BACTERIAL GENETICS

Phenogenomics of *Mycobacterium abscessus*

Multidimensional phenotyping links sequencing data and bacterial phenotypic diversity to uncover crucial virulence pathways in the emerging pathogen *Mycobacterium abscessus*.

Iñaki Comas and Miguel Moreno-Molina

Identifying genomic determinants and their presence in circulating variants can have clinical and epidemiological applications (Fig. 1), and can help predict drug resistance in pathogens such as *Mycobacterium tuberculosis* at a large scale[1]. The unprecedented expansion of genomic sequencing over recent decades has not yet been matched by large-scale phenotypic diversity datasets, hampering our ability to link pathogen diversity with phenotypes for relevant traits such as drug resistance and virulence. Reporting in *Nature Microbiology*, Boeck and colleagues[2] combine genomic and functional approaches to phenotypically characterize clinical *Mycobacterium abscessus* isolates and reveal key pathobiology determinants.

Traditional functional genomics approaches rely on in vitro generated mutant libraries of a reference strain that are used to link genotypes to phenotypes. Transposon mutagenesis, or CRISPR silencing-based libraries, allow the identification of domains, genes and pathways that interact to produce a phenotype. However, these approaches are limited by the use of reference strains and fail to integrate the genetic and phenotypic diversity of the 'wild' bacterial population in order to understand how traits are regulated in different strains of the same pathogen. Recent work has shown the benefits of combining high-throughput functional information and genomic data from clinical isolates to narrow down genomic determinants of antibiotic resistance and virulence[3–5]. As an alternative, large datasets of clinical strains can be sequenced, enabling the association between pathogen genomic diversity and phenotypic diversity when available. Bacterial genome-wide association studies (GWAS) are also useful, but present many limitations even when corrected by population structure[6,7]. Evolutionary approaches, such as the identification of convergent evolutionary events[8] or signals of selection[9], have helped to pinpoint candidate genomic determinants that can be later associated with phenotypes of interest. Alternative approaches based on machine learning or structure-based predictions have
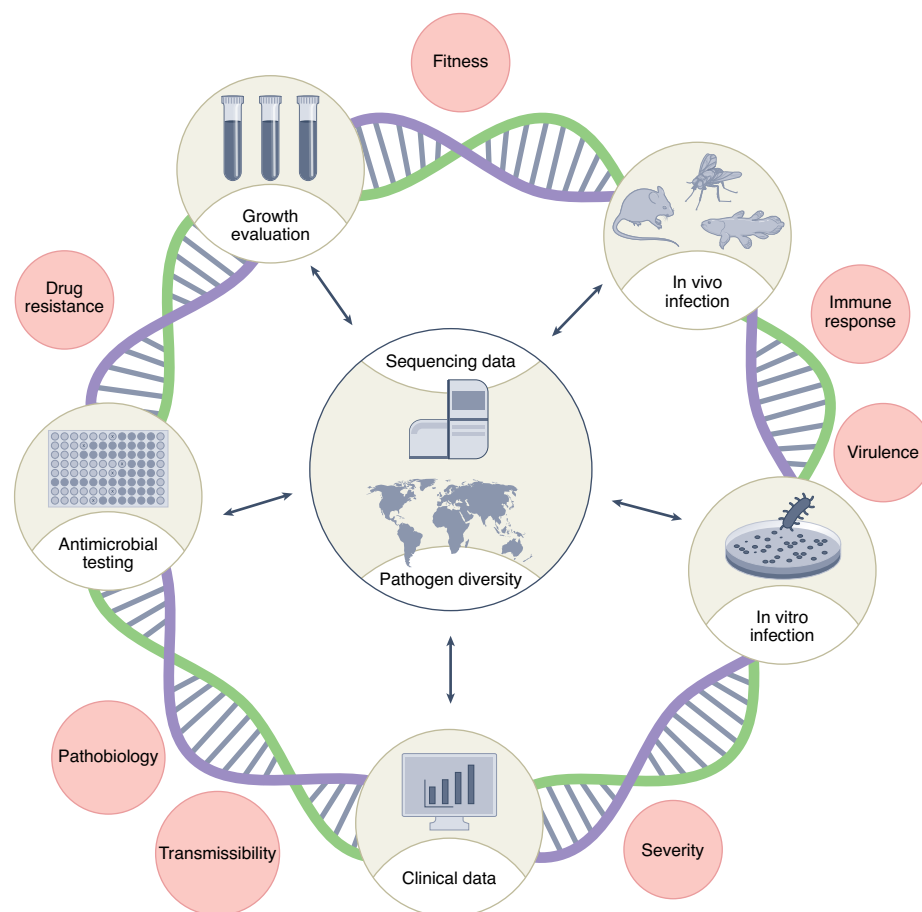
Fig. 1 | Integrating data for genotype–phenotype association and forecasting. A robust framework to associate genetic and phenotypic diversity in microbial pathogens needs a multidimensional and systematic approach to obtain good quality data. The integration of experimental observations and clinical data with variants obtained by whole-genome sequencing can give meaningful insights into the pathogen biology, improving patient prognosis and pointing out pathogen genomic determinants of transmissibility, severity or new drug targets with the goal of improving diagnosis, treatment and surveillance.

also been applied successfully[10,11]. Variation in bacteria is also defined by a diverse gene content across strains of the same species, known as the pangenome, and there are ways to link gene content to phenotypes. All of these approaches have intrinsic limitations, the most obvious one is that many traits are the result of the interaction of different mutations, a phenomenon called epistasis. Thus, individual approaches are limited to reveal the whole range of relevant genomic variants in a microbial pathogen even when large-scale and diverse associated phenotypic datasets exist, which is rarely the case.

