



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

Current Opinion in  
Microbiology

# High-throughput sequencing for the study of bacterial pathogen biology

Paul R McAdam, Emily J Richardson and J Ross Fitzgerald



A revolution in sequencing technologies in recent years has led to dramatically increased throughput and reduced cost of bacterial genome sequencing. An increasing number of applications of the new technologies are providing broad insights into bacterial evolution, epidemiology, and pathogenesis. For example, the capacity to sequence large numbers of bacterial isolates is enabling high resolution phylogenetic analyses of bacterial populations leading to greatly enhanced understanding of the emergence, adaptation, and transmission of pathogenic clones. In addition, RNA-seq offers improved quantification and resolution for transcriptomic analysis, and the combination of high-throughput sequencing with transposon mutagenesis is a powerful approach for the identification of bacterial determinants required for survival *in vivo*. In this concise review we provide selected examples of how high throughput sequencing is being applied to understand the biology of bacterial pathogens, and discuss future technological advances likely to have a profound impact on the field.

## Addresses

The Roslin Institute and Edinburgh Infectious Diseases, University of Edinburgh, Easter Bush Campus, Edinburgh EH25 9RG, United Kingdom

Corresponding author: Fitzgerald, J Ross ([Ross.Fitzgerald@ed.ac.uk](mailto:Ross.Fitzgerald@ed.ac.uk))

**Current Opinion in Microbiology** 2014, **19**:106–113

This review comes from a themed issue on **Novel technologies in microbiology**

Edited by **Emmanuelle Charpentier** and **Luciano Marraffini**

For a complete overview see the [Issue](#) and the [Editorial](#)

Available online 14th July 2014

<http://dx.doi.org/10.1016/j.mib.2014.06.002>

1369-5274/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

## Introduction

The development of new technologies enabling rapid, inexpensive, and high-throughput DNA sequencing (HTS) that offer clear advantages over traditional Sanger sequencing has revolutionized the field of bacterial genomics [1\*,2]. Furthermore, the recent development of high-throughput ‘benchtop’ sequencers is empowering laboratories to sequence their bacteria of interest independently of specialist sequencing centres [2,3]. An array of different technologies have been developed with the common feature that they parallelize the sequencing

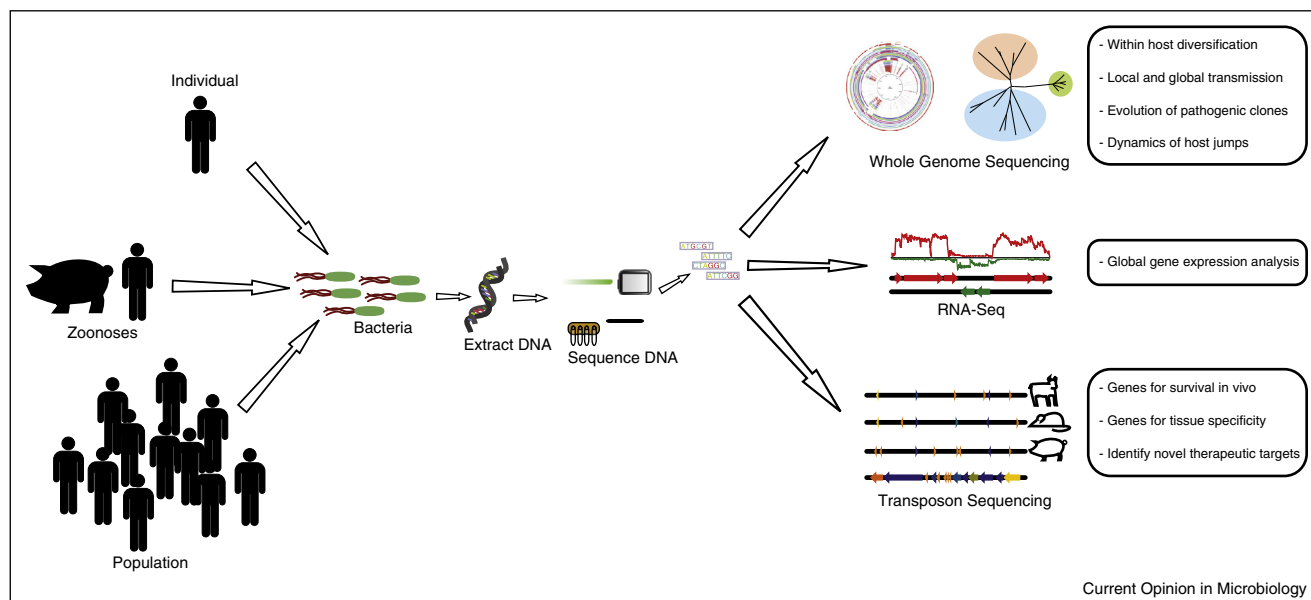
process, leading to the production of thousands or millions of sequence reads concurrently (for a review of the HTS technologies see [1\*,2]). HTS is being applied in a myriad of ways to address fundamental questions concerning the biology of infectious diseases. The high resolution offered by HTS allows the inference of transmission pathways during global pandemics and localized outbreaks, identification of molecular mechanisms underpinning the emergence of pathogenic clones, and the evolutionary analysis of bacterial populations during infection of individual patients. HTS also provides the potential for transcriptomic analyses with advantages over traditional hybridization approaches, including genome-wide coverage, accurate quantification, and single nucleotide resolution (for recent comprehensive reviews, see [4,5\*]). In addition, the combination of HTS with transposon mutagenesis leading to the development of approaches such as Tn-seq [6], transposon-directed insertion site sequencing (TraDIS) [7\*], insertion sequencing (INseq) [8], and high-throughput insertion tracking by deep sequencing (HITS) [9] has facilitated the screening of libraries of hundreds of thousands of bacterial mutants to identify determinants required for survival during growth *in vivo* or in other specific growth conditions (for recent comprehensive reviews see [10,11]).

