

REVIEW

Open Access



Next-generation microbiology: from comparative genomics to gene function

Carolyn M. Kobras¹ , Andrew K. Fenton^{1*}  and Samuel K. Sheppard^{2*} 

* Correspondence: A.K.Fenton@sheffield.ac.uk; S.K.Sheppard@bath.ac.uk

¹Department of Molecular Biology & Biotechnology, University of Sheffield, The Florey Institute for Host-Pathogen Interactions, Sheffield, UK

²Department of Biology & Biochemistry, University of Bath, Milner Centre for Evolution, Bath, UK

Abstract

Microbiology is at a turning point in its 120-year history. Widespread next-generation sequencing has revealed genetic complexity among bacteria that could hardly have been imagined by pioneers such as Pasteur, Escherich and Koch. This data cascade brings enormous potential to improve our understanding of individual bacterial cells and the genetic basis of phenotype variation. However, this revolution in data science cannot replace established microbiology practices, presenting the challenge of how to integrate these new techniques. Contrasting comparative and functional genomic approaches, we evoke molecular microbiology theory and established practice to present a conceptual framework and practical roadmap for next-generation microbiology.

Introduction

Experimental approaches for studying bacteria have changed dramatically over the last 20 years [1]. Shifting in response to public interest and fuelled by technological advances, understanding of these remarkable organisms continues to rapidly advance. We now know more than ever before about the metabolism, environmental context and host interactions of microbes, and the rate of discovery shows little sign of slowing. Among the most influential shifts in technology has been the increasing use of large sequencing datasets in research practice. These contemporary research approaches continue to gain momentum, expanding into ever more ingenious ways of using sequencing data to discover complex patterns of behaviour and reach a deeper understanding of the bacterial cell. This rapid advancement has many conceptual benefits but has also come at a significant cost, as laboratories struggle to integrate these techniques and apply best research practices to new types of data.

The magnitude and complexity of large sequencing datasets can make them appear abstract to the non-specialist, potentially leading to subjective judgements about whether to believe the analyses or not. This can, in turn, risk general disenfranchisement of microbiology researchers away from genomic data, promoting an over-reliance on outside proofs or validations to give meaning to sequencing-based datasets. Here, we argue for an integrated future for microbiology that combines the strengths



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of traditional microbiology with the promise of emergent sequencing technologies. Addressing the widening gap in research practice, we discuss some of the most influential methodologies, the validation of findings from large sequencing datasets, and how comparative and functional genomics can be integrated to advance microbiology from fundamental discovery to contemporary microbiology research practice.

Using a data deluge for qualitative and quantitative microbial genomics

It has been well over a decade since next-generation sequencing (NGS) platforms became widely available for microbial genomics. The cost of sequencing has continued to fall to a point where large sequence datasets are within the budget of most research groups. This democratisation of technology was not driven by a fundamental change in how DNA is sequenced. In fact, the major shift came through the upscaling of bridge amplification in the Illumina sequencing-by-synthesis process [2, 3]. This allowed the simultaneous sequencing of millions of individual DNA molecules in parallel by NGS machines generating huge amounts of data. While new single-molecule sequencing technologies developed by Oxford Nanopore and Pacific Biosystems gather momentum [4], the massively parallel Illumina NGS approach remains a major driver in the generation of large-scale DNA sequencing datasets. Key to the widespread use of NGS methodology are the diverse applications. Broadly, the functionality can be described under two contrasting modes. The first is a high-accuracy DNA sequencing function best applied on either *de novo* genome sequencing or making detailed comparisons between genomes. Here, the huge numbers of individual sequencing reads are combined to remove errors in base calling and generate high-confidence ensemble averages. The second mode is a counting function used to survey mixed populations of DNA or RNA molecules. Here, each sequencing read is examined individually, separated into groups and scored. This approach can, for example, be applied to measure the relative frequencies of mRNA levels in a cell or to capture the composition of a bacterial population from an environmental sample.

Transformative sequencing technologies and the genetics of phenotype variation

Determining the genetic basis of phenotype variation is among the most pervasive aims in microbiology. This is a major challenge and requires understanding of how changes to genes, and their constituent DNA sequences, can alter gene function and affect a phenotypic change over time. Two of the most transformative techniques that address this in bacteria are genome-wide association studies (GWAS) [5–7] and transposon insertion sequencing methods (here referred to as Tn-Seq, but also known as HITS, InSeq or TraDIS) [8–11]. Both techniques are powered by NGS, but each uses different functions of DNA sequencing technologies. GWAS requires genomes from multiple strains within a population to identify genomic elements that are statistically associated with a given phenotype or environmental condition [12, 13] and therefore uses the high-accuracy function of NGS. In contrast, Tn-seq profiles fewer strains and uses the DNA counting function to identify transposon insertions in populations of mutants to identify the contribution each gene makes to bacterial survival within the specific experimental context [14, 15].

Describing population-wide genomic variation

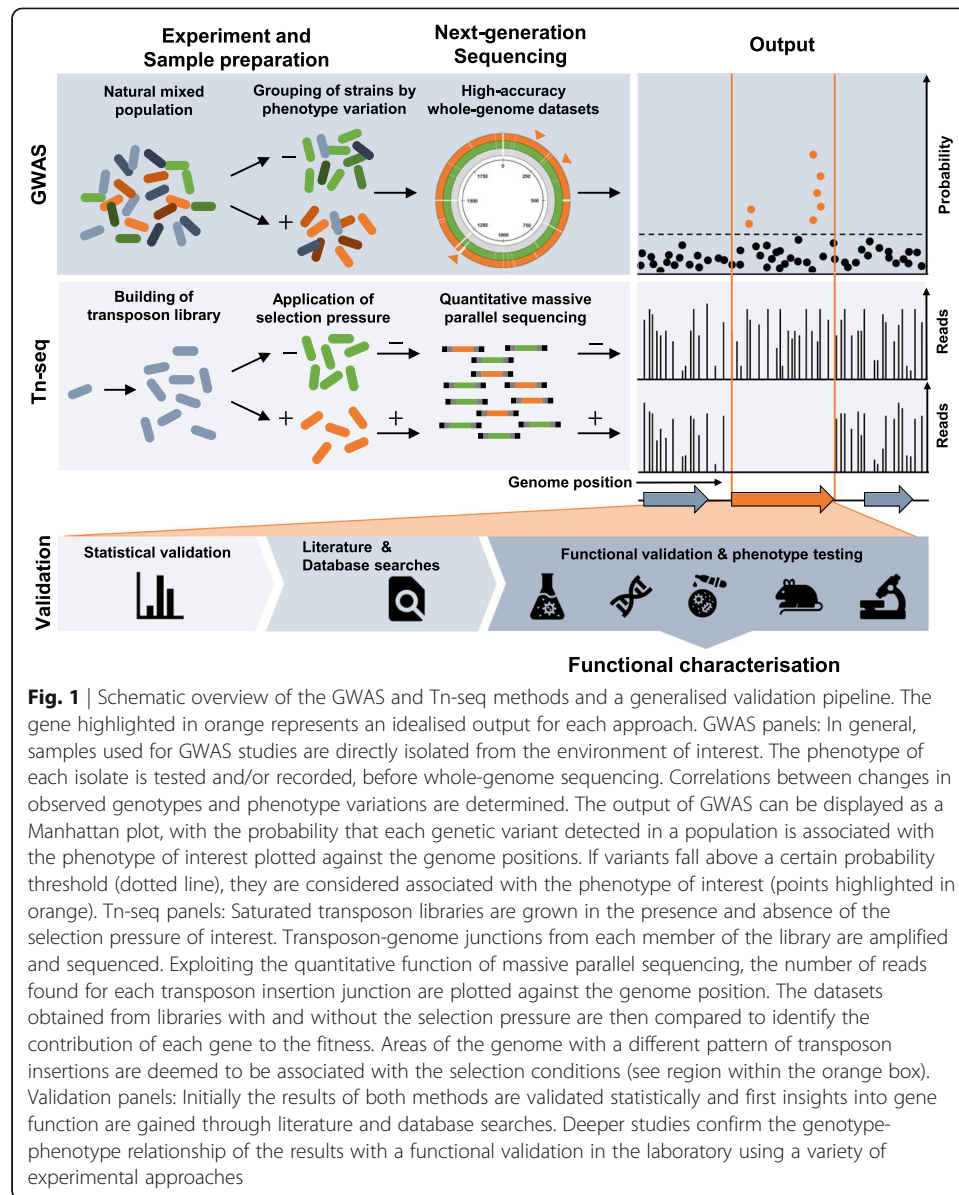
The availability of numerous high-quality bacterial genomes representing the extraordinary complexity of phenotypic and genotypic variation in natural populations has inevitably led microbiologists to new analytical techniques. Drawing on methods that were pioneered in human genetics, early bacterial GWAS approaches [5] have been adapted to become an important *in silico* tool for population-wide genomic screening [16]. Studies typically involve sampling and genome sequencing of hundreds of isolates from different environments or conditions and identifying genetic elements (e.g. single nucleotide polymorphisms (SNPs), k-mers or accessory genetic elements) that are significantly associated with a phenotype in question (Fig. 1). Now widely used, bacterial GWAS have successfully identified candidate genes involved in host specificity [5, 17], virulence [6, 18–24], the duration of pathogen carriage [25], and antibiotic resistance [7, 26–29].

