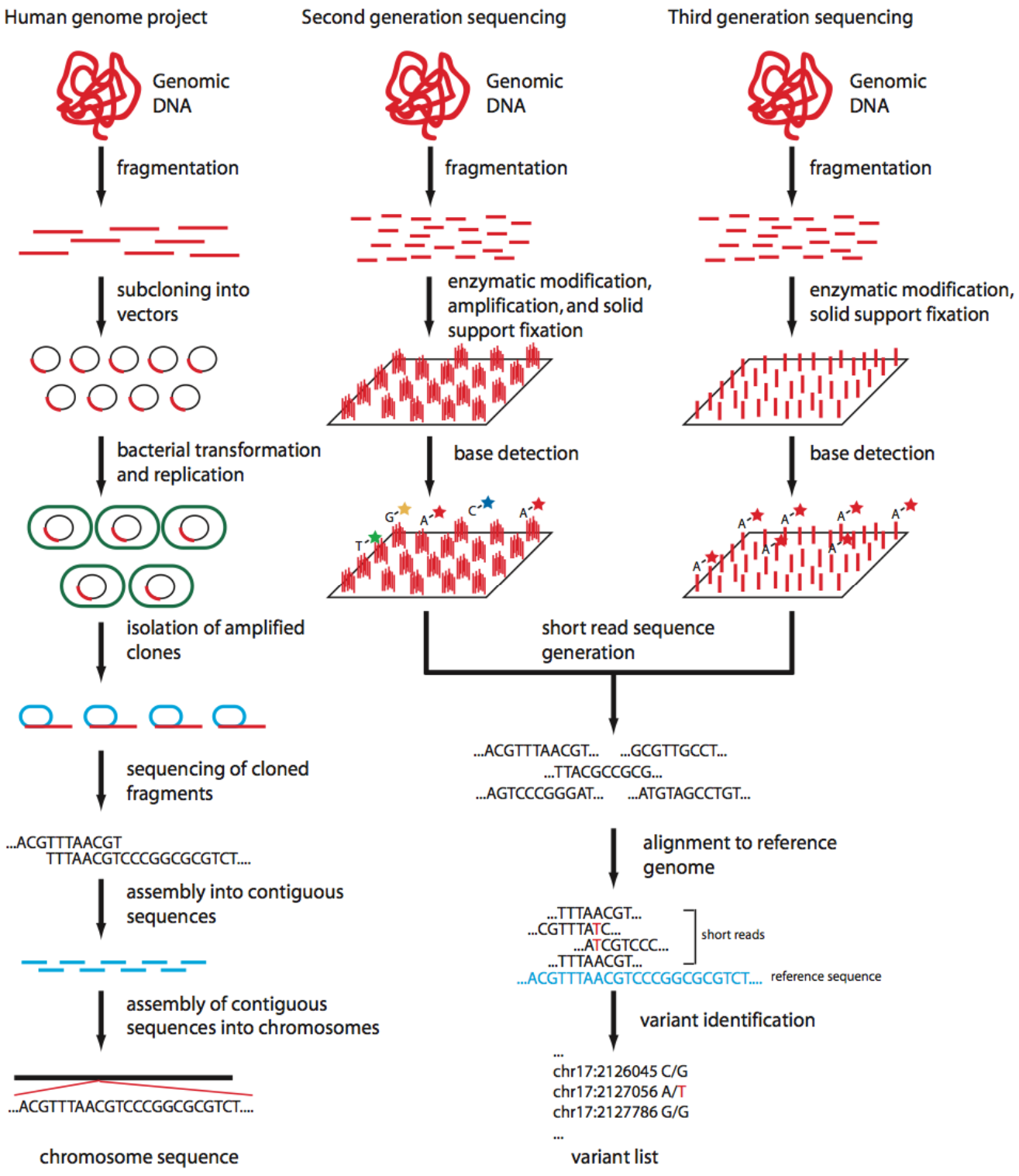


Fig. 1: Comparison of basic principles behind each generation of sequencing (2).

Next Generation Sequencing

Common to all NGS platforms is the ability to generate an enormous amount of data in a short time by sequencing in parallel. The steps involved are similar for each platform, with sample extraction followed by generation of a library, then sequencing, and finally bioinformatics analysis (4).

Library generation involves the random fragmentation and ligation of platform-specific "universal" adaptors on the sample(s). The adaptors allow multiplexing of many samples, a necessary enrichment (amplification) of library fragments, and the attachment of those fragments to the sequencer's solid surface. The immobilization of many uniquely labeled fragments allows the identification of the multiple sequencing reactions that occur(4).