In the current concise review, we will summarize selected recent studies that have applied HTS to answer important questions regarding the success of major bacterial pathogens. We provide an overview of some of the insights which can be derived from the application of these techniques.

## Study of bacterial evolution during infection

The progression and outcome of infectious disease is determined by the dynamics of host–pathogen interactions, and recent studies employing HTS have offered novel insights into the evolution of bacterial pathogens during the course of colonization and infection [12\*,13,14\*\*,15]. For example, an emerging theme in infectious disease research is the extent of genetic and phenotypic diversification that may occur among the infecting bacterial population within an individual host. In particular, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Mycobacterium abscessus*, *Mycobacterium tuberculosis*, and *Burkholderia dolosa* have been demonstrated to undergo considerable diversification during infection, resulting in ‘clouds of diversity’ that originated from a single or closely related group of infecting bacteria [12\*,14\*\*,15,16,17\*,18]. During infection, random advantageous mutations may

Figure 1



Schematic diagram summarizing the applications of high throughput sequencing for studies of the epidemiology, evolution and pathogenesis of bacterial infections.

become fixed within sub-populations, due to selective pressures such as co-infection with other microorganisms, the host immune response, and antimicrobial chemotherapy [12\*,14\*\*,15,17\*,18,19]. Of note, cystic fibrosis (CF) patients are at particular risk of pulmonary infections, and a number of studies have utilized HTS to examine the genetic diversification of bacterial populations during long-term infection of CF patients [12\*,15,18,20]. For example, convergent evolution represented by independent mutations affecting O-antigen switching has been identified in chronic *B. dolosa* infections [15]. Furthermore, mutations influencing the smooth to rough morphotype transition of *M. abscessus* spp. during infection of the CF lung have been identified [20,21], and distinct polymorphisms of loci influencing the alternative sigma factor (SigB) of *S. aureus* were identified in multiple sub-lineages of *S. aureus* within a single CF patient [18]. The study also revealed mutations underlying antibiotic resistance which occurred during infection as demonstrated in other studies of *S. aureus* chronic infection [22,23]. Loss of virulence factor production has also been described during therapeutic *Escherichia coli* colonization of human patients with recurrent urinary tract infections [24\*\*]. Numerous mutations associated with reduced virulence and adaptation to oxidative stresses, and the recurrence of mutations within individual patients strongly suggested that host-specific selective pressures influence microevolution during infection [24\*\*]. In addition, a long term *S. aureus* carriage study identified an overall pattern of purifying selection within asymptotically colonized hosts [13]. An enrichment of premature stop codons

was observed in invasive bloodstream isolates when compared to carriage isolates from an individual who developed a fatal bacteraemia, implying that specific genetic changes in carriage isolates may be functionally important in pathogenesis [14\*\*].

Differences in the composition of resident bacterial populations between healthy and disease states are increasingly being described, and deep sequencing metagenomic methods can capture greater diversity from the microbiota in comparison to traditional methods relying on PCR amplification and Sanger sequencing [25]. For example, decreased microbial diversity in CF patients in comparison to healthy controls is associated with more severe inflammation [26] and distinct shifts in metabolic pathways have been identified [27]. Additionally, the effects of antimicrobial therapies on the gut microbiota have been investigated revealing increased phage mobilization [28,29], and profound shifts in composition that persist after the cessation of therapy [30] (Figure 1).

### Identification of bacterial determinants required for pathogenesis

The application of HTS for measuring the bacterial transcriptome (RNA-seq) in different environmental conditions has considerable advantages over traditional hybridization-based techniques. The method requires isolation of RNA followed by reverse transcription to cDNA to allow library construction before HTS. Specifically, the inclusion of strain-specific genetic material,

single base-pair resolution, and more accurate quantification of relative levels of gene expression are improvements on previous approaches using microarrays. Although RNA-seq technology has been relatively slow to be employed for bacterial transcriptomic studies due in part to technical challenges of contaminating rRNA, an increasing number of studies are now being published. For example, comparative gene expression analysis of the opportunistic human pathogen *Aggregatibacter actinomycetemcomitans* *in vivo* and in a biofilm model *in vitro* revealed differential expression of 14% of the transcriptome, providing new information relevant to the metabolic pathways important for infection [31]. Other selected studies have shown differential expression of master regulatory genes in *Streptococcus pneumoniae* [6], genes essential for the survival of *Haemophilus influenzae* outside of the host [32], and survival of *Salmonella* Typhi in bile [7]. In addition, RNA-seq identified the extent of the *S. Typhi ompR* regulon [33] and revealed that the *E. coli* plasmid pO157\_sal regulates the expression of chromosomal genes associated with the stress response, antibiotic resistance, and virulence [34].

