

PERSPECTIVE

For reprint orders, please contact: reprints@futuremedicine.com

Metabolic competition as a driver of bacterial population structure

Eleanor R Watkins^{*1}, Martin CJ Maiden¹ & Sunetra Gupta¹

Understanding the processes whereby diversity arises and is maintained in pathogen populations is pivotal for designing disease control interventions. A particular problem is the maintenance of strain structure in bacterial pathogen populations despite frequent genetic exchange. Although several theoretical frameworks have been put forward to explain this widespread phenomenon, few have focused on the role of genes encoding metabolic functions, despite an increasing recognition of their importance in pathogenesis and transmission. In this article, we review the literature for evidence of metabolic niches within the host and discuss theoretical frameworks which examine ecological interactions between metabolic genes. We contend that metabolic competition is an important phenomenon which contributes to the maintenance of population structure and diversity of many bacterial pathogens.

First draft submitted: 27 April 2016; Accepted for publication: 14 July 2016;

Published online: 23 September 2016

Bacterial population structure

Populations of many bacterial pathogens exhibit high levels of genetic diversity. Much of this variation, however, is structured into discrete 'strains' which persist for long periods despite frequent horizontal genetic transfer (HGT) both within and among species (**Box 1**) [1]. The definition of a strain varies widely depending upon the methods used for discrimination and the question being asked: for example, the variation in metabolic enzymes is structured among many bacteria into a limited number of electrophoretic types (ETs) on the basis of differences in electrophoretic mobilities (i.e., by multilocus enzyme electrophoresis or MLEE [2]). Classification within bacterial species is more typically performed by multilocus sequence typing (MLST) of fragments of metabolic house-keeping genes dispersed around the bacterial chromosome; the sequence types (STs) thus defined tend to be organized into clonal complexes [3]. It is commonly observed that only a subset of clonal complexes is significantly associated with a heightened ability to cause invasive disease. Within the bacterial pathogen *Neisseria meningitidis* (the meningococcus), for example, a number of such hyper-invasive clonal complexes have been observed to persist over several decades and achieved global spread [4,5]. Strains can also be defined through their antigenic properties: the distinct serotypes of *Streptococcus pneumoniae* (the pneumococcus) provide one such example.

There are a number of theoretical frameworks that seek to explain the observed patterns of population level diversity in bacterial pathogen populations. Several of these assume that bacterial pathogen populations are structured primarily through neutral processes, with little or no selection occurring [6]. It has been demonstrated that purely neutral mutational drift and recombination are

KEYWORDS

- bacterial pathogens
- competition
- mathematical models
- metabolism • vaccination

¹Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK

^{*}Author for correspondence: eleanor.rose.watkins@gmail.com

Box 1. Glossary of terms.

- Antigenic type: classification based on variation in genes encoding targets of immunity
- Bacterial population structure: patterns and organization of genetic diversity observed among bacterial populations, over both temporal and spatial scales
- Clonal complex: a group of closely related sequence types, defined by multilocus sequence typing of fragments of metabolic housekeeping genes
- Horizontal genetic transfer: the transfer of genetic material from one organism to another, other than through vertical descent
- Metabolic competition: any mechanism through which bacteria compete for nutrients and energy sources within the host environment, where a lack of such substances inhibits growth
- Metabolic type: constellation of genes within a bacterium which function in the uptake of nutrients and metabolic processes
- Nonoverlapping associations: in a population, allelic variants predominantly associate such that a certain allelic variant of one gene associates uniquely with one particular allelic variant from another gene. This means that the most common strains have unique combinations of allelic variants which are not shared by others
- Overlapping associations: the allelic variants do not associate with each other in unique combinations, and pathogen strains overlap in their allelic repertoires
- Strain structure: the existence of groups of organisms sharing distinct phenotypes and genotypes within a population

not sufficient to account for the population structure of a number of bacterial species, including *N. meningitidis*, *S. pneumoniae* and *Staphylococcus aureus* [7]. A modified model which includes the effects of localized transmission, the neutral microepidemic model, is able to reproduce the observed population genetic structure of these pathogens, based on distributions of pairwise distances of MLST allelic data [7]. However, Jolley *et al.* used a coalescent-based approach to demonstrate that the number of meningococcal STs generated as expected through neutral processes alone significantly underestimated the number of STs observed in a sample of meningococci from the Czech Republic in 1993 [8]. Furthermore, the majority of allele combinations persisted for 1 year, and none persisted for more than 7 years, suggesting that purely neutral processes may not provide an adequate explanation for the patterns observed [9].