All NGS platforms use the principle of "sequencing-by-synthesis', whereby synthesis of a strand, complementary to the fragment, reveals the base sequence; however the sequencing chemistry is different for each platform. The Illumina platform immobilizes fragments onto a glass slide. Washes of nucleotides, together with DNA polymerase and fluorescence-labeled reversible terminators, generate strands complementary to the immobilized fragments. Each time a reversible labeled terminator is incorporated, the reaction is temporarily halted, photographed, then allowed to repeat, gradually revealing the fragment's sequence (4,5). On the Ion Torrent platform fragments are immobilized onto magnetic beads such that only one molecule is present on each bead. The beads with fragments are deposited into many thousands of wells where one bead sits in each well. Sequencing by synthesis occurs in the wells, with loss of a hydrogen ion each time extension occurs (measured by a pH meter, a semiconductor chip) and converted into sequence data (6). The Roche 454 platform similarly uses library fragments immobilized onto magnetic beads, again one molecule per bead. The beads are deposited onto a picotitre plate containing multiple wells. Each well contains one bead, where sequencing by synthesis using pyrosequencing occurs. In pyrosequencing, incorporation of a nucleotide results in pyrophosphate release, an emission of light proportional to the number of bases incorporated. These discharges of light from multiple wells are photographed and converted to reads (4,5). Each platform has benchtop versions which are cheaper and provide faster turnaround times, thus affording sequencing to individual researchers; these are the Illumina MiSeq, Ion Torrent Personal Genome Machine, and 454 GS Jr (7). Each platform's advantages and disadvantages are outlined in Table 1.

Technology	Advantages	Disadvantages	Error rates	Instrument cost (benchtop sequencers)	Run time (benchtop sequencers)
Capillary / Sanger	Low cost for very small studies	Very high cost for large amounts of data	0.1-1%	USD$376k	2 hours
454	Long read length	High cost per megabase	1%	USD$124k	10 hours
Illumina	Moderate cost instrument and runs	Cost per megabase of MiSeq reads is higher compared to larger Illumina machines	~0.1%	USD$125k	39 hours
Ion Torrent	Low cost instrument upgraded through disposable chips; very simple machine with few moving parts	Higher error rate and needs more hands on time than Illumina	1%	USD$81k	4 – 7 hours

Table 1: Advantages and disadvantages of the common NGS machines (adapted from (7) with updated information from http://www.molecularecologist.com/next-gen-fieldguide-2013/)

Bioinformatics in NGS is the computational analysis of the sheer volume of data generated and can be the most difficult step(8). There are multiple steps, including assessing the quality of the data generated (and filtering the low quality reads), aligning reads to a reference genome and/or assembling the reads, and annotation. There is a plethora of open-source and proprietary tools available (8). Each platform supplies bioinformatics software; however vendor-supplied software may be perceived as inferior, less efficient in alignment, or lacking in features specific to the project or organism sequenced. For example, when vendor-supplied Illumina alignment software was compared to MAQ (Mapping and Alignment with Qualities algorithm, an early mapping tool) MAQ was shown to have several advantages, including taking into account the read quality score during sequence alignment, and assigning a score to quantify the quality of the algorithm's mapping (9).

Data quality control and validation is a barrier to the routine clinical application of NGS (2). Inherent errors of the sequencing technology used and errors introduced during library amplification steps must be filtered using quality score metrics and genome reference data. Alignment (using any of numerous tools available) is often integrated with the ability to identify genetic variants in comparison to the reference, in the form of small or large variants. Small variants include single nucleotide polymorphisms (SNPs) or insertions and deletions (indels); whilst structural variants (SV) are large structural rearrangements >1 kilobase (2).

THE CLINICAL MICROBIOLOGY LABORATORY

NGS in the clinical microbiology laboratory has great potential, with virtual resistance testing based on resistance-associated sequences, high-resolution epidemiological typing and microbiome analysis as possible routine uses (10). The market-leading benchtop Illumina and Ion Torrent platforms generate very similar datasets when sequencing microbial genomes, though the Ion Torrent platform generates severe coverage bias when sequencing the AT-rich *P falciparum* genome, resulting in absent data for 30% of the genome (11). An evaluation of NGS within the various sub-disciplines of microbiology is discussed below, with examples summarized in Table 2.

Apart from the efficient (re)sequencing, the most immediate uses of NGS in bacterial genomes may be diagnostic identification of resistant strains (by detecting resistance genes), and in public health (evaluating the evolutionary history of and identifying specific outbreak strains)(12). Bacterial diversity can be researched through metagenomics, utilising either 16S rRNA targeted sequencing or whole community sequencing (13). NGS of chromatin immunoprecipitation (ChIP), known as ChIP-seq, is a method of studying microbial protein expression and has the potential to identify new serologic tests, vaccines and new therapeutic targets (12).