A particularly powerful application is the combination of HTS with transposon mutagenesis. Briefly, libraries made up of bacteria each with single transposon insertions into non-essential genes are used as an inoculum for experimental infections or for culture in defined conditions. Subsequently, input and output populations are subjected to HTS allowing the relative quantification of mutants in each population. Accordingly, the complement of genes required for survival can be identified. For example, the application of transposon-directed insertion-site (TraDIS) sequencing to *Salmonella enterica* serovars *S. Typhi* and *S. Typhimurium* revealed a conserved core of 281 genes required for growth in both serovars [35]. Attenuation of O-antigen genes and ferric [Fe(III)] genes was observed in *S. Typhimurium* and *S. Typhi* respectively [35], which may partly reflect the distinct host tropisms of each serovar. A Tn-seq study of adaptation of *S. pneumoniae* to the nasopharynx and the lung indicated an array of genes associated with response to hydrogen peroxide which were important during lung infection, while genes associated with response to temperature and sucrose were involved in survival in the nasopharynx [36]. In contrast, a study of the interactions of *Moraxella catarrhalis* with the host respiratory epithelia reported no difference in expression of adhesion factors between adhesive and planktonic states [37].

Finally, in a novel application of Tn-seq, the dynamics of *H. influenzae* and influenza A virus co-infection were examined in a mouse model of respiratory infection. Of note, *H. influenzae* genes associated with oxidative stress were required during co-infection with influenza A compared to *H. influenzae* infection alone [38]. The use of HTS-transposon approaches is revealing much about the

mechanisms of bacterial survival *in vivo* and in doing so may indicate novel therapeutic targets for controlling infection.

### Identifying the source and transmission routes of disease outbreaks

Traditional bacterial typing methods are often of limited use for epidemiological investigations of infectious disease outbreaks due to their low level of discriminatory power [39]. The comparison of whole genome sequences can provide the ultimate level of nucleotide resolution between bacterial isolates and accordingly has the potential to identify transmission events within and between hospitals and in the community [3,17,40,41,42,43,44,45–47]. Importantly, the application of rapid sequencing of suspected outbreak strains can inform both clinical practice and infection control procedures [17]. The level of genetic variation observed among strains epidemiologically linked to an outbreak can help determine whether a point source or multiple strains are responsible [3,40,44,48]. For example, based on multiple locus tandem repeats typing analysis, a clonal population of *M. tuberculosis* was implicated as the cause of a sustained outbreak of tuberculosis in Canada. However, in a seminal study by Gardy *et al.*, integration of genome sequence information and social network analysis led to the discovery that the outbreak was caused by two circulating lineages sustained by crack cocaine users [44]. In addition, HTS has been applied retrospectively to analyse *S. aureus* epidemics in hospitals revealing the dynamics of outbreaks confined to single hospital wards [3,40], and transmission between the hospital and community settings [17]. The directionality of transmission, although often ambiguous, may sometimes be inferred by combining epidemiological data and analysis of bacterial population genome sequences [12,15]. For example, in a study of *Burkholderia dolosa* associated with chronic infections in a CF patient cohort, transmission events could be inferred from the phylogeny based on patterns of shared polymorphisms [15]. Also, a study of patients with *M. abscessus* subspecies *massiliense* pulmonary infections showed greater genetic diversity among isolates from a single patient than from different patients, suggestive of inter-patient transmission events [12]. Recently, the development of novel bioinformatic algorithms has facilitated the identification of transmission events while accounting for the heterogeneity present in infecting bacterial populations [49].

### Understanding the molecular basis for the emergence of pathogenic clones

The capacity to sequence large numbers of closely related isolates allows high-resolution phylogenies to be reconstructed, which may provide insights into the processes underlying the emergence and spread of pathogenic clones. Bacterial populations accumulate