The widespread application of bacterial GWAS has been made possible by adapting the methodological and analytical assumptions of human GWAS in two important ways. First, bacterial GWAS not only targets homologous sequence variation but also aims to identify the numerous accessory genetic elements and genes that may be found in some, but not all, isolate genomes [5, 30]. Second, and most importantly, it accounts for the strong linkage disequilibrium resulting from the clonal mode of bacterial reproduction. Accounting for this population structure is particularly important when considering the genetics underlying phenotype variation as causal variants will be co-inherited with linked loci that may have no adaptive function [12, 13]. In highly structured bacterial populations entire clusters of strains may share elements that have facilitated their expansion as well as those that simply reflect common ancestry. To address this, population subsampling [20, 31], linear mixed models [27, 32] and phylogenetic trees [21] can be incorporated into analyses to account for the clonal frame of the population. Resultant associations that cannot be explained by the effect of shared ancestry can represent convergent genomic signatures in groups of divergent strains. This provides clues to the evolutionary forces acting on the bacterial genome.

Sophisticated bioinformatics analyses of ever larger genome collections are: (i) incorporating quantitative trait variation [28]; (ii) conditioning on multiple genomic or phenotypic determinants [20, 29]; (iii) using machine learning to quantify the relative importance of associated elements in explaining the observed phenotype [20, 31, 33]. However, while bacterial GWAS approaches benefit from retaining the natural population setting of a given phenotype, they often return many thousands of genetic elements associated with complex traits such as host association or virulence [34]. In such cases, it can be extremely difficult to identify the role of individual genes and unravel the myriad interacting selective effects that shape the observed genomic variation. For this it may be necessary to move beyond *in silico* statistical associations and understand the function and importance of specific genes under more carefully controlled conditions.

Studying gene function through modification and inactivation

Observing how small genomic differences, in otherwise isogenic strains, influence the phenotype provides evidence about the functional consequence of sequence variation.



Over decades, microbiologists have identified the function of numerous genes across multiple species mainly through investigating the effect of gene loss. It is possible to infer gene function by inactivating specific genes, usually through introduction of a specific mutation into the genome of an organism and comparing the resultant phenotype to that of a ‘wild-type’ strain. While this is relatively laborious compared to observing genomic variation in natural populations *in silico*, it provides much greater control of the genomic variation and the conditions in which the gene function is being tested. Extending the principle of gene inactivation for genome-wide functional studies, ordered gene deletion libraries have been generated for several model laboratory strains [35–38]. In these libraries, all non-essential genes have been disrupted by the insertion of antibiotic markers, allowing the rapid screening of phenotypes under different selective conditions. Furthermore, random chemical or UV mutagenesis have been used to generate ordered mutant libraries, without requiring any a priori genetic manipulation

of the bacterial strain [39]. While this allows investigations into bacterial species that are hard to manipulate genetically, it may be difficult to generate sufficient mutations for complete gene coverage in the screen, particularly as the whole genome of each strain needs to be sequenced to locate a mutation.

Whole-genome fitness profiling using transposon insertion mutagenesis

Building on the concept of using large-scale gene deletion libraries that cover the entire genome of the bacterium, a revolution in these methodologies began just over a decade ago with the integration of quantitative high-throughput NGS technologies that capture the complexity of a large transposon-insertion library in one sequencing step [14, 15]. Developed around the same time, conceptually similar techniques including Tn-seq [8], TraDIS [9], HITS [10], and INSeq [11], all use large transposon insertion libraries, across which all or most non-essential genes contain transposon insertions. Selection pressure is applied to these library strains by growing them in defined *in vitro* or *in vivo* conditions (Fig. 1). Subsequent amplification and sequencing of the transposon-genome junctions in the libraries allows the insertion location of each transposon to be determined for each condition. The key feature of these approaches are the resulting ‘profiles’ of transposon insertions that reflect the fitness contribution each gene had under the selective conditions of the experiment. Specifically, regions of the genome where transposon insertions are statistically underrepresented likely contain genes that are essential for the bacteria to survive in the experimental conditions [8, 40–42].

This whole-genome fitness profiling method has linked many genes with metabolic pathways [43] and important phenotypes including stress response and antibiotic resistance [44–46], virulence and survival in the host environment [9–11, 47–51]. Furthermore, by deleting specific query genes it may be possible to identify gene interactions [8, 52, 53] and to examine the role of non-coding and regulatory DNA [54]. Most Tn-seq approaches rely on negative selection via gene inactivation. However, transposons carrying outward facing promoters can result in the upregulation of neighbouring genes. This allows controlled analysis of functional gene upregulation, an approach applied to the study of antibiotic resistance [55–57].

Recently, CRISPR interference (CRISPRi)-based methods have been added to the assortment of functional genomic tools [58]. Here, a small guide RNA forms a complex with the inactivated DNA-binding protein Cas9 and together they bind a specific region of the genome. The complex blocks RNA polymerase at the targeted site through steric hindrance, and represses transcription of the targeted gene [59, 60]. In a genome-wide screen, large libraries of mutants, each containing a different CRISPRi construct, can be captured by high-throughput NGS [61–66]. In contrast to Tn-seq, CRISPRi libraries have the potential of covering all genes in a genome, including essential genes. However, secondary and off-target effects still have to be carefully considered.