The small genomes in virology allow for efficient sequencing using barcoding of multiple samples, and also the ability to examine viral quasispecies, especially in highly error-prone RNA viruses. Epidemiologically, the population diversity of viruses can be explored and, clinically, minor resistance variants can be monitored at a higher sensitivity than has been traditionally possible with PCR. Viral metagenomics, using direct nontargeted sequencing of samples can lead to the discovery of novel and emerging viruses in clinical and environmental samples. NGS in virology is also increasingly used in the evaluation of genome expression, or transcriptomics. This assessment of transcribed RNA is known as RNA-seq, and can be performed on viral or host transcriptome (14).

Transcriptomics and whole genome resequencing projects are similarly underway for fungal genomes(15). Fungal genomes are relatively larger and contain repetitive sequences, which can make *de novo* assembly in whole genome sequencing difficult. NGS of the internal transcribed spacer (ITS) region has been suggested to assist with taxonomy and fungal microbiome (16). Complex extraction methods, and the availability of a simpler homokaryon phase, with single nuclei in the fungal asexual phase may complicate sequencing (17).

In parasitology, much work has been in sequencing the *Plasmodium falciparum* genome which is extremely AT-rich, making bacterial clone-based sequencing difficult (18). Genome analysis has been used to identify mutations related to drug resistance, as well as characterization of the *P. falciparum* transcriptome.

NGS is an emerging technology, which holds great potential in the clinical microbiology laboratory. In the era of genomic medicine where one human genome can be sequenced for $1000, routine whole genome sequencing of microorganisms will become commonplace.

Next page (5): Table 2: A selection of next generation sequencing projects already performed

Organism	Tissue / starting material	Library preparation	Sequencing platform	Bioinformatics - assembly	Aim of sequencing	Clinical correlation	Conclusions	Ref
MRSA (Methicillin Resistant *Staphylococcus aureus*)	Frozen isolate stock from clinical samples	Nextera DNA Sample Prep Kit	Illumina MiSeq	SMALT	Whole genome	MRSA outbreak in neonatal ICU	1. Cluster of outbreak Isolates identified 2. Antimicro-bial resistance profile generated	(19)
Escherichia coli O104:H4	Stool sample cultures	As per standard protocols	Ion Torrent Illumina HiSeq PacBio RS	Newbler SOAPdenovo BLASR ALLORA	Whole genome	Nationwide foodborne outbreak (Germany)	Whole genome sequencing and analysis occurred within days	(20, 21)
Mycobacterium tuberculosis	MGIT 960 tubes	Ion Xpress Plus	Ion Torrent 314	Not detailed ? DNAStar	Full-length resistance genes	Genotypic drug resistance in 26 random samples from South Africa	1. Several uncommon amino acid changes discovered 2. Mixed / novel mutations discovered	(22)
HIV	Serum	As per standard protocols	Roche 454	GS Amplicon variant analyzer	p24 (viral capsid) p41 (viral envelope)	HIV super-infection	2 cases of superinfection identified	(23)
Influenza	Lung (autopsy tissue)	Illumina genomic DNA kit	Illumina GA II	EulerSR Velvet	*de novo* assembly of pathogenic organisms	Detection of pathogens responsible for respiratory failure (single patient)	1. Influenza A and *S pnuemo-niae* detected by NGS 2. Quasispe-cies identified	(24)
Cryptococcus gatii	Colony media	As per protocol	Illumina GA II	bwa MUMmer MEGA5 Seqman NGEN	Whole genome	Epidemiologic investigation	1. Superior genomic resolution 2. Unique genotypes within “clonal” subtypes	(25)
Plasmodium falciparum	Leucocyte depleted blood	As per protocol	Illumina GA II	Bwa SNP-o-matic	Whole genome	Analyze diversity	Evolutionary changes can be monitored	(26)
Clinical microbiology laboratory	Loop of colony material from solid agar (~5 colonies)	NexteraXT	Illumina MiSeq	Velvet MiSeq reporter v2.0	*de novo* assembly (whole genome) of organisms	Routine use in the clinical microbiology laboratory	1. Mixed samples characterised 2. *Mycobacte-rium* sp. identi-fied before conventional identification 3. Organisms absent from reference database – no identification possible	(27)