random mutations over time through inherent mutation rates specific for the organism and its ecology. Estimates of the mutation rate for a given bacterial population allow the construction of time-calibrated phylogenetic trees [50]. In particular, the application of Bayesian phylogenetic methods to sequences of closely related bacteria, can allow a time-scaled reconstruction of their evolutionary history and geographic dissemination, and may also reveal genetic events that correlate with the emergence of successful clones. The evolutionary history of an increasing array of bacterial pathogens has been examined using this approach including *S. aureus*, *Shigella sonnei*, *S. Typhimurium* and *Vibrio cholerae* [42,43\*,51–56]. In a concise review of this nature we can provide just a few recent examples. Methicillin-resistant *S. aureus* (MRSA) is an important cause of nosocomial infection in the United Kingdom, and recent studies have described genetic events preceding the emergence of the major hospital-adapted lineages EMRSA-15 and EMRSA-16, including polymorphisms associated with increased antimicrobial resistance and reduced virulence [42,43\*]. The reduction in virulence likely represents a fitness compensation in response to the increased energy costs associated with antibiotic resistance. In addition, a recent HTS study of invasive *S. Typhimurium* circulating in sub-Saharan Africa revealed the existence of two distinct epidemics that correlated strongly with peaks in HIV incidence in the region, suggesting that the increased number of immunocompromised hosts in sub-Saharan Africa has been a major contributor to the success of invasive *S. Typhimurium* [53]. Similarly, a study of *S. Typhimurium* DT104 from livestock and human hosts employed Bayesian phylogenetic methods with extended analysis of host-tropism using Markov jump methods to demonstrate that circulating DT104 strains represent independent epidemics. The data indicated that there had been limited cross-species transmission, contradicting a long-standing assumption regarding the zoonotic spread of *S. Typhimurium* [54].

In another HTS-based evolutionary study, it was demonstrated that the 7th cholera pandemic comprised of three independent waves, with the latter two being driven through acquisition of the SXT antibiotic resistance element by *V. cholerae*. This resulted in resistance to the commonly used anti-cholera therapies [52]. In the aftermath of the 2010 earthquake in Haiti, the country suffered from a prolonged outbreak of cholera. Phylogenetic reconstruction based on whole genome sequences of *V. cholerae* from affected patients identified several intercontinental transmission events originating from a source population in the Bay of Bengal [52]. Further analysis revealed that Haitian epidemic strains are related to strains circulating in Nepal, implicating movements of Nepali peacekeepers in the emergence of the epidemic [55] following a single introduction to Haiti [56].

The impact of human clinical interventions on the emergence of successful bacterial clones has been demonstrated for the *S. pneumoniae* PMEN1 lineage [57,58]. Following introduction of a conjugate polysaccharide vaccine, capsule switching was observed [57] leading to the emergence of the vaccine escape serotype 19A. It was revealed that the geographic distribution of serotype 19A has rapidly expanded to replace vaccine-susceptible serotypes [58]. Further work based on HTS of 616 *S. pneumoniae* isolates identified variation in recombination rates between branches of the species' phylogeny and also revealed an association between specific genetic determinants of *S. pneumoniae* and the age of the host. These data are consistent with adaptation of *S. pneumoniae* in response to the maturing host immune response [59\*\*].

HTS also played a central role in the near real-time characterization of a high mortality outbreak of haemolytic-uraemic syndrome (HUS) in central Europe during 2011 [60–62], rapidly revealing that the increased virulence of the O104:H4 outbreak was likely due to acquisition of a Shiga-toxin encoding prophage [62], and that the strain was refractory to antibiotic therapy due to carriage of an extended spectrum  $\beta$ -lactamase [60].

### The molecular basis of bacterial host switches

The high resolution phylogenetic analysis possible using HTS has enhanced our understanding of the capacity of bacterial pathogens to switch host species and adapt to survive and spread among novel host populations [51,63–66]. In particular, several studies have identified livestock as reservoirs for emerging bacterial strains capable of causing disease in humans [51,63,65]. For example, Price *et al.* examined the multi-host association of the ST398 clone of *S. aureus* by HTS providing evidence for the emergence of antibiotic resistant strains resulting from the use of antibiotics in the livestock industry [63]. Similarly, Spoor *et al.* demonstrated that cows are a potential reservoir of new strains of *S. aureus* with the capacity for pandemic spread in humans [51].

In addition to understanding the dynamics of cross-species transmission, HTS allows high-resolution analysis of genetic correlates of host specificity. For example, in one of the few genome-wide association studies (GWAS) of bacteria to be carried out, Shepphard *et al.* identified a genomic region encoding vitamin B<sub>5</sub> biosynthesis components which is associated with adaptation of *Campylobacter jejuni* to the bovine host [67\*]. Also, a *S. Typhimurium* ST4/74 transposon mutant pool was used to infect avian, bovine, and porcine hosts revealing host-specific gene repertoires, including genes associated with anaerobic growth in the avian host [68\*\*]. Finally, HTS of isolates of the host-restricted DT2 lineage of *S. Typhimurium* and closely related strains suggested that adaptation of the DT2 lineage to the rock pigeon was

not mediated by acquisition of novel genetic material, but rather by polymorphisms in the pre-existing genetic repertoire [69].