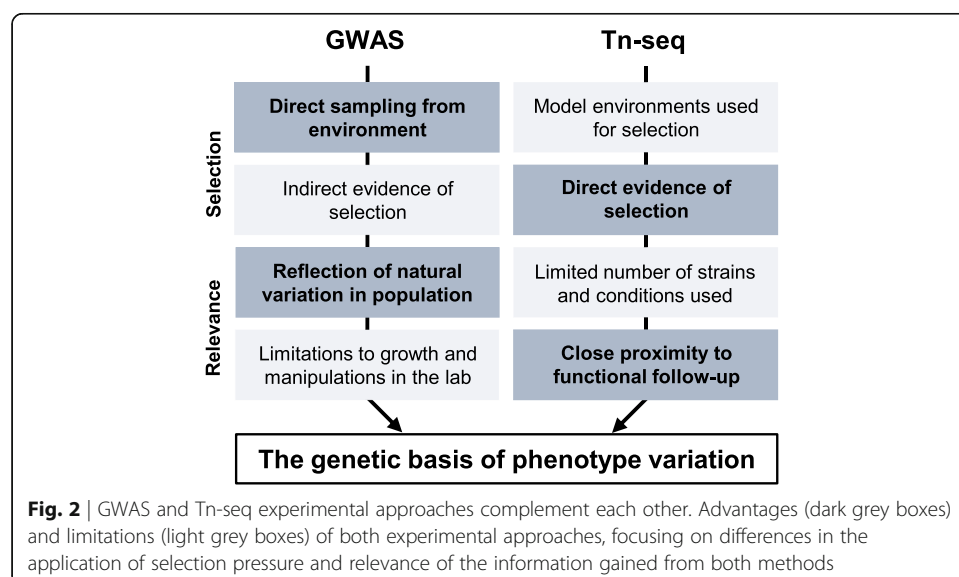
Understanding bacteria in the wild

When trying to understand the genomics underlying trait variation in bacteria microbiologists must make compromises. The major challenge is to balance the experimental control needed to understand the function of specific genes with the requirement for data that is relevant in natural populations. This is illustrated by contrasting the

selective conditions in Tn-Seq and GWAS approaches (Fig. 2). By deliberately limiting the selection pressures, Tn-seq studies provide a clear path to the functional validation of genes, often involving recreation of the initial experimental conditions used for the Tn-seq screen and measuring the fitness of genetically modified bacteria in competition-based assays [8, 44, 67]. However, in some cases there is surprisingly little overlap among the genes required for growth in particular conditions when comparing datasets between different laboratories [68]. This may be because of differences in: the precise experimental conditions; the transposons used; false-positives resulting from polar effects of transposon insertion; library selection and handling methods [69]. While the high-throughput nature of these methods, and appropriate validation, can largely overcome the challenge of reproducibility, the major strength of Tn-seq can also be considered its limitation. Specifically, while the deliberate constraint of the selection conditions acting on the bacteria facilitates functional genomics, these studies can also be criticised for lacking ‘real-world’ insight into genotype-phenotype relationships.

To enhance the relevance of laboratory findings for natural bacterial populations, recent multi-strain Tn-seq studies have included clinical or environmental isolates [42, 44, 46, 57]. However, there is a clear benefit to inference from bacteria in the wild. In this respect, GWAS is a powerful approach, as it directly surveys natural genotype-phenotype associations. This inevitably means that bacteria are sampled from dynamic systems and will have been exposed to a complex set of selection pressures, not all of which are directly related to the primary condition of interest. Structured sampling, replication and statistical tests, and *in silico* validations can strengthen assumptions about causal genetic variations [20], but recreating the experimental conditions in a laboratory setting may be extremely difficult, leading to difficulty when trying to understand the value of GWAS for lab-based microbiologists.

The complementary strengths and limitations of population-wide screens and laboratory fitness profiling methods provide a means to identify the genetic basis of complex bacterial traits (Fig. 2). Therefore, in combination, techniques such as GWAS and Tn-seq could provide insights into the behaviours of bacteria in natural environments in



such a way as to be experimentally tractable in a laboratory setting. The initial steps taken to ensure data quality for both methods are similar, involving technical repeats and statistical validation. Yet, it is the experimental proof that specific genetic variants cause observable phenotypes that makes these studies so impactful. The challenges are how to achieve these proofs, what experimental methods should be used, and against what guidelines might we measure the evidence.

Functional validation in the post-genomics era

To the data analyst, functional genomic inference may be considered ‘validated’ if associations are proven robust against a series of statistical challenges. However, for laboratory-based researchers, in silico findings are typically considered ‘validated’ only when their effects can be reproduced using a complementary experimental approach. This requirement for experimental reproducibility has been a central tenet in microbiology since the publication of Koch’s postulates [70]. Adapting this conceptual framework in 1988, Stanley Falkow established a set of rules to prove causality of molecular genetic changes to disease phenotypes (Fig. 3) [70–72]. Subsequently adjusted to fit different research areas [73–76], these Molecular Koch’s postulates remain engrained in molecular microbiology best practice because of the scientific rigour they promote.

A major limitation of population-scale and genome-wide genetic screens is that a relatively small fraction of candidate genes are functionally validated. In some cases, in silico and laboratory-based genomic screens have employed follow-up gene inactivation and phenotype investigations to link the function of specific genes to pathogenicity [6, 18, 20, 31, 48, 77], survival and transmission [17, 19, 51] and antimicrobial resistance [29, 44, 57]. This experimental confirmation can be a challenging task given the large numbers of genes involved in complex phenotypes but it remains important for robust genotype-phenotype association. To address this we propose revised Molecular Koch’s postulates for functional genomic validation of NGS analyses (Fig. 3).

Molecular Koch’s postulates	Next-generation Koch’s postulates
1. The phenotype or property under investigation should be associated with pathogenic members of a genus or pathogenic strains of a species.	1. The genetic variant under investigation should be significantly associated with the phenotype. Preferably a strong association, which is true across several bacterial strains or species.
2. Specific inactivation of the gene(s) associated with the suspected virulence trait should lead to a measurable loss in pathogenicity or virulence, or the gene(s) associated with the supposed virulence trait should be isolated by molecular methods. Specific inactivation or deletion of the gene(s) should lead to loss of function in the clone.	2. Specific recreation of the genetic variant associated with the suspected phenotype should lead to a measurable change thereof. In most cases this is best met through gene inactivation. In the case of an essential gene, a method of depletion is often used to investigate changes in phenotype.
3. Reversion or allelic replacement of the mutated gene should lead to restoration of pathogenicity, or the replacement of the modified gene(s) for its allelic counterpart in the strain of origin should lead to loss of function and loss of pathogenicity or virulence. Restoration of pathogenicity should accompany the reintroduction of the wild-type gene(s).	3. Reversion or allelic replacement of the genetic variant should lead to restoration of the phenotype.

Fig. 3 | Molecular Koch’s postulates, including a generalised revision for the purposes of this review. Stanley Falkow’s adaptation of Koch’s postulates (left) have been the gold standard to support causal links between genotypes and bacterial phenotypes for decades [71, 72]. To provide wider accessibility and application to microbiology research more generally, we have revised these postulates (right). Importantly, these postulates are adapted to guide functional validation and the authors acknowledge the full set might not be fulfilled in all cases

In most cases, genomic screening approaches will automatically fulfil the first postulate - that bacterial strains with an identified genetic variant should display the phenotype of interest (Fig. 3). Ideally, the relationship between genetic change and phenotype of interest would be as direct as possible, shifting only one experimental variable at a time and showing a large effect size. In this way, focusing on strong correlations can be useful as it provides the best experimental proofs.