REFERENCES

1. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain -terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 1977 Dec;74(12):5463–7.
2. Dewey FE, Pan S, Wheeler MT, Quake SR, Ashley E a. DNA sequencing: clinical applications of new DNA sequencing technologies. Circulation. 2012 Feb 21;125(7):931–44.
3. McGinn S, Gut IG. DNA sequencing - spanning the generations. New biotechnology. Elsevier B.V.; 2012 Nov 16;00 (00).
4. Metzker ML. Sequencing technologies — the next generation. Nature Reviews Genetics. Nature Publishing Group; 2009;11(1):31–46.
5. Mardis ER. Next-generation DNA sequencing methods. Annual review of genomics and human genetics. 2008 Jan;9:387–402.
6. Rothberg J, Hinz W, Rearick T, Schultz J. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;3–7.
7. Glenn TC. Field guide to next-generation DNA sequencers. Molecular ecology resources. 2011 Sep;11(5):759–69.
8. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. Journal of genetics and genomics. Elsevier Limited and Science Press; 2011 Mar 20;38(3):95–109.
9. Koboldt DC, Ding L, Mardis ER, Wilson RK. Challenges of sequencing human genomes. Briefings in bioinformatics. 2010 Sep;11(5):484–98.
10. Dunne WM, Westblade LF, Ford B. Next-generation and whole-genome sequencing in the diagnostic clinical microbiology laboratory. European journal of clinical microbiology & infectious diseases: official publication of the European Society of Clinical Microbiology. 2012 Aug;31(8):1719–26.
11. Quail M a, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics. BMC Genomics; 2012 Jan;13(1):341.
12. Relman D a. Microbial genomics and infectious diseases. The New England journal of medicine. 2011 Jul 28;365 (4):347–57.
13. Forde BM, O'Toole PW. Next-generation sequencing technologies and their impact on microbial genomics. Briefings in functional genomics. 2013 Jan 11;
14. Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. Application of next-generation sequencing technologies in virology. The Journal of general virology. 2012 Sep;93(Pt 9):1853–68.
15. Chaturvedi V, Nierman WC. Cryptococcus gattii comparative genomics and transcriptomics: a NIH/NIAID White Paper. Mycopathologia. 2012 Jun;173(5-6):367–73.
16. Nilsson RH, Ryberg M, Abarenkov K, Sjökvist E, Kristiansson E. The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. FEMS microbiology letters. 2009 Jul;296(1):97–101.
17. Ma L-J, Fedorova N. A practical guide to fungal genome projects: strategy, technology, cost and completion. Mycology. 2010 Jan;1(1):9–24.
18. Le Roch KG, Chung D-WD, Ponts N. Genomics and integrated systems biology in Plasmodium falciparum: a path to malaria control and eradication. Parasite immunology. 2012;34(2-3):50–60.
19. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. The New England journal of medicine. 2012 Jun 14;366(24):2267–75.
20. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. The New England journal of medicine. 2011 Aug 25;365(8):709–17.
21. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, et al. Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4. The New England journal of medicine. 2011 Aug 25;365(8):718–24.
22. Daum LT, Rodriguez JD, Worthy S a, Ismail N a, Omar S V, Dreyer AW, et al. Next-generation ion torrent sequencing of drug resistance mutations in Mycobacterium tuberculosis strains. Journal of clinical microbiology. 2012 Dec;50 (12):3831–7.
23. Redd AD, Collinson-Streng A, Martens C, Ricklefs S, Mullis CE, Manucci J, et al. Identification of HIV superinfection in seroconcordant couples in Rakai, Uganda, by use of next-generation deep sequencing. Journal of clinical microbiology. 2011 Aug;49(8):2859–67.
24. Kuroda M, Katano H, Nakajima N, Tobiume M, Ainai A, Sekizuka T, et al. Characterization of quasispecies of pandemic 2009 influenza A virus (A/H1N1/2009) by de novo sequencing using a next-generation DNA sequencer. PloS one. 2010 Jan;5(4):e10256.
25. Gillece JD, Schupp JM, Balajee SA, Harris J, Pearson T, Yan Y, et al. Whole genome sequence analysis of Cryptococcus gattii from the Pacific Northwest reveals unexpected diversity. PloS one. 2011 Jan;6(12):e28550.
26. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature. Nature Publishing Group; 2012 Jul 19;487(7407):375–9.
27. Long SW, Williams D, Valson C, Cantu CC, Cernoch P, Musser JM, et al. A genomic day in the life of a clinical microbiology laboratory. Journal of clinical microbiology. 2013 Apr;51 (4):1272–7.

Applications of Whole Genome Sequencing: Tuberculosis

Grant A. Hill-Cawthorne

Evolution of sequencing technologies

Ever since Leroy Hood invented the automated DNA sequencer in 1986 the importance of genomics in research and clinical microbiology has grown. The ABI 373/377 led to a sea change in the Human Genome Project with the tedious and laborious manual method of reading terminated nucleotide sequences from gels being replaced with capillaries and fluorescence. Craig Venter estimated that the later ABI 3700 could sequence the same amount of DNA in a day that was previously being done in a year. Use of these automated sequencing technologies shaved at least 5 years off the length of the Human Genome Project.