### Future applications of HTS for understanding the biology of bacterial pathogens

The pace of development of sequencing technologies shows no sign of slowing, and platforms capable of single-molecule sequencing and ever increasing read lengths offer the possibility of highly accurate assemblies of individual pathogens within a microbial community [70]. While bacterial culture has previously been an essential step for the isolation of enough genomic DNA for whole genome sequencing, novel culture-free methodologies offer the ability to sequence un-culturable organisms. Such applications have relevance for investigating the cause of infectious diseases of unknown aetiology. Furthermore, the rapid diagnosis and *in silico* determination of sensitivity profiles of pathogens without the necessity for culture have obvious benefits for the treatment of clinical infections [71,72]. Furthermore, transcriptomic analysis of complex populations of bacteria within the microbiota will be theoretically feasible. To date, transcriptomic studies have focused on either the pathogen or the host. However, a comprehensive understanding of host–pathogen interactions would require simultaneous analysis of gene expression of both parties during infection. Dual RNA-seq offers the potential for transcriptomic analysis of both host and pathogen during the course of colonization and infection but there are technical difficulties to overcome before routine implementation of this technology is feasible, including removal of both bacterial and host rRNA, and large scale differences in relative amounts of RNA for host and pathogen [73]. Finally, the availability of genome sequences for large numbers of well-defined clinical isolates lends itself to the application of GWAS to bacteria, a method originally developed for human genetic association analysis which has the potential for unbiased identification of genetic determinants associated with a given phenotype [74]. For example molecular correlates of virulence or host-specificity may be determined through simultaneous analysis of genome data and the results from virulence assays [67,75,76].

Overall the development of HTS technologies has revolutionized how we approach fundamental research into infectious diseases. Without doubt the new approaches will result in a very enhanced understanding of the biology of bacterial pathogens which will ultimately lead to improved infection control.

### Acknowledgements

We are grateful to current and past members of the LBEP lab for their contributions to research on the evolutionary genomics of bacterial pathogens, and to the Biotechnology and Biological Sciences Research Council (UK) for ongoing funding.

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
  - of outstanding interest
1. Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ: **High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity.** *Nat Rev Microbiol* 2012, **10**:599-606.
  - Comparison of high throughput sequencing platforms and their performance.
  2. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ: **Performance comparison of benchtop high-throughput sequencing platforms.** *Nat Biotechnol* 2012, **30**:434-439.
  3. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CLC, Wilson DJ, Didelot X, O'Connor L *et al.*: **A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance.** *BMJ Open* 2012, **2**.
  4. Filiatrault MJ: **Progress in prokaryotic transcriptomics.** *Curr Opin Microbiol* 2011, **14**:579-586.
  5. Croucher NJ, Thomson NR: **Studying bacterial transcriptomes • using RNA-seq.** *Curr Opin Microbiol* 2010, **13**:619-624.
  - A good review of RNA-seq technology.
  6. Van Opijnen T, Bodi KL, Camilli A: **Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms.** *Nat Methods* 2009, **6**:767-772.
  7. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G *et al.*: **Simultaneous assay of every *Salmonella Typhi* gene using one million transposon mutants.** *Genome Res* 2009, **19**:2308-2316.
  - One of the first examples of transposon mutagenesis on a high-throughput scale.
  8. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI: **Identifying genetic determinants needed to establish a human gut symbiont in its habitat.** *Cell Host Microbe* 2009, **6**:279-289.
  9. Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ: **Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung.** *Proc Natl Acad Sci U S A* 2009, **106**:16422-16427.
  10. Van Opijnen T, Camilli A: **Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms.** *Nat Rev Microbiol* 2013, **11**.
  11. Barquist L, Boinett CJ, Cain AK: **Approaches to querying bacterial genomes with transposon-insertion sequencing.** *RNA Biol* 2013, **10**:1161-1169.
  12. Bryant JM, Grogono DM, Greaves D, Foweraker J, Roddick I, Inns T, Reacher M, Haworth CS, Curran MD, Harris SR *et al.*: **Whole-genome sequencing to identify transmission of *Mycobacterium abscessus* between patients with cystic fibrosis: a retrospective cohort study.** *Lancet* 2013, **381**:1551-1560.
  - Reveals how an understanding of intra-host and inter-host bacterial diversity can be used to infer transmission events.
  13. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Lerner-Svensson H, Fung R, Godwin H, Knox K, Votintseva A *et al.*: **Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage.** *PLoS ONE* 2013, **8**:e61319.
  14. Young BC, Golubchik T, Batty EM, Fung R, Lerner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG *et al.*: **Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease.** *Proc Natl Acad Sci U S A* 2012 <http://dx.doi.org/10.1073/pnas.1113219109>.
  - Seminal paper highlighting the level of genetic diversity in colonising *S. aureus* isolates, and genetic changes associated with disease progression.