Non-neutral frameworks invoke competitive interactions among strains as the primary driver of bacterial population structure. In these models, strains within pathogen populations can compete directly within the host – for example, through depleting shared nutritional resources or producing antimicrobial compounds – or indirectly, through the host immune response. In the latter case, competition between pathogen strains can be mediated by immunological cross-protection: the degree to which infection by one strain prevents successful infection by another. Where the dominant immune response to a

pathogen is against a single variable antigenic determinant, strong variant-specific immune responses will maintain antigenic diversity in the population, whereas strong cross-protective responses to a number of pathogen strains will act to decrease the number of circulating antigenic types [10]. A balance between these two conflicting selection pressures has been shown to be important in generating observed levels of serotype diversity in *S. pneumoniae* [11,12]. When dominant immune responses target multiple antigenic determinants, high levels of variant-specific immunity can cause pathogen populations to segregate into discrete strains which do not share antigenic determinants [13]. A number of well-characterized examples of nonoverlapping antigenic determinants have been shown in *N. meningitidis*. The two variable regions of the outer membrane antigen PorA [10], in addition to the antigenic iron transporter FetA [5,14], have been observed to associate in discordant allelic groupings in samples from both carriage and invasive disease worldwide. PorA and, to a lesser extent, FetA are important vaccine components of a number of meningococcal vaccines which have been deployed in several countries over the past 20 years [15–17]. Variants of the *Neisseria* opacity-associated proteins, which are used in adhesion and invasion, also manifest a nonoverlapping pattern [18]. Group A Streptococci manifest nonoverlapping associations among variants of the surface-presented M protein and T protein antigens [19]. Within Group B Streptococci, the

expression of a number of immunodominant surface proteins associates in a nonoverlapping way with capsular serotype [20]. Similarly, coinfection by the bacterial pathogen *Anaplasma marginale*, a prevalent vector-borne pathogen of cattle, relies on the expression of nonoverlapping antigenic variants of the immunologically dominant *msp2* antigen [21].

Whereas most theoretical frameworks of pathogen population structure have modeled the effects of immune-mediated competition, few frameworks have considered the effects of direct short-term competition among distinct strains in the host environment. There is, however, *in vivo* experimental evidence as well as epidemiological data which suggest that such direct competition may play an important role in colonization dynamics in a number of distinct anatomical sites. For instance, longitudinal studies in Kenya and Denmark show that the rate of acquisition of pneumococcal strains in the nasopharynx is higher for noncarriers compared with carriers [22,23]. It has also been shown in mouse models that the carriage of vaccine-type strains of *S. pneumoniae* inhibits the subsequent colonization of particular nonvaccine type strains [24]. Indirect evidence of ecological competitive interactions can be found among organisms colonizing the human oral cavity: for example, clinical studies have found that patients colonized with *Streptococcus oligofermentans* have a reduced incidence of dental caries caused by *Streptococcus mutans* [25]. A large body of experimental work with enteric bacteria suggests that there is also intense competition within the human gut [26]. Resident microbial communities in the intestine are considered to be diverse and also stable in composition, and together act to protect against the invasion of other microorganisms, including many pathogens: a phenomenon referred to as colonization resistance [27]. This helps to explain the observation that healthy human volunteers are not able to be colonized by *Escherichia coli* strains isolated from their own feces, upon being fed these strains directly [28].

We propose that metabolic competition is an important component of direct ecological competition operating in bacterial pathogen populations. Although genetic variation in metabolic genes has conventionally been viewed as primarily neutral, an increasing body of evidence indicates that such variation has important functional and evolutionary consequences for bacterial populations.

Diversity of metabolic genes

The conventional paradigm of pathogen evolution proposes that, whereas antigen-encoding genes should display genetic variation as a result of diversifying selection to avoid recognition by the host immune response, the metabolic genes should display relatively little variation as they are under stabilizing selection for conservation of function [29]. However, although the core genomes of bacterial pathogens are primarily composed of metabolic genes [30], there is an increasing number of studies which show that metabolic and transport genes are also part of the variable accessory genome, and that they contribute significantly to the diversity observed across strains within bacterial populations [31,32].