However, despite automation there was a lull period where the maximum output that could be produced by any one machine run was around 100 kilo base-pairs (kbp). This changed at the end of 2004 with the development and marketing of the 454 pyrosequencer. Output per run jumped to 10^4kbp per run and the age of "deep sequencing" was born.

Meanwhile Shankar Balasubramanian and David Klenerman at the University of Cambridge had invented a rapid sequencing-by-synthesis technology that generated millions of clusters of short reads and used a reversible chain-termination technique, together with a dye and four-colour imaging, to sequence all of the reads at once. Although the read lengths were initially tiny at 36bp when their company Solexa launched their first machine, the output per run was approximately 10^{12}kbp (*1*). This made the sequencing of whole organisms, including humans, possible within a short space of time. The requirement to fragment DNA and then only sequence 36bp reads required a significant expansion in bioinformatic algorithms and expertise with the big sequencing centres rising to this challenge.

Currently the Illumina HiSeq 2500 can produce 600 gigabases of sequence output per run, which takes 11 days. This translates into 6 billion quality-filter passed reads of 2x100bp in length. You need to produce 330Mb of sequence in order to sequence *Mycobacterium tuberculosis* to a high enough coverage depth for good variant detection sensitivity and specificity. With current commercial kits 96 isolates can be multiplexed onto one lane of an Illumina flowcell – with 16 lanes available on one run. This translates into the ability to sequence 1536 MTb isolates in a single 11 day run. It is certainly a far cry from the construction of large-insert clones (cosmids and bacterial artificial chromosomes) and random small-insert clones required for the first shotgun library of MTb completed only 15 years ago (*2*).

So what can all these output be used for?

Exploring phylogeny and the global distribution of clades

MTb phylogeny was first described by Sreevatsan *et al*. by looking at non-synonymous single nucleotide polymorphisms (SNPs) present in 26 structural genes. They categorised MTb into three principle genetic groups (*3*). Baker *et al*. expanded upon this by sequencing seven housekeeping genes which they could use to group isolates into four distinct lineages (*4*). However, the classification that has remained with us the longest was Sebastian Gagneux's use of large sequence polymorphisms (LSPs) to classify MTb into six lineages: Indo-Oceanic, East Asian, East African Indian and Euro-American, together with the West African 1 and 2 lineages (previously *M. africanum*) (*5*). The phylogeny produced using this method maps nicely onto the phylogenies produced by Baker and Sreevatsan. Filliol *et al*. published a study looking at 159 synonymous SNPs present in 219 isolates of MTb and they identified ten distinct clusters (*6*). These clusters are more distinct than those produced using the six lineages proposed on the basis of LSP data; with the Euro-American lineage (lineage 4) covering six distinct Filliol clusters.

Expanding the number of SNPs studied will allow for greater granularity within these clusters. Typically there are 2000 SNP differences between lineages and 800 SNP differences within a lineage (*7*). This level of resolution enables distinct clades within a lineage to be identified – with individual phylogenetic clades likely to represent expansion of a clone or common ancestor and therefore identify linked cases. Certainly by the end of 2013 enough genomic data will be available to allow for more detailed lineage designations to be made and an MTb-specific genomic nomenclature to be established.

Applications of Whole Genome Sequencing: Tuberculosis

Antimicrobial susceptibilities

Antimicrobial resistance-inducing SNPs are present throughout key genes in the MTb genome, many of which have been previously described. However producing PCR primers to amplify and capillary sequence individual target genes is laborious and is currently mostly done in research. For some anti-tuberculous drugs a handful of SNPs are responsible for the majority of resistance, but in some populations resistance against isoniazid can be caused by a number of different SNPs in genes or gene promoters, and often multiple SNPs for the same drug are present. These aggregations of mutations lead to accumulating levels of drug resistance. Which genes to sequence, which positions to focus on and knowing when to stop looking are all difficult decisions when using capillary sequencing methods. Having the whole genome sequence in hand allows you to quickly identify all genotypic causes of phenotypic drug resistance. It also enables you to identify possible causative mutations when mismatches between phenotypic and genotypic resistances are present.

During clonal expansion and transmission of multidrug resistant forms of MTb, theoretical studies have predicted that relative fitness is a significant factor in the spread of MDR-TB (*8*). The accumulation of ever more drug resistance mutations, often in essential genes, is thought to cause a loss of fitness. Compensatory mechanisms exist in MTb for loss of fitness due to isoniazid, aminoglycoside and rifampicin resistance (*9-11*). Understanding the nature of compensatory evolution in drug resistant MTb strains is important in the prediction of future drug resistant outbreaks and the relative transmission risk of an individual. Knowing where and when compensatory mutations occur also allows us to identify mechanisms with which we may be able to target the evolution of drug resistance in MTb.