15. Lieberman TD, Michel J-B, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB *et al.*: **Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes.** *Nat Genet* 2011, **43**:1275-1280.
16. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M *et al.*: **Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection.** *Nat Genet* 2011, **43**:482-486.
17. Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J *et al.*: **Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study.** *Lancet Infect Dis* 2013, **13**:130-136.
- A good example of how sequencing of bacterial genomes during a hospital outbreak can be used to inform infection control strategies.
18. McAdam PR, Holmes A, Templeton KE, Fitzgerald JR: **Adaptive evolution of *Staphylococcus aureus* during chronic endobronchial infection of a cystic fibrosis patient.** *PLoS ONE* 2011, **6**:e24301.
19. Lieberman TD, Flett KB, Yelin I, Martin TR, McAdam AJ, Priebe GP, Kishony R: **Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures.** *Nat Genet* 2013, **46**(1):82-87.
20. Kreuzfeldt KM, McAdam PR, Claxton P, Holmes A, Seagar AL, Laurenson IF, Fitzgerald JR: **Molecular longitudinal tracking of *Mycobacterium abscessus* spp. during chronic infection of the human lung.** *PLoS ONE* 2013, **8**:e63237.
21. Pawlik A, Garnier G, Orgeur M, Tong P, Lohan A, Le Chevalier F, Sapriel G, Roux A-L, Conlon K, Honoré N *et al.*: **Identification and characterization of the genetic changes responsible for the characteristic smooth-to-rough morphotype alterations of clinically persistent *Mycobacterium abscessus*.** *Mol Microbiol* 2013 <http://dx.doi.org/10.1111/mmi.12387>.
22. Mwangi MM, Wu SW, Zhou Y, Sieradzki K, de Lencastre H, Richardson P, Bruce D, Rubin E, Myers E, Siggia ED *et al.*: **Tracking the *in vivo* evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing.** *Proc Natl Acad Sci U S A* 2007, **104**:9451-9456.
23. Howden BP, McEvoy CR, Allen DL, Chua K, Gao W, Harrison PF, Bell J, Coombs G, Bennett-Wood V, Porter JL *et al.*: **Evolution of multidrug resistance during *Staphylococcus aureus* infection involves mutation of the essential two component regulator WalKR.** *PLoS Pathog* 2011, **7**:e1002359.
24. Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D, Voigt B, Grönberg-Hernandez J, Ragnarsdottir B, Hecker M, Ron EZ *et al.*: **Host imprints on bacterial genomes – rapid, divergent evolution in individual patients.** *PLoS Pathog* 2010, **6**:e1001078.
- A unique study demonstrating human host-specific adaptations acquired by an *E. coli* strain used for therapeutic management of recurrent urinary tract infections.
25. Salipante SJ, Sengupta DJ, Rosenthal C, Costa G, Spangler J, Sims EH, Jacobs MA, Miller SI, Hoogestraat DR, Cookson BT *et al.*: **Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections.** *PLoS ONE* 2013, **8**:e65226.
26. Blainey PC, Milla CE, Cornfield DN, Quake SR: **Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis.** *Sci Transl Med* 2012, **4**:153ra130.
27. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S, Nosrat B, Conrad D, Rohwer F: **Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals.** *PLoS ONE* 2009, **4**:e7370.
28. Modi SR, Lee HH, Spina CS, Collins JJ: **Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome.** *Nature* 2013, **499**:219-222.
29. Fancello L, Desnues C, Raoult D, Rolain JM: **Bacteriophages and diffusion of genes encoding antimicrobial resistance in cystic fibrosis sputum microbiota.** *J Antimicrob Chemother* 2011, **66**:2448-2454.
30. Dethlefsen L, Relman DA: **Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.** *Proc Natl Acad Sci U S A* 2011, **108**(Suppl):4554-4561.
31. Jorth P, Trivedi U, Rumbaugh K, Whiteley M: **Probing bacterial metabolism during infection using high-resolution transcriptomics.** *J Bacteriol* 2013, **195**:4991-4998.
32. Langereis JD, Zomer A, Stunnenberg HG, Burghout P, Hermans PWM: **Nontypeable *Haemophilus influenzae* carbonic anhydrase is important for environmental and intracellular survival.** *J Bacteriol* 2013, **195**:2737-2746.
33. Perkins TT, Davies MR, Klemm EJ, Rowley G, Wileman T, James K, Keane T, Maskell D, Hinton JCD, Dougan G *et al.*: **ChIP-seq and transcriptome analysis of the OmpR regulon of *Salmonella enterica* serovars Typhi and Typhimurium reveals accessory genes implicated in host colonization.** *Mol Microbiol* 2013, **87**:526-538.
34. Zhao HQ, Chen C, Xiong YW, Xu XF, Lan RT, Wang HY, Yao XY, Bai XN, Liu XT, Meng Q *et al.*: **Global transcriptional and phenotypic analyses of *Escherichia coli* O157:H7 strain Xuzhou21 and its pO157\_Sal cured mutant.** *PLoS ONE* 2013, **8**.
35. Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK, Bateman A, Parkhill J, Wain J, Gardner PP: **A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium.** *Nucleic Acids Res* 2013, **41**:4549-4564.
36. Van Opijnen T, Camilli A: **A fine scale phenotype-genotype virulence map of a bacterial pathogen.** *Genome Res* 2012, **22**:2541-2551.
- The study uses transposon mutagenesis in more than 17 different environments to compare genes required for adaptation in different conditions.
37. De Vries SPW, Eleveld MJ, Hermans PWM, Bootsma HJ: **Characterization of the molecular interplay between *Moraxella catarrhalis* and human respiratory tract epithelial cells.** *PLoS ONE* 2013, **8**.
38. Wong SM, Bernui M, Shen H, Akerley BJ: **Genome-wide fitness profiling reveals adaptations required by *Haemophilus* in coinfection with influenza A virus in the murine lung.** *Proc Natl Acad Sci U S A* 2013, **110**:15413-15418.
- The first application of Tn-seq for analysis of bacterial genes involved in co-infection.
39. Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, Holden MTG, Dougan G, Bentley SD, Parkhill J *et al.*: **Routine use of microbial whole genome sequencing in diagnostic and public health microbiology.** *PLoS Pathog* 2012, **8**:e1002824.
40. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR *et al.*: **Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak.** *N Engl J Med* 2012, **366**:2267-2275.
- One of the first applications of HTS to tracing a hospital outbreak of infection.
41. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, Pallen MJ: **High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak.** *J Hosp Infect* 2010, **75**:37-41.
42. McAdam PR, Templeton KE, Edwards GF, Holden MTG, Feil EJ, Aanensen DM, Bargawi HJA, Spratt BG, Bentley SD, Parkhill J *et al.*: **Molecular tracing of the emergence, adaptation, and transmission of hospital-associated methicillin-resistant *Staphylococcus aureus*.** *Proc Natl Acad Sci U S A* 2012, **109**:1-6.
43. Holden MTG, Hsu L-Y, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H *et al.*: **A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic.** *Genome Res* 2013, **23**:653-664.