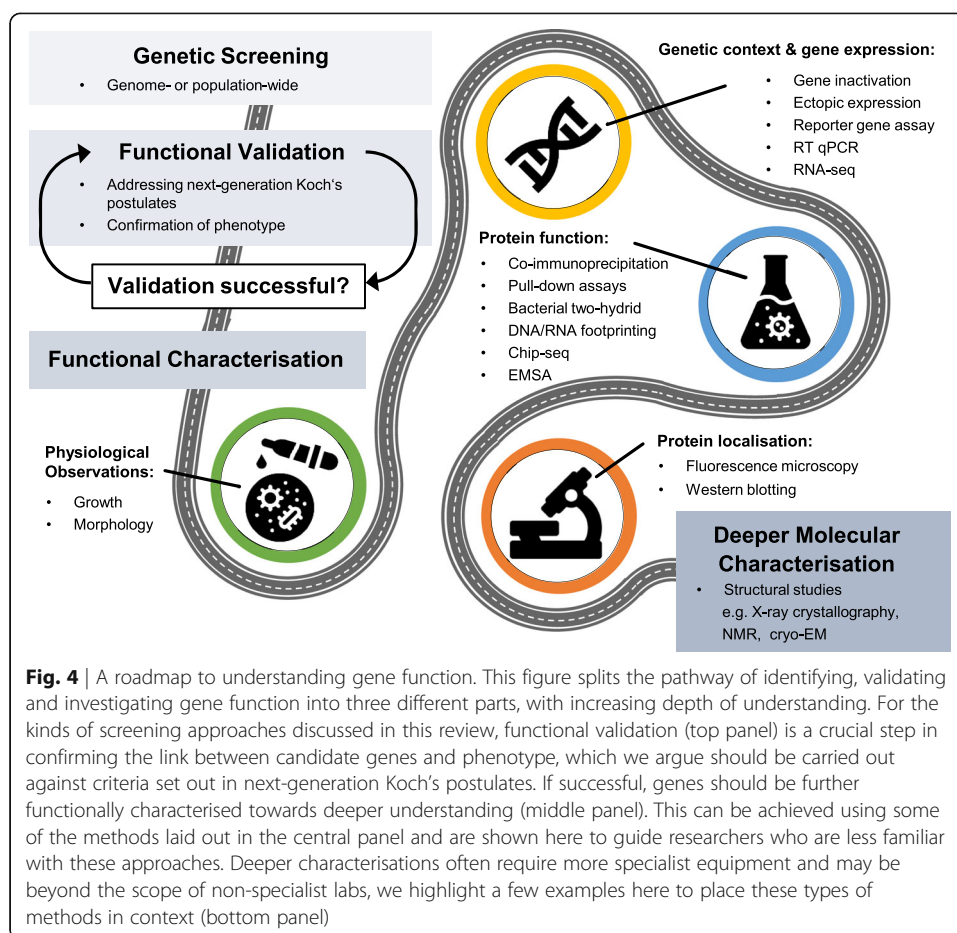
According to the second postulate, specific changes made to the gene of interest should result in a change to the phenotype in question (Fig. 3). While not all genetic variants will result in the loss of the gene function, the generation of targeted gene deletions and reproduction of the experimental conditions leading to the expected phenotype is an approach often used to meet this postulate. To achieve this, tools for marked and markerless gene deletions have been developed [78–80], with the availability of ordered transposon or single-gene deletion libraries expediting this process for some bacterial species [35, 38, 81]. In cases where a gene is essential for the survival of a bacterium, depletion systems can be used but often require an established set of genetic tools to function. A popular example for sequence-specific repression of gene expression is CRISPRi [59, 60].

Finally, the third postulate focusses on restoring the observed phenotype through genetic complementation (Fig. 3). This is an essential step to close the loop and prove causation but may require a more sophisticated set of genetic tools to achieve. The most direct example is complementation via a conditional expression system, usually on a plasmid or at an ectopic locus in the genome. Alternatively, more subtle genetic manipulates can be used to meet this postulate, for example the introduction of a base pair change into the genome that complements the phenotype, an approach often achieved by site-directed mutagenesis [82].

The functional validation of genomic screens, potentially based on these revised Molecular Koch's postulates, provides an ambitious target. In fact, it may be extremely difficult in practice to identify, remove and reinstate the genetics underlying trait variations. For example, where multiple independent variations cause subtle phenotypic changes or where genes are part of interactive networks and co-vary because of epistasis [83, 84]. As such, Molecular Koch's postulates exist principally as a 'gold standard' rather than a definitive list of experimental criteria. In practice, integrated microbiology should layer multiple experimental approaches, using the minimum number of methods required to meet the burden of proof and focus validation effort for functional follow-up studies.

A roadmap for next-generation functional microbiology

Identifying and then proving that a genetic variant causes a phenotypic change is a considerable step towards understanding bacterial genomics. However, determining the precise role of the gene and how the encoded protein may function requires further study. This can be challenging, particularly for researchers with no background in molecular microbiology, even when the genetic methods or biochemical assays require fairly basic laboratory equipment. Therefore, just as laboratory microbiologists are encouraged to embrace contemporary genomic approaches, so bioinformaticians might consider the central role of functional microbiology in various ways (Fig. 4).



Physiological observations

Although considered rather simplistic, basic physiological observations are an important starting point for studies of gene function in microbiology. Differences in growth rate, signs of growth arrest, and early lysis are often tell-tale signs and should not be overlooked. Live-cell microscopy can also provide insights into whether a gene of interest affects cell morphology.

Genetic context and gene expression

In bacteria, genes with related functions often cluster together in operons. Therefore, deciphering the genetic context of candidate genes, as well as understanding where and when a gene is expressed in a bacterial cell, can provide insights into how they function. Determining the pattern and timing of gene expression can be accomplished by replacing the coding sequence of candidate genes with a reporter construct, allowing expression to be monitored through fluorescence, luminescence or enzymatic activity [85]. A more direct measure of gene expression is the determination of relative amounts of mRNA by reverse-transcription quantitative PCR (RT-PCR) [86]. Broadening this approach, techniques such as microarrays or RNA-seq represent powerful methods that may allow genome-wide functional transcriptomic analysis [87–89].

Protein function

When considering the biological role of a gene it is important to understand the function of the protein it encodes, especially if the protein is thought to have an enzymatic function or is likely to interact with binding partners in a wider network. This can include protein–protein interactions, protein–DNA/RNA binding, and other substrate or ligand interactions. Commonly used methods to identify protein-protein interactions are co-immunoprecipitation and protein pull-down assays. These conceptually similar approaches use antibodies to specifically recognise and isolate the (tagged) query protein from a cell lysate, bringing any binding partners with it [90]. Putative binding partners are then identified by mass spectrometry or other similar techniques. Alternative approaches include bacterial two-hybrid systems that are designed to report protein-protein interactions in vivo through expression of a reporter gene, most commonly beta-galactosidase or luciferase. This can be a very rapid approach to discovering or confirming individual predicted interactions, but may also be used to screen libraries of potential protein partners [91, 92].