The sequencing of multiple isolates of MTb in a single geographic location can allow us to understand and map the acquisition of drug resistance and assess the clonality of isolates. Sequencing of Beijing isolates from the Samara Oblast region of Russia identified two separate clades (Figure 1)(*7*). One clade acquired resistance to isoniazid, rifampicin and streptomycin early and then subsequently developed ethambutol, kanamycin and capreomycin resistance. These multi-resistant isolates then clonally expanded, with pyrazinamide and fluoroquinolone resistance only being acquired late on, and by separate methods for each patient. However, a second clade showed much less clonality with isoniazid and streptomycin resistance again being acquired early but with resistance to ethambutol and the second-line injectables developing multiple independent times within the clade. This pattern is commonly seen in high resistance populations and while traditional molecular epidemiology techniques are relatively stable, the resolution provided by whole genome sequencing (WGS) identifies these sporadic events and is therefore a useful tool for epidemiological tracking.

Outbreak investigation

The earliest and best study to demonstrate the utility of WGS in the investigation of outbreaks was Gardy *et al*. (*12*). An outbreak of 41 cases of TB occurred over a 3-year period in a medium sized community in British Columbia. Traditional techniques (mycobacterial interspersed repetitive units - variable number tandem repeats [MIRU-VNTR] and restriction fragment length polymorphisms) suggested that the outbreak was clonal; with the implication that there was one source. Use of a social-network questionnaire identified one patient (MT0001) as the possible source patient because he was linked to all bar two early cases. However hierarchical clustering of 204 SNPs that differentiated one strain from another identified two distinct lineages; suggesting two simultaneous but coalescing outbreaks. Combining the results from the social-network analysis with the WGS data showed that nine of the transmission events estimated by MIRU were biologically impossible with the divergence occurring at least 5 years prior to MT0001 being infected (Figure 2). MT0001 was indeed a superspreader, causing one of the two outbreaks, but the sources for the second outbreak were likely to be two other patients. Quick and accurate identification of outbreak sources will become increasingly important and is necessary to reduce and stop the spread of disease. We are likely to continue to find that our previous assumptions about outbreaks may have been wrong – with multiple source patients and complicated routes of transmission through communities becoming the norm.

Where to from here?

Whole genome sequencing has now been democratised by the introduction of cheaper and more reliable benchtop sequencers. Library preparation techniques have improved to the point where as little as 10ng of genetic material can be used for sequencing. The need for extended culturing and drug susceptibility testing is likely to be seen as increasingly cost ineffective compared to the greater wealth of information that can be gleaned through sequencing. However, on a note of caution, we do run the risk of not being able to see the wood for the trees and we need to ensure that we have a question to be answered through sequencing and not simply going on a fishing trip. However, with this in mind clinical microbiology will be undergoing a significant shift during this decade and the onus is on individual laboratories to acquire the technology and expertise to meet the new era head on.