Papers 42–43 highlighting the genetic changes preceding the global emergence of hospital associated MRSA.

44. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodtkin E, Rempel S, Moore R, Zhao Y, Holt R *et al.*: **Whole-genome sequencing and social-network analysis of a tuberculosis outbreak.** *N Engl J Med* 2011, **364**:730–739.
- A landmark study combining genome sequences and epidemiological patient information to identify transmission of tuberculosis during a sustained outbreak.
45. Baker S, Holt KE, Clements ACA, Karkey A, Arjyal A, Boni MF, Dongol S, Hammond N, Koirala S, Duy PT *et al.*: **Combined high-resolution genotyping and geospatial analysis reveals modes of endemic urban typhoid fever transmission.** *Open Biol* 2011, **1**:110008.
46. Holt KE, Thieu Nga TV, Thanh DP, Vinh H, Kim DW, Vu Tra MP, Campbell JI, Hoang NVM, Vinh NT, Minh PV *et al.*: **Tracking the establishment of local endemic populations of an emergent enteric pathogen.** *Proc Natl Acad Sci U S A* 2013, **110**:17522–17527.
47. Holt KE, Baker S, Weill F-X, Holmes EC, Kitchen A, Yu J, Sangal V, Brown DJ, Coia JE, Kim DW *et al.*: **Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe.** *Nat Genet* 2012, **44**:1056–1059.
48. Reuter S, Harrison TG, Köser CU, Ellington MJ, Smith GP, Parkhill J, Peacock SJ, Bentley SD, Török ME: **A pilot study of rapid whole-genome sequencing for the investigation of a Legionella outbreak.** *BMJ Open* 2013, **3**.
49. Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TEA, Walker AS, Wilson DJ: **Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in Clostridium difficile transmission.** *PLoS Comput Biol* 2013, **9**:e1003059.
50. Gray RR, Tatem AJ, Johnson JA, Alekseyenko AV, Pybus OG, Suchard MA, Salemi M: **Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant Staphylococcus aureus ST239 genome-wide data within a Bayesian framework.** *Mol Biol Evol* 2011, **28**:1593–1603.
51. Spoor LE, McAdam PR, Weinert LA, Rambaut A, Hasman H, Aarestrup FM, Kearns AM, Larsen AR, Skov RL, Fitzgerald JR: **Livestock origin for a human pandemic clone of community-associated methicillin-resistant Staphylococcus aureus.** *MBio* 2013, **4** e00356–13.
52. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M *et al.*: **Evidence for several waves of global transmission in the seventh cholera pandemic.** *Nature* 2011 <http://dx.doi.org/10.1038/nature10392>.
53. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, Kariuki S, Msefula CL, Gordon MA, de Pinna E *et al.*: **Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa.** *Nat Genet* 2012 <http://dx.doi.org/10.1038/ng.2423>.
54. Mather AE, Reid SWJ, Maskell DJ, Parkhill J, Fookes MC, Harris SR, Brown DJ, Coia JE, Mulvey MR, Gilmour MW *et al.*: **Distinguishable epidemics of multidrug-resistant Salmonella Typhimurium DT104 in different hosts.** *Science* 2013, **341**:1514–1517.
55. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolala V, Pearson T, Waters AE, Prasad Upadhyay B *et al.*: **Population genetics of Vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak.** *MBio* 2011, **2** e00157–11.
56. Katz LS, Petkau A, Beaulaurier J, Tyler S, Antonova ES, Turnsek MA, Guo Y, Wang S, Paxinos EE, Orata F *et al.*: **Evolutionary dynamics of Vibrio cholerae O1 following a single-source introduction to Haiti.** *MBio* 2013, **4** e00398–13.
57. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS *et al.*: **Rapid pneumococcal evolution in response to clinical interventions.** *Science* 2011, **331**:430–434.
58. Golubchik T, Brueggemann AB, Street T, Gertz RE, Spencer CCA, Ho T, Giannoulatos E, Link-Gelles R, Harding RM, Beall B *et al.*: **Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event.** *Nat Genet* 2012, **44**:352–355.
59. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP, Lipsitch M: **Population genomics of post-vaccine changes in pneumococcal epidemiology.** *Nat Genet* 2013, **45**:656–663.