Efficient bacterial replication is essential for colonization and transmission; thus understanding bacterial metabolism within the host is essential to understanding the spread of pathogens among hosts and host–pathogen interactions [33]. Differences in metabolic machinery may lead to differences in growth rates between strains – for example, if different strains assimilate resources at different rates. A number of experimental studies suggest highly significant links between metabolic genes and bacterial growth rates. For example, in a series of experiments by Helling, the lack of the glutamate-synthesizing enzyme glutamate dehydrogenase in *Escherichia coli* was shown at first to result in decreased glucose-limited growth, but was subsequently compensated for by mutations in the *ndh*, *cyo* and *cyd* genes, which code for NADH dehydrogenases and terminal oxidases used in ATP synthesis [34]. This indicates the existence of alternative metabolic pathways in *E. coli* with differing levels of efficiency and cost. In another study, Sabarly and colleagues assayed the growth yield of *E. coli* strains on 95 carbon sources and correlated their growth capacities with the presence/absence of enzyme-coding genes. They found that most of the variation in growth rates was explained by the presence/absence of metabolic pathways, and was largely independent of phylogeny [35].

In addition to its importance in colonization and transmission, bacterial *in vivo* metabolism is a fundamental aspect of virulence and pathogenesis [33]. Much evidence for the importance of metabolism in virulence and pathogenesis in a range of bacterial pathogens and host sites has accumulated in recent years [36–38]. Most bacterial pathogens encounter a variety of different environments within the host during infection,

and must adapt accordingly to make use of alternative nutrient sources: a phenomenon which has been termed ‘nutritional virulence’ [39]. Various experimental studies have shown differences in expression of metabolic genes within different host environments, including differences between commensal and invasive strains of *N. meningitidis*, *S. agalactiae* and *S. pneumoniae* in the nasopharynx and blood, and differences in sugar use between commensal and pathogenic *E. coli* strains [36,40–42]. Other screening and transcription studies have revealed a number of metabolic genes that are essential to virulence, including *S. pneumoniae*, *Mycobacterium tuberculosis*, *Helicobacter pylori*, *S. aureus*, *Vibrio cholera* and *Salmonella typhimurium*, amongst others (reviewed by [33,37]).

Mechanisms of metabolic competition

Ecological competition among bacteria involving the metabolic genes could occur through exploitative competition – whereby bacteria compete for limited resources but do not directly interact – or interference competition, whereby bacteria interact antagonistically. The results of a number of experimental studies suggest that interference competition occurs among several bacterial pathogens, through which bacteria produce harmful substances to inhibit the growth of others, such as secondary metabolites, or actively restrict or remove a nutrient from its competitors. Mechanisms of interference competition both within and among bacterial species have been well described elsewhere [33,43] and exploitative competition among pathogenic strains for metabolic resources shall be the focus of this article. Evidence is accumulating to suggest that there is a large variety of nutritional resources available to bacteria which colonize and invade mammalian hosts, with a particular wealth of information for colonization of the airway and intestine (Box 2). Through surveying the literature, we surmise three principal ways through which bacterial pathogens may experience competition for metabolic resources within the host. Many of the supporting studies cited here use *S. pneumoniae* and *E. coli*, as such processes in the host are particularly well studied for these pathogens.

First, two strains may not be able to co-infect a host if they utilize an identical repertoire of substrates; thus bacterial strains may evolve to occupy distinct metabolic niches by utilizing different resources (Figure 1A). Strain-specific

differences in the presence and expression of transporters of particular metabolic substrates have been shown in a number of studies of *S. pneumoniae* [38,49]. Linke *et al.* showed that strain-to-strain variation in the ability to utilize different lengths of fructooligosaccharide chains is determined by diversity at the *sus* transporter locus, with 60–79% of pneumococci able to utilize the fructooligosaccharide inulin [50]. Buckwalter and King noted strain-specific differences in the absence/presence of 12 distinct carbohydrate transporters across pneumococcal genomes, from seven independent studies [38]. The construction of pan genome-scale metabolic models for *E. coli* has also shown a large number of strain-specific differences in the number and functional classification of metabolic coding sequences and reactions [31]. The metabolic pan genome – constructed from the interrogation of 16 finished genomes – contained 79 ORFs and 32 reactions not present in the core metabolic genome, possibly facilitating the breakdown of alternative substrates. Indeed, the results of a series of experiments in mice suggest that distinct *E. coli* strains are able to use different nutrients for growth in the intestine. Different pathogenic and commensal *E. coli* strains each used a different repertoire of approximately 6–7 sugars to colonize the intestine, of the 18 that *E. coli* is capable of using *in vivo* [42,51–52]. This supports the idea that virulent, invasive *E. coli* strains are able to overcome colonization resistance by taking advantage of nutrients that are not used by resident commensal strains [51].