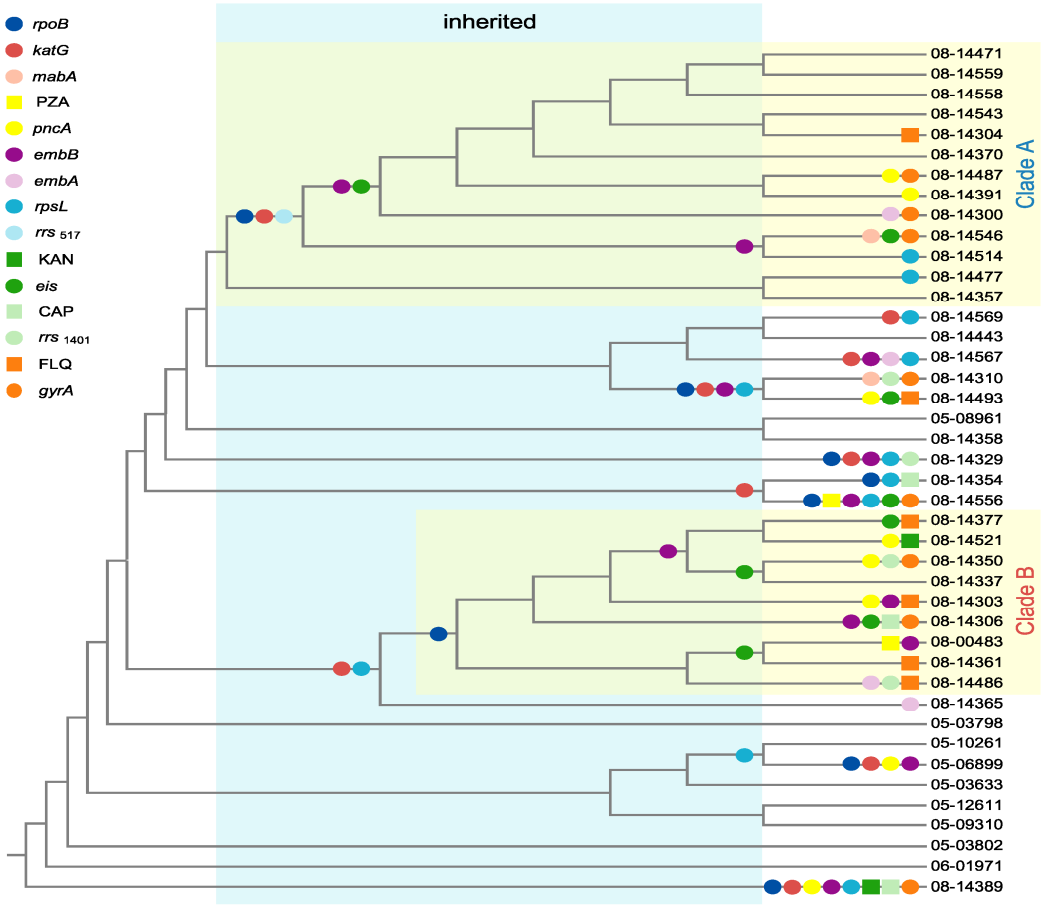


Figure 1. ***Model of acquisition of drug resistance in*** *Beijing isolates. SNPs causative for drug resistance are shown as circles, squares denote drug resistance where no causative SNP was identified. PZA=pyrazinamide, KAN=kanamycin, CAP=capreomycin and FLQ=fluoroquinolone. From* ***N. Casali*** *et al., Genome Research 22, 735–745 (2012).*

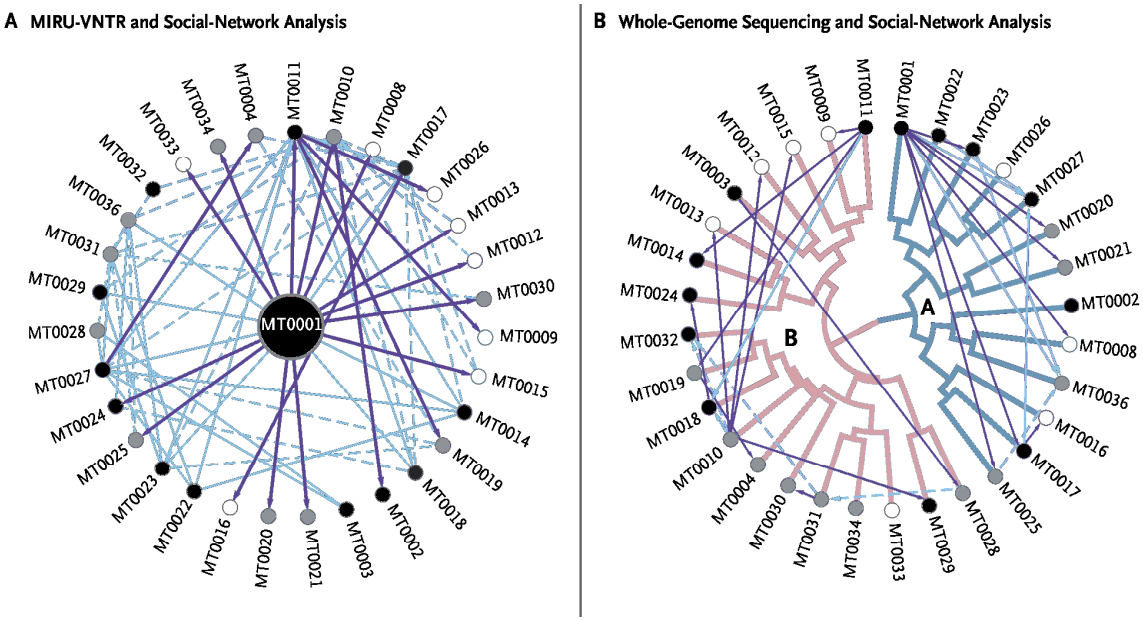


Figure 2. Putative transmission networks based on MIRU-VNTR versus whole-genome sequencing data. Genotyping data based on MIRU-VNTR is shown in panel A; combination with the social-network analysis suggests that MT0001 is the source patient. Application of whole-genome sequencing (panel B) identifies two separate clades, making the conclusion from panel A biologically impossible. Instead clade A is likely to be from the source patient MT0001, whilst clade B is derived from two separate sources MT0010 and MT0011. From ***J. L. Gardy*** *et al., N Engl J Med 364, 730–739 (2011).*