- 3 excellent papers 57–59 highlighting the genetic changes in *S. pneumoniae* populations following the introduction of a vaccine.
60. Frank C, Werber D, Cramer JP, Askar M, Faber M, an der Heiden M, Bernard H, Fruth A, Prager R, Spode A *et al.*: **Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany.** *N Engl J Med* 2011, **365**:1771–1780.
61. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin C-S, Iliopoulos D *et al.*: **Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany.** *N Engl J Med* 2011, **365**:709–717.
62. Rohde H, Qin J, Cui Y, Li D, Loman NJ, Hentschke M, Chen W, Pu F, Peng Y, Li J *et al.*: **Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4.** *N Engl J Med* 2011, **365**:718–724.
63. Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J *et al.*: **Staphylococcus aureus CC398: host adaptation and emergence of methicillin resistance in livestock.** *MBio* 2012, **3**.
64. Harrison EM, Paterson GK, Holden MTG, Larsen J, Stegger M, Larsen AR, Petersen A, Skov RL, Christensen JM, Bak Zeuthen A *et al.*: **Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC.** *EMBO Mol Med* 2013, **5**:509–515.
65. Lebreton F, van Schaik W, McGuire AM, Godfrey P, Griggs A, Mazumdar V, Corander J, Cheng L, Saif S, Young S *et al.*: **Emergence of epidemic multidrug-resistant Enterococcus faecium from animal and commensal strains.** *MBio* 2013, **4** e00534–13.
66. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A, Cartwright RA, Simpson AJ, Rambaut A, Nübel U, Fitzgerald JR: **Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus.** *Proc Natl Acad Sci U S A* 2009, **106**:19545–19550.
67. Sheppard SK, Didelot X, Meric G, Torralba A, Jolley KA, Kelly DJ, Bentley SD, Maiden MCJ, Parkhill J, Falush D: **Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter.** *Proc Natl Acad Sci U S A* 2013, **110**:11923–11927.
- One of the first applications of GWAS to bacteria: identification of genetic correlates of host specificity.
68. Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang JH, van Diemen PM, Buckley AM, Bowen AJ *et al.*: **Comprehensive assignment of roles for Salmonella Typhimurium genes in intestinal colonization of food-producing animals.** *PLoS Genet* 2013:9.
- An excellent study demonstrating the use of transposon mutagenesis *in vivo* to identify genes associated with survival in multiple host species.
69. Kingsley RA, Kay S, Connor T, Barquist L, Sait L, Holt KE, Sivaraman K, Wileman T, Goulding D, Clare S *et al.*: **Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted Salmonella enterica serovar Typhimurium pathovar.** *MBio* 2013, **4** e00565–13.
70. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE *et al.*: **Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.** *Nat Methods* 2013, **10**:563–569.
71. Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZM, Quick J, Weir JC, Quince C, Smith GP, Betley JR *et al.*: **A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxicogenic Escherichia coli.** *JAMA* 2013, **309**:1502.
- Demonstration of the potential of metagenomic analysis of complex samples for investigating outbreaks in a culture-independent manner.

72. Seth-Smith HMB, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C *et al.*: **Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture.** *Genome Res* 2013 <http://dx.doi.org/10.1101/gr.150037.112>.
73. Westermann AJ, Gorski SA, Vogel J: **Dual RNA-seq of pathogen and host.** *Nat Rev Microbiol* 2012, **10**:618-630.
74. Falush D, Bowden R: **Genome-wide association mapping in bacteria?** *Trends Microbiol* 2006, **14**:353-355.
75. Priest NK, Rudkin JK, Feil EJ, van den Elsen JMH, Cheung A, Peacock SJ, Laabei M, Lucks DA, Recker M, Massey RC: **From genotype to phenotype: can systems biology be used to predict *Staphylococcus aureus* virulence?** *Nat Rev Microbiol* 2012, **10**:791-797.
76. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, Williams P, Endres JL, Bayles KW, Fey PD *et al.*: **Predicting the virulence of MRSA from its genome sequence.** *Genome Res* 2014, **24**:839-849.