Classically, DNA- or RNA-protein interactions are identified using DNA/RNA footprinting. Here, protein-bound DNA or RNA molecules are protected from cleavage by nuclease enzymes causing gaps in the digestion patterns when compared to DNA/RNA only controls [93]. More recently, genomic footprinting techniques based on NGS have been established, replacing the final gel separation steps with sequencing [94]. Genome-wide profiles of protein-DNA interactions can be further studied through combining chromatin immunoprecipitation with NGS (ChIP-seq) [95, 96]. Here, the query protein is fixed to the interacting DNA through chemical crosslinking in vivo. These complexes are enriched by immunoprecipitation and, after crosslink reversal, the DNA fragments are released and identified by NGS. ChIP-seq is especially useful for proteins with multiple binding sites, e.g. transcription factors. More focused methods for identifying specific protein-DNA/RNA binding, such as electrophoretic mobility shift assays (EMSA), exploit slower migration rates of protein-nucleic acid complexes in gels compared to nucleic acids alone [97]. Whatever the scale, protein interaction studies are essential for further molecular characterisation of genomic approaches, adding depth and context to candidate genes and advancing our understanding of the bacterial cell.

Protein localisation

Important clues to the specific function of a protein can be derived by examining its subcellular localisation. Separation of the bacterial cell into simple fractions (such as: membrane, cytoplasm, cell wall) can, in combination with western blotting, give a first indication. However, more specific insights into protein localisation can be achieved by fluorescence microscopy. This method usually involves the creation of protein-reporter fusions and allows tracking of the fluorescent product inside the cell [98–101]. In addition, protein tags, which are made fluorescent through the introduction of a small molecule, have become increasingly popular [102, 103]. While we highlight live-cell microscopy here, the expertise and special equipment required for microscopy methods beyond this quickly scale in complexity with the need for higher resolution.

Deeper molecular characterisation

The methods described so far should be practicable in most molecular microbiology laboratories. However, for many laboratories, this is where the research endeavour begins. For example, structural biologists acquire a deeper mechanistic understanding of protein function using methods such as X-ray crystallography [104], nuclear magnetic resonance (NMR) spectroscopy [105] and advanced optical methods such as (cryo-) electron microscopy [106], capitalising on the atomic resolution these methods bring.

Examples for integrated next-generation microbiology

We have described a direction of travel for next-generation microbiology. Other researchers share the same vision and some studies have begun to bridge the gap and integrate large sequencing datasets with molecular microbiology. For example, early bacterial GWAS made simple comparisons between the putative function of the genes containing associated elements and basic bacterial growth assays with defined substrates [5]. In a more sophisticated approach, and consistent with next-generation Koch's postulates, several studies have used mutagenesis and complementation to test the causality of hits and their predicted phenotype [6, 17, 18, 29, 51, 77]. Going still further, some studies have characterised the function of candidate genes and their genetic context [48, 107, 108], sometimes over multiple publications [84, 109]. For example, combining protein localisation microscopy with protein-protein interaction studies defined the role of candidate genes in the regulation of cell wall biosynthesis [52, 53]. Consistent with this, putative antimicrobial resistance determinants have been investigated in multiple bacterial strain backgrounds to provide information that is increasingly relevant to natural systems [44, 46, 57].

Future directions

Microbiology research seeks to understand the workings of the bacterial cell in natural environments. Advances in DNA sequencing technologies have touched all aspects of microbiology research but this comes at a price. The specialism required to use these sequencing-based research methods risks a disconnect between bioinformatics and fundamental microbiology. This is because sequencing information is typically viewed through a series of analytical lenses to give it meaning. This means sequencing datasets are frequently abstract in nature and often the mathematical methods used to generate them are challenging to understand for non-specialists. With continued revisions to analytical methods and ever-increasing sample sizes, bioinformatic analysis has the potential to outpace fundamental microbiology investigations by orders of magnitude, exacerbating the analytical disconnect.

There are currently two principal solutions to this problem. First, bioinformaticians can collaborate widely, often contributing specific analytical expertise to each investigation. Second, bioinformatics in some areas has become a specialist data-service where researchers pay for analyses. In both cases, non-specialist researchers are divorced from the data, potentially leaning on internal controls or pre-assumed expectations to guide their interpretations. Similarly, bioinformaticians are removed from deeper understanding and characterisation of their initial discoveries. The typical conclusion is to argue for validated standardised analysis pipelines. This is sometimes necessary, such as in

clinical settings, but in a rapidly evolving field it is vital that researchers are given flexibility for future innovation.

The pace of bioinformatics has moved microbiology research towards the study of natural populations. However, fundamental molecular microbiology approaches continue to focus on laboratory-adapted model strains for consistency across research groups. In the future, it will be important to merge these contrasting approaches to deepen the impact of research studies. Already, association studies (e.g. GWAS) or genome-wide profiles (e.g. Tn-seq) are commonly challenged to validate the gene function of at least one hit, while more fundamental laboratory studies are often compelled to contextualise their findings among more representative strains. In each case, this rising demand places strain on specialist laboratories. Here, we argue for an integrated future for next-generation microbiology, embracing new analysis techniques and placing *in silico* findings in a microbiological context. The spirit of integrated research is captured by large research collaborations or consortia but integration does not have to be collaboration in the strictest sense, it can be embodied by a small number of individuals who understand complementary research methodologies, provided they find a way to meet the burden of proof set out in next-generation Koch's postulates. With careful reflection on current microbiology practice and a new awareness of the value of communicating complex datasets across disciplines, microbiology has never been in a better position to drive deeper understanding of the natural world.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02344-9>.

Additional file 1. Review history.

Acknowledgements

Icons used in Figs. 1 and 4 are taken from iconfinder.com and used under a creative commons licence and we acknowledge the creators Ivan Boyko, Daniel Bruce, Valera Zvonko, Keris Maker, Yogi Apprelliyanto, PICOL, Graphic Mall and Arana Graphics.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 1.

Authors' contributions

CMK, AKF and SKS conceived the review and jointly wrote the paper. The authors read and approved the final manuscript.

Authors' information

Twitter handles: @carokobras (Carolin M. Kobras); @AndrewKFenton (Andrew K. Fenton); @sheppard_lab (Samuel K. Sheppard).

Funding

CMK and AKF were funded by Medical Research Council (MRC) grant MR/S009280/1. SKS was supported by MRC grants MR/M501608/1 and MR/L015080/1.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 5 January 2021 Accepted: 8 April 2021