Applications of Whole Genome Sequencing: Tuberculosis

REFERENCES

1. E. R. Mardis, *Nature* **470**, 198–203 (2011).
2. S. T. Cole *et al.*, *Nature* **393**, 537–544 (1998).
3. S. Sreevatsan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9869–9874 (1997).
4. L. Baker, T. Brown, M. C. Maiden, F. Drobniewski, *Emerg Infect Dis* **10**, 1568–1577 (2004).
5. S. Gagneux, P. M. Small, *The Lancet Infectious Diseases* **7**, 328–337 (2007).
6. I. Filliol *et al.*, *J. Bacteriol.* **188**, 759–772 (2006).
7. N. Casali *et al.*, *Genome Research* **22**, 735–745 (2012).
8. T. Cohen, M. Murray, *Nat Med* **10**, 1117–1121 (2004).
9. I. Comas *et al.*, *Nat Genet* (2011), doi:10.1038/ng.1038.
10. D. R. Sherman *et al.*, *Science* **272**, 1641–1643 (1996).
11. D. Shcherbakov *et al.*, *Molecular Microbiology* (2010), doi:10.1111/j.1365-2958.2010.07218.x.
12. J. L. Gardy *et al.*, *N Engl J Med* **364**, 730–739 (2011).

EVENTS

Whole Genome Sequencing in Diagnostic Microbiology Workshop
Westmead Hospital
Friday, 31 May 2013

Visit the CIDM-Public Health or SEIB website for a full copy of the WGS program and registration form. Registrations close 29 May 2013.

Neglected Zoonotic Diseases Workshop
University of Sydney
Friday, 2 August 2013
More information: www.sydney.edu.au/seib/

SEIB/ SIBRN Colloquium
New Law School, University of Sydney
Thursday, 5 December 2013
More information: www.sydney.edu.au/seib/

CONTACT US

For more information on any articles or CIDM-PH & SEIB events, or to join the e-lists and receive regular updates, please contact us at:

CIDM-PH

Centre for Infectious Diseases and Microbiology - Public Health
Mailing Address: PO Box 533, Wentworthville NSW 2145
Phone: (612) 9845 6255
Website: www.cidmpublichealth.org
Email: lou.orszulak@swahs.health.nsw.gov.au

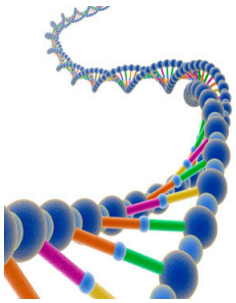
Contact: Lou Orszulak

SEIB

Sydney Emerging Infections and Biosecurity Institute
Mailing Address: ICPMR Bldg Rm 3114, Westmead Hospital, Westmead, NSW 2145
Phone: (612) 9845 9864
Website: www.sydney.edu.au/seib/
Email: seib@sydney.edu.au

Contact: Christine Aitken

WHOLE GENOME SEQUENCING IN CLINICAL AND PUBLIC HEALTH MICROBIOLOGY



Friday 31st May 2013

CIDM-PH & SEIB Workshop

Loewenthal Auditorium,
Westmead Hospital, Sydney

Invited Speakers:

Professor Eddie Holmes, University of Sydney and Fogarty International Center, National Institute of Health, USA

Dr Grant Hill-Cawthorne, SEIB and the University of Cambridge, UK

Dr Tanya Golubchik, Departments of Medicine and Statistics, University of Oxford, UK

A/Prof Ruiting Lan, University of New South Wales

Dr Sebastian van Hal, Royal Prince Alfred Hospital, Sydney

CIDM
centre for infectious diseases & microbiology
Public Health

The aim of this workshop is to give microbiologists an overview of the applications of next generation sequencing (NGS) in diagnosis and surveillance of infectious diseases. The emphasis will be on methods of analysis of bacterial and viral genome sequencing data generated by NGS platforms, principally by benchtop Illumina and Ion Torrent, and on the utility of NGS for infection control and public health.

A Centre for Infectious Diseases & Microbiology – Public Health (CIDM-PH) and Sydney Emerging Infections and Biosecurity Institute (SEIB), University of Sydney, Workshop

Registration & Program available at:

www.cidmpublichealth.org
www.sydney.edu.au/seib

Workshop Enquiries

Ph: (02) 9845 9870

lou.orszulak@swahs.health.nsw.gov.au