Boeck and colleagues combined genomic, phenotypic and GWAS approaches to phenotypically characterize a large collection

of 331 clinical *M. abscessus* isolates and to link those phenotypes to causal variants. This phenotyping was performed across five dimensions: planktonic growth using different carbon sources, antimicrobial resistance across several time points, in vitro infection of human macrophages, in vivo infection of *Drosophila melanogaster* and clinical outcomes of patients. Using correlation analysis, they revealed three phenotypic clusters with different virulence characteristics. One of these groups showed the fastest growth in culture and high mortality in both the macrophage and fly models, while another group had the opposite features. A third group had intermediate characteristics and was associated with the best clinical outcomes. As these three groups were independent of colony morphotype, subspecies and levels of macrolide resistance, Boeck and colleagues argue that they could represent different evolutionary trajectories and highlight the importance of assessing multiple phenotypic characteristics for patient prognosis.

The authors next deployed a GWAS analysis to delve into the genetic basis of the variation in *M. abscessus* pathobiology. Using corrected and uncorrected models for population structure, they identified previously known genetic resistance-related determinants, as well as other hits such as mycobactin synthesis genes that are responsible for intracellular iron intake. The authors performed proteome-wide computational structural modelling to assess the impact of non-synonymous variants, and found that one particular single-nucleotide polymorphism in the mycobactin polyketide synthetase (*MbtD*) gene was predicted to result in a loss of protein function, thus affecting bacterial access to iron during infection. They experimentally confirmed this with a *MbtD* knockout mutant that showed decreased growth in the macrophage infection model, showcasing this gene as a potential therapeutic target. The authors also explored potential epistatic interactions at the genome-wide scale to discover genes that might have co-evolved

and to potentially uncover functionally linked protein networks. By applying correlation-compressed direct coupling analysis, they identified co-selection and established highly connected clusters, such as the mammalian cell entry gene family, genes involved in secretion systems or, again, the mycobactin synthesis genes previously flagged by the structural-guided GWAS analysis. To integrate the results of all previous analyses, the authors tested the effect of two homoplastic variants — a deletion in the *MAB_0471* secretion system component, and a non-synonymous single-nucleotide polymorphism in the *MAB_3317c* non-ribosomal peptide synthase — in vivo in a *Drosophila* infection model. Both variants increased the survival of infected flies and were associated with more persistent clinical infection, suggesting a meaningful role of those two genes in *M. abscessus* virulence regulation.

Defining adequate phenotypes is not always straightforward when applying genotype–phenotype diversity mapping, and it is not clear which in vitro and in vivo models are more adequate for many pathogens. Boeck and colleagues' work is a great example of how strains can be grouped into clinically relevant clades when combining different relevant in vitro and in vivo models. However, models cancel noise at the expense of reducing the complexity of real-world observations. Relevant phenotypes in clinical settings are multifactorial and rarely have correlates in the laboratory and thus need to be measured in the clinical setting. Recent studies combine paired microbial and host genomics to clinical data[9], but linking pathogen diversity to clinical and epidemiological phenotypes is probably the last and more challenging frontier in genotype–phenotype diversity mapping.

In the past two years, predictions for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pathogenicity have advanced more than for any other pathogen. Deep mutation scanning coupled with antibody neutralization assays has allowed us to predict mutations associated

with antibody escape[12]. Epidemiological and phylodynamic modelling have helped to predict the epidemiological fitness of variants and of individual residues[11]. For variants such as Omicron or Delta, researchers and public-health agencies have been able to predict and measure their impact on complex traits such as transmissibility, severity, immune responses or vaccine effectiveness.

The above is an unmet challenge for other pathogens with similar genome sizes, and for those with much bigger and more complex genomes, such as bacteria. The capacity to link variants to complex biological and clinical phenotypes in real time as shown for SARS-CoV-2 and by Boeck and colleagues for *M. abscessus* maps out one future direction for epidemic control. Importantly, it is also a blueprint for researchers to incorporate clinical and epidemiological data for other pathogens.

We believe that the work by Boeck and colleagues heralds the dawn of a large-scale genotype–phenotype diversity mapping era in microbial genomics. ❐

Iñaki Comas [ID][1,2][✉] and
Miguel Moreno-Molina [ID][1]

¹Instituto de Biomedicina de Valencia IBV-CSIC, Valencia, Spain. ²CIBER in Epidemiology and Public Health, Madrid, Spain.
✉e-mail: *icomas@ibv.csic.es*

References
1. Walker, T. M. et al. *Lancet Microbe* **3**, e265–e273 (2022).
2. Boeck, L. et al. *Nat. Microbiol.* https://doi.org/10.1038/s41564-022-01204-x (2022).
3. Furió, V. et al. *Commun. Biol.* **4**, 1322 (2021).
4. Li, S. et al. *Nat. Microbiol.* **7**, 766–779 (2022).
5. Liu, X. et al. *Cell Host Microbe* **29**, 107–120.e6 (2021).
6. Earle, S. G. et al. *Nat. Microbiol.* **1**, 16041 (2016).
7. Didelot, X. *Methods Mol. Biol.* **2242**, 205–220 (2021).
8. Chiner-Oms, Á., López, M. G., Moreno-Molina, M., Furió, V. & Comas, I. *Proc. Natl Acad. Sci. USA* **119**, e2113600119 (2022).
9. Lees, J. A. et al. *Nat. Commun.* **10**, 2176 (2019).
10. Goodswen, S. J. et al. *FEMS Microbiol. Rev.* **45**, fuab015 (2021).
11. Starr, T. N. et al. *Science* **371**, 850–854 (2021).
12. Obermeyer, F. et al. *Science* **376**, 1327–1332 (2022).