Published online: 29 April 2021

References

1. Bennett JW: Microbiology in the 21st Century. In: Designing the Microbial Research Commons: Proceedings of an Interantional Symposium. Volume 2. Edited by National Research Council (US) Board on Research Data and Information. Editor: P.F. Uhler. Washington (DC): National Academies Press (US); 2011.
2. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* 2008;9(1):387–402. <https://doi.org/10.1146/annurev.genom.9.081307.164359>.
3. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods.* 2008;5(1):16–8. <https://doi.org/10.1038/nmeth1156>.
4. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. *Nat Biotechnol.* 2020;38(6):701–7. <https://doi.org/10.1038/s41587-020-0422-6>.
5. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, et al. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. *PNAS.* 2013;110(29):11923–7. <https://doi.org/10.1073/pnas.1305559110>.
6. Laabei M, Recker M, Rudkin JK, Aldeljawi M, Gulay Z, Sloan TJ, et al. Predicting the virulence of MRSA from its genome sequence. *Genome Res.* 2014;24(5):839–49. <https://doi.org/10.1101/gr.165415.113>.
7. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, et al. Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. *Plos Genet.* 2014;10(8):e1004547. <https://doi.org/10.1371/journal.pgen.1004547>.
8. 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(10):767–72. <https://doi.org/10.1038/nmeth.1377>.
9. Langridge GC, Phan M-D, Turner DJ, Perkins TT, Parts L, Haase J, et al. Simultaneous assay of every *Salmonella typhi* gene using one million transposon mutants. *Genome Res.* 2009;19(12):2308–16. <https://doi.org/10.1101/gr.097097.109>.
10. Gawronski JD, Wong SM, 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. *PNAS.* 2009;106(38):16422–7. <https://doi.org/10.1073/pnas.0906627106>.
11. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, et al. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe.* 2009;6(3):279–89. <https://doi.org/10.1016/j.chom.2009.08.003>.
12. Read TD, Massey RC. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Med.* 2014;6(11):109. <https://doi.org/10.1186/s13073-014-0109-z>.
13. Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. *Curr Opin Microbiol.* 2015;25:17–24. <https://doi.org/10.1016/j.mib.2015.03.002>.
14. van Opijnen T, Camilli A. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol.* 2013;11(7):435–42. <https://doi.org/10.1038/nrmicro3033>.
15. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. A decade of advances in transposon-insertion sequencing. *Nat Rev Genet.* 2020;21(9):526–40. <https://doi.org/10.1038/s41576-020-0244-x>.
16. Power RA, Parkhill J, de Oliveira T. Microbial genome-wide association studies: lessons from human GWAS. *Nat Rev Genet.* 2017;18(1):41–50. <https://doi.org/10.1038/nrg.2016.132>.
17. Yahara K, Méric G, Taylor AJ, de Vries SP, Murray S, Pascoe B, et al. Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. *Environ Microbiol.* 2017;19(1):361–80. <https://doi.org/10.1111/1462-2920.13628>.
18. Laabei M, Uhlemann A-C, Lowy FD, Austin ED, Yokoyama M, Ouadi K, et al. Evolutionary trade-offs underlie the multi-faceted virulence of *Staphylococcus aureus*. *Plos Biol.* 2015;13(9):e1002229. <https://doi.org/10.1371/journal.pbio.1002229>.
19. Pascoe B, Méric G, Murray S, Yahara K, Mageiros L, Bowen R, et al. Enhanced biofilm formation and multi-host transmission evolve from divergent genetic backgrounds in *Campylobacter jejuni*. *Environ Microbiol.* 2015;17(11):4779–89. <https://doi.org/10.1111/1462-2920.13051>.
20. Méric G, Mageiros L, Pensar J, Laabei M, Yahara K, Pascoe B, et al. Disease-associated genotypes of the commensal skin bacterium *Staphylococcus epidermidis*. *Nat Commun.* 2018;9(1):5034. <https://doi.org/10.1038/s41467-018-07368-7>.
21. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *Plos Comp Biol.* 2018;14(2):e1005958. <https://doi.org/10.1371/journal.pcbi.1005958>.
22. Berthenet E, Yahara K, Thorell K, Pascoe B, Meric G, Mikhail JM, et al. A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. *BMC Biol.* 2018;16(1):84. <https://doi.org/10.1186/s12915-018-0550-3>.
23. Lees JA, Ferwerda B, Kremer PHC, Wheeler NE, Serón MV, Croucher NJ, et al. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. *Nat Commun.* 2019;10(1):2176. <https://doi.org/10.1038/s41467-019-09976-3>.
24. Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, et al. Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nat Genet.* 2019;51(3):548–59. <https://doi.org/10.1038/s41588-018-0343-1>.
25. Lees JA, Croucher NJ, Goldblatt D, Nosten F, Parkhill J, Turner C, et al. Genome-wide identification of lineage and locus specific variation associated with pneumococcal carriage duration. *eLife.* 2017;6:e26255. <https://doi.org/10.7554/eLife.26255>.
26. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet.* 2013;45(10):1183–9. <https://doi.org/10.1038/ng.2747>.
27. Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC, Walker TM, et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol.* 2016;1(5):16041. <https://doi.org/10.1038/nmicrobiol.2016.41>.

28. Farhat MR, Freschi L, Calderon R, loerger T, Snyder M, Meehan CJ, et al. GWAS for quantitative resistance phenotypes in *Mycobacterium tuberculosis* reveals resistance genes and regulatory regions. *Nat Commun.* 2019;10(1):2128. <https://doi.org/10.1038/s41467-019-10110-6>.
29. Ma KC, Mortimer TD, Duckett MA, Hicks AL, Wheeler NE, Sánchez-Busó L, et al. Increased power from conditional bacterial genome-wide association identifies macrolide resistance mutations in *Neisseria gonorrhoeae*. *Nat Commun.* 2020;11(1):5374. <https://doi.org/10.1038/s41467-020-19250-6>.
30. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J: pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics.* 2018;34(24):4310–2. <https://doi.org/10.1093/bioinformatics/bty539>.
31. Mageiros LMG, Bayliss SC, Pensar J, Pascoe B, Mourkas E, Calland JK, et al. Genome evolution and emergence of pathogenicity in avian *Escherichia coli*. *Nat Commun.* 2021;12:1–13
32. Lees JA, Vehkala M, Välimäki N, Harris SR, Chewapreecha C, Croucher NJ, et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nat Commun.* 2016;7(1):12797. <https://doi.org/10.1038/ncomms12797>.
33. Lees JA, Mai TT, Galardini M, Wheeler NE, Horsfield ST, Parkhill J, et al. Improved prediction of bacterial genotype-phenotype associations using interpretable pangenome-spanning regressions. *mBio.* 2020;11:e01344–20.
34. Sheppard SK, Guttman DS, Fitzgerald JR. Population genomics of bacterial host adaptation. *Nat Rev Genet.* 2018;19(9):549–65. <https://doi.org/10.1038/s41576-018-0032-z>.
35. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol.* 2006;2:2006.0008.
36. de Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, Cruaud C, et al. A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. *Mol Syst Biol.* 2008;4(1):174. <https://doi.org/10.1038/msb.2008.10>.
37. Porwollik S, Santiviago CA, Cheng P, Long F, Desai P, Fredlund J, et al. Defined single-gene and multi-gene deletion mutant collections in *Salmonella enterica* vs *typhimurium*. *Plos One.* 2014;9(7):e99820. <https://doi.org/10.1371/journal.pone.0099820>.
38. Koo BM, Kritikos G, Farelli JD, Todor H, Tong K, Kimsey H, et al. Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis*. *Cell Syst.* 2017;4(3):291–305 e297. <https://doi.org/10.1016/j.cels.2016.12.013>.
39. Kokes M, Dunn JD, Granek JA, Nguyen BD, Barker JR, Valdivia RH, et al. Integrating chemical mutagenesis and whole-genome sequencing as a platform for forward and reverse genetic analysis of *Chlamydia*. *Cell Host Microbe.* 2015;17(5):716–25. <https://doi.org/10.1016/j.chom.2015.03.014>.
40. Remmele CW, Xian Y, Albrecht M, Faulstich M, Fraunholz M, Heinrichs E, et al. Transcriptional landscape and essential genes of *Neisseria gonorrhoeae*. *Nucleic Acids Res.* 2014;42(16):10579–95. <https://doi.org/10.1093/nar/gku762>.
41. Goodall ECA, Robinson A, Johnston IG, Jabbari S, Turner KA, Cunningham AF, et al. The essential genome of *Escherichia coli* K-12. *mBio.* 2018;9:e02096–17.
42. Poulsen BE, Yang R, Clatworthy AE, White T, Osmulski SJ, Li L, et al. Defining the core essential genome of *Pseudomonas aeruginosa*. *PNAS.* 2019;116(20):10072–80. <https://doi.org/10.1073/pnas.1900570116>.
43. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio.* 2015;6:e00306–15.
44. van Opijnen T, Dedrick S, Bento J. Strain dependent genetic networks for antibiotic-sensitivity in a bacterial pathogen with a large pan-genome. *Plos Path.* 2016;12(9):e1005869. <https://doi.org/10.1371/journal.ppat.1005869>.
45. Price MN, Wetmore KM, Waters RJ, Callaghan M, Ray J, Liu H, et al. Mutant phenotypes for thousands of bacterial genes of unknown function. *Nature.* 2018;557(7706):503–9. <https://doi.org/10.1038/s41586-018-0124-0>.
46. Carey AF, Rock JM, Krieger IV, Chase MR, Fernandez-Suarez M, Gagneux S, et al. TnSeq of *Mycobacterium tuberculosis* clinical isolates reveals strain-specific antibiotic liabilities. *Plos Path.* 2018;14(3):e1006939. <https://doi.org/10.1371/journal.ppat.1006939>.
47. van Opijnen T, Camilli A. A fine scale phenotype-genotype virulence map of a bacterial pathogen. *Genome Res.* 2012;22(12):2541–51. <https://doi.org/10.1101/gr.137430.112>.
48. Fu Y, Waldor MK, Mekalanos JJ. Tn-Seq analysis of *Vibrio cholerae* intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. *Cell Host Microbe.* 2013;14(6):652–63. <https://doi.org/10.1016/j.chom.2013.11.001>.
49. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *PNAS.* 2015;112(13):4110–5. <https://doi.org/10.1073/pnas.1419677112>.
50. Duncan MC, Gillette RK, Maglasang MA, Corn EA, Tai AK, Lazinski DW, et al. High-throughput analysis of gene function in the bacterial predator *Bdellovibrio bacteriovorus*. *mBio.* 2019;10:e01040–19.
51. Rowe HM, Karlsson E, Echlin H, Chang T-C, Wang L, van Opijnen T, et al. Bacterial factors required for transmission of *Streptococcus pneumoniae* in mammalian hosts. *Cell Host Microbe.* 2019;25:884–891.e886.
52. Fenton AK, El Mortaji L, Lau DTC, Rudner DZ, Bernhardt TG. CozE is a member of the MreCD complex that directs cell elongation in *Streptococcus pneumoniae*. *Nat Microbiol.* 2016;2:16237.
53. Fenton AK, Manuse S, Flores-Kim J, Garcia PS, Mercy C, Grangeasse C, et al. Phosphorylation-dependent activation of the cell wall synthase PBP2a in *Streptococcus pneumoniae* by MacP. *PNAS.* 2018;115(11):2812–7. <https://doi.org/10.1073/pnas.1715218115>.
54. Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Collier JA, et al. The essential genome of a bacterium. *Mol Syst Biol.* 2011;7(1):528. <https://doi.org/10.1038/msb.2011.58>.
55. Wang H, Claveau D, Vaillancourt JP, Roemer T, Meredith TC. High-frequency transposition for determining antibacterial mode of action. *Nat Chem Biol.* 2011;7(10):720–9. <https://doi.org/10.1038/nchembio.643>.
56. Santiago M, Lee W, Fayad AA, Coe KA, Rajagopal M, Do T, et al. Genome-wide mutant profiling predicts the mechanism of a lipid II binding antibiotic. *Nat Chem Biol.* 2018;14(6):601–8. <https://doi.org/10.1038/s41589-018-0041-4>.
57. Coe KA, Lee W, Stone MC, Komazin-Meredith G, Meredith TC, Grad YH, et al. Multi-strain Tn-Seq reveals common daptomycin resistance determinants in *Staphylococcus aureus*. *PLoS Pathog.* 2019;15(11):e1007862. <https://doi.org/10.1371/journal.ppat.1007862>.
58. Todor H, Silvis MR, Osadnik H, Gross CA. Bacterial CRISPR screens for gene function. *Curr Opin Microbiol.* 2021;59:102–9. <https://doi.org/10.1016/j.mib.2020.11.005>.

59. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*. 2012; 482(7385):331–8. <https://doi.org/10.1038/nature10886>.
60. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013;8(11):2180–96. <https://doi.org/10.1038/nprot.2013.132>.
61. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell*. 2016;165(6):1493–506. <https://doi.org/10.1016/j.cell.2016.05.003>.
62. Wang T, Guan C, Guo J, Liu B, Wu Y, Xie Z, et al. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat Commun*. 2018;9(1):2475. <https://doi.org/10.1038/s41467-018-04899-x>.
63. Lee HH, Ostrov N, Wong BG, Gold MA, Khalil AS, Church GM. Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat Microbiol*. 2019;4(7):1105–13. <https://doi.org/10.1038/s41564-019-0423-8>.
64. Liu X, Kimmey JM, Matarazzo L, de Bakker V, Van Maele L, Sirard J-C, et al. Exploration of bacterial bottlenecks and *Streptococcus pneumoniae* pathogenesis by CRISPRi-Seq. *Cell Host Microbe*. 2020;29:107–120.e106.
65. Hawkins JS, Silvis MR, Koo B-M, Peters JM, Osadnik H, Jost M, et al. Mismatch-CRISPRi reveals the co-varying expression-fitness relationships of essential genes in *Escherichia coli* and *Bacillus subtilis*. *Cell Syst*. 2020;11:523–535.e529.
66. Jiang W, Oikonomou P, Tavazoie S. Comprehensive genome-wide perturbations via CRISPR adaptation reveal complex genetics of antibiotic sensitivity. *Cell*. 2020;180:1002–1017.e1031.
67. Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A. Gene fitness landscapes of *Vibrio cholerae* at important stages of its life cycle. *Plos Path*. 2013;9(12):e1003800. <https://doi.org/10.1371/journal.ppat.1003800>.
68. Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, et al. General and condition-specific essential functions of *Pseudomonas aeruginosa*. *PNAS*. 2015;112(16):5189–94. <https://doi.org/10.1073/pnas.1422186112>.
69. Burby PE, Nye TM, Schroeder JW, Simmons LA. Implementation and data analysis of Tn-seq, whole-genome resequencing, and single-molecule real-time sequencing for bacterial genetics. *J Bacteriol*. 2017;199:e00560–16.
70. Koch R: Über bakteriologische Forschung. In: Verhandlungen des X. Internationalen Medizinischen Kongresses, Berlin 1890. Volume 1. Berlin: Verlag von August Hirschwald; 1892.
71. Falkow S. Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis*. 1988;10(Suppl 2):S274–6. https://doi.org/10.1093/cid/10.Supplement_2.S274.
72. Falkow S. Molecular Koch's postulates applied to bacterial pathogenicity—a personal recollection 15 years later. *Nat Rev Microbiol*. 2004;2(1):67–72. <https://doi.org/10.1038/nrmicro799>.
73. Evans AS. Causation and disease: the Henle-Koch postulates revisited. *Yale J Biol Med*. 1976;49(2):175–95.
74. Fredericks DN, Relman DA. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev*. 1996;9(1):18–33. <https://doi.org/10.1128/CMR.9.1.18>.
75. Gradmann C. A spirit of scientific rigour: Koch's postulates in twentieth-century medicine. *Microb Infect*. 2014;16(11): 885–92. <https://doi.org/10.1016/j.micinf.2014.08.012>.
76. Byrd AL, Segre JA. Adapting Koch's postulates. *Science*. 2016;351(6270):224–6. <https://doi.org/10.1126/science.aad6753>.
77. Mourkas E, Taylor AJ, Méric G, Bayliss SC, Pascoe B, Mageiros L, et al. Agricultural intensification and the evolution of host specialism in the enteric pathogen *Campylobacter jejuni*. *PNAS*. 2020;117(20):11018–28. <https://doi.org/10.1073/pnas.1917168117>.
78. Thomason LC, Sawitzke JA, Li X, Costantino N, Court DL. Recombineering: genetic engineering in Bacteria using homologous recombination. *Curr Protocols Mol Biol*. 2014;106:1.16.11–11.16.39.
79. Selle K, Barrangou R. Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends Microbiol*. 2015;23(4):225–32. <https://doi.org/10.1016/j.tim.2015.01.008>.
80. McClure EE, Chávez ASO, Shaw DK, Carlyon JA, Ganta RR, Noh SM, et al. Engineering of obligate intracellular bacteria: progress, challenges and paradigms. *Nat Rev Microbiol*. 2017;15(9):544–58. <https://doi.org/10.1038/nrmicro.2017.59>.
81. Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, et al. A genetic resource for rapid and comprehensive phenotype screening of nonessential *Staphylococcus aureus* genes. *mBio*. 2013;4:e00537–12.
82. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*. 1989;77(1):51–9. [https://doi.org/10.1016/0378-1119\(89\)90358-2](https://doi.org/10.1016/0378-1119(89)90358-2).
83. Arnold BJ, Gutmann MU, Grad YH, Sheppard SK, Corander J, Lipsitch M, et al. Weak epistasis may drive adaptation in recombining bacteria. *Genetics*. 2018;208(3):1247–60. <https://doi.org/10.1534/genetics.117.300662>.
84. Yokoyama M, Stevens E, Laabei M, Bacon L, Heesom K, Bayliss S, et al. Epistasis analysis uncovers hidden antibiotic resistance-associated fitness costs hampering the evolution of MRSA. *Genome Biol*. 2018;19(1):94. <https://doi.org/10.1186/s13059-018-1469-2>.
85. Bronstein I, Fortin J, Stanley PE, Stewart GS, Kricka LJ. Chemiluminescent and bioluminescent reporter gene assays. *Anal Biochem*. 1994;219(2):169–81. <https://doi.org/10.1006/abio.1994.1254>.
86. Nolan T, Hands RE, Bustin SA. Quantification of mRNA using real-time RT-PCR. *Nat Protoc*. 2006;1(3):1559–82. <https://doi.org/10.1038/nprot.2006.236>.
87. Croucher NJ, Thomson NR. Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol*. 2010;13(5):619–24. <https://doi.org/10.1016/j.mib.2010.09.009>.
88. Sorek R, Cossart P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet*. 2010;11(1):9–16. <https://doi.org/10.1038/nrg2695>.
89. Saliba A-E, C Santos S, Vogel J: New RNA-seq approaches for the study of bacterial pathogens. *Curr Opin Microbiol* 2017;35:78–87. doi: <https://doi.org/10.1016/j.mib.2017.01.001>.
90. Lin J-S, Lai E-M. Protein–Protein Interactions: Co-Immunoprecipitation. In: Journet L, Cascales E, editors. *Bacterial Protein Secretion Systems: Methods and Protocols*. New York: Springer New York; 2017. p. 211–9.
91. Karimova G, Pidoux J, Ullmann A, Ladant D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *PNAS*. 1998;95(10):5752–6. <https://doi.org/10.1073/pnas.95.10.5752>.
92. Karimova G, Gauliard E, Davi M, Ouellette SP, Ladant D. Protein–Protein Interaction: Bacterial Two-Hybrid. In: Journet L, Cascales E, editors. *Bacterial Protein Secretion Systems: Methods and Protocols*. New York: Springer New York; 2017. p. 159–76.

93. Galas DJ, Schmitz A. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* 1978;5(9):3157–70. <https://doi.org/10.1093/nar/5.9.3157>.
94. Vierstra J, Stamatoyannopoulos JA. Genomic footprinting. *Nat Methods.* 2016;13(3):213–21. <https://doi.org/10.1038/nmeth.3768>.
95. Wade JT, Struhl K, Busby SJW, Grainger DC. Genomic analysis of protein–DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol Microbiol.* 2007;65(1):21–6. <https://doi.org/10.1111/j.1365-2958.2007.05781.x>.
96. Bonocora RP, Wade JT. ChIP-Seq For Genome-Scale Analysis Of Bacterial DNA-Binding Proteins. In: Artsimovitch I, Santangelo TJ, editors. *Bacterial Transcriptional Control: Methods and Protocols.* New York: Springer New York; 2015. p. 327–40.
97. Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc.* 2007;2(8):1849–61. <https://doi.org/10.1038/nprot.2007.249>.
98. Phillips GJ. Green fluorescent protein – a bright idea for the study of bacterial protein localization. *FEMS Microbiol Lett.* 2001;204(1):9–18. <https://doi.org/10.1111/j.1574-6968.2001.tb10854.x>.
99. Lippincott-Schwartz J, Patterson GH. Development and use of fluorescent protein markers in living cells. *Science.* 2003; 300(5616):87–91. <https://doi.org/10.1126/science.1082520>.
100. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat Methods.* 2005;2(12):905–9. <https://doi.org/10.1038/nmeth819>.
101. Wang S, Moffitt JR, Dempsey GT, Xie XS, Zhuang X. Characterization and development of photoactivatable fluorescent proteins for single-molecule–based superresolution imaging. *PNAS.* 2014;111(23):8452–7. <https://doi.org/10.1073/pnas.1406593111>.
102. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, et al. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem Biol.* 2008;3(6):373–82. <https://doi.org/10.1021/cb800025k>.
103. Gautier A, Juillerat A, Heinis C, Corrêa IR, Kindermann M, Beauflis F, et al. An engineered protein tag for multiprotein labeling in living cells. *Chem Biol.* 2008;15(2):128–36. <https://doi.org/10.1016/j.chembiol.2008.01.007>.
104. Shi Y. A glimpse of structural biology through X-ray crystallography. *Cell.* 2014;159(5):995–1014. <https://doi.org/10.1016/j.cell.2014.10.051>.
105. Montelione GT, Zheng D, Huang YJ, Gunsalus KC, Szyperski T. Protein NMR spectroscopy in structural genomics. *Nat Struct Biol.* 2000;7:982–5. <https://doi.org/10.1038/80768>.
106. Fernandez-Leiro R, Scheres SHW. Unravelling biological macromolecules with cryo-electron microscopy. *Nature.* 2016; 537(7620):339–46. <https://doi.org/10.1038/nature19948>.
107. Schuster CF, Wiedemann DM, Kirsebom FCM, Santiago M, Walker S, Gründling A. High-throughput transposon sequencing highlights the cell wall as an important barrier for osmotic stress in methicillin resistant *Staphylococcus aureus* and underlines a tailored response to different osmotic stressors. *Mol Microbiol.* 2020;113(4):699–717. <https://doi.org/10.1111/mmi.14433>.
108. Meeske AJ, Rodrigues CDA, Brady J, Lim HC, Bernhardt TG, Rudner DZ. High-throughput genetic screens identify a large and diverse collection of new sporulation genes in *Bacillus subtilis*. *PLoS Biol.* 2016;14(1):e1002341. <https://doi.org/10.1371/journal.pbio.1002341>.
109. Duggan S, Laabei M, Alnahari AA, O'Brien EC, Lacey KA, Bacon L, et al. A small membrane stabilizing protein critical to the pathogenicity of *Staphylococcus aureus*. *Infect Immun.* 2020;88:e00162–20.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