Second, bacterial strains may occupy distinct metabolic niches through binding different host structures in order to access the same nutritional resource; thus, even if two strains have identical substrate repertoires, they may be able to occupy distinct niches (Figure 1B). For example, variants of family 98 glycoside hydrolases involved in fucose utilization in *S. pneumoniae* bind distinct host carbohydrate antigens, showing selectivity for either Lewis or group A/B antigens [53]. Similarly, meningococci encode several surface receptors to acquire iron or heme from specific iron-binding proteins in the host, including hemoglobin, lactoferrin and transferrin, and there is a variable distribution of such receptors among meningococci. The hemoglobin receptor Hmbr, for example, is significantly over-represented in invasive isolates [54], and the iron transporter FetA is absent from a minority of strains [55].

Box 2. Nutrient sources in the mammalian gut and airway.

- The concentration of free sugars in the normal human airway is low [44], yet the human airway still provides a wide range of energy sources for bacterial pathogens, including the mucus which coats the epithelial layer of the respiratory tract. Many of the components of the host immune system secreted into mucus (including immunoglobulins, cytokines) are sources of sugars and amino acids. Mucin glycoproteins in mucus are also utilized by some bacteria, and dead respiratory cells potentially provide lipids, nucleic acids and proteins [45]. Many bacterial pathogens and commensals are also able to access sugars through the breakdown of a variety of diverse sugar structures and glycoconjugates which are attached to the surface of host epithelial cells and some are also able to lyse host cells directly [46]
- In contrast, there is an abundance of fermentable substrates available to enteric bacteria in the GI tract. In addition to nutrients originating from the host diet itself, multiple metabolic intermediates are formed during the breakdown of many dietary components, and protein can also be derived from host enzymes and other secretions [47]. As in the respiratory tract, mucus is a rich source of substrates, providing a variety of mucins, shed epithelial cells and several smaller metabolites. The outer mucus layer of the GI tract comprises mainly commensals which are able to cleave glycans or glycoconjugates from epithelial cells; many pathogens exploit the sugars thereupon released when passing through the mucus barrier to the epithelial cells [48]

Third, cocolonizing bacteria may have different rates of uptake for particular resources, and/or different utilization hierarchies for their energy sources (**Figure 1C**). Even if two strains utilize an identical repertoire of nutritional resources, ecological competition could be reduced if they require different amounts of particular substrates (e.g., if strain 1 required a large amount of substrate A but only a minimal amount of substrate B, and strain 2 required only a small amount of substrate A but a lot of substrate B). Bidossi *et al.* used a functional genomics approach to demonstrate strain-specific differences among pneumococci in the ability to absorb a wide range of carbohydrate resources [49]. As well as variable distributions in the specific uptake systems present, the relative intensity at which different substrates were fermented was observed to differ between strains. Furthermore, a transcriptional analysis by Pagliarulo *et al.* identified fourfold differences in expression of the *gdhA* gene between different hyperinvasive lineages of *N. meningitidis*, which is involved in ammonia assimilation [56]. In the context of the GI tract, the ‘nutrient niche’ hypothesis proposed by Freter [57] states that in order to colonize the intestine, each strain must use at least one limiting nutrient better than its competitors. In support of this, Fabich and colleagues found that *flhD* mutants of *E. coli* K-12 were superior colonizers of the mouse intestine by using the same sugars more efficiently than its wild-type parent, rather than through using different sugars [58].

In a series of analyses by Watkins *et al.* [59], allelic variants of the metabolic genes of 616 whole genomes of *S. pneumoniae* were

demonstrated to show a significantly higher level of association than variants of functional coding genes not belonging to metabolic processes, using linkage disequilibrium and mutual information metrics. These findings lend support to the hypothesis that the metabolic profile of pneumococci encodes a number of tightly linked and interacting proteins, and it is these successful constellation of alleles that allows them to exploit a particular metabolic niche.

Theoretical frameworks incorporating metabolic competition

The majority of theoretical frameworks of pathogen population structure to date have modeled the effects of immunity-mediated competition. Few have considered the effects of ecological, short-term competition between strains in the host environment, including competition at the metabolic genes.

Lipsitch assumed strong ecological competition among serotypes of *S. pneumoniae* in a model simulating the effects of vaccination, the results of which were consistent with available data [60]. In another model simulating pneumococcal vaccination, Zhang *et al.* investigated the effects of both direct competition operating during carriage in the nasopharynx in addition to antibody-mediated immunity [61]. The results showed that serotype replacement – a well-documented phenomenon in which nonvaccine serotypes increase at a population level following vaccination – is only observed when strong ecological competition occurs between strains.

The interplay between ecological competition manifested specifically as metabolic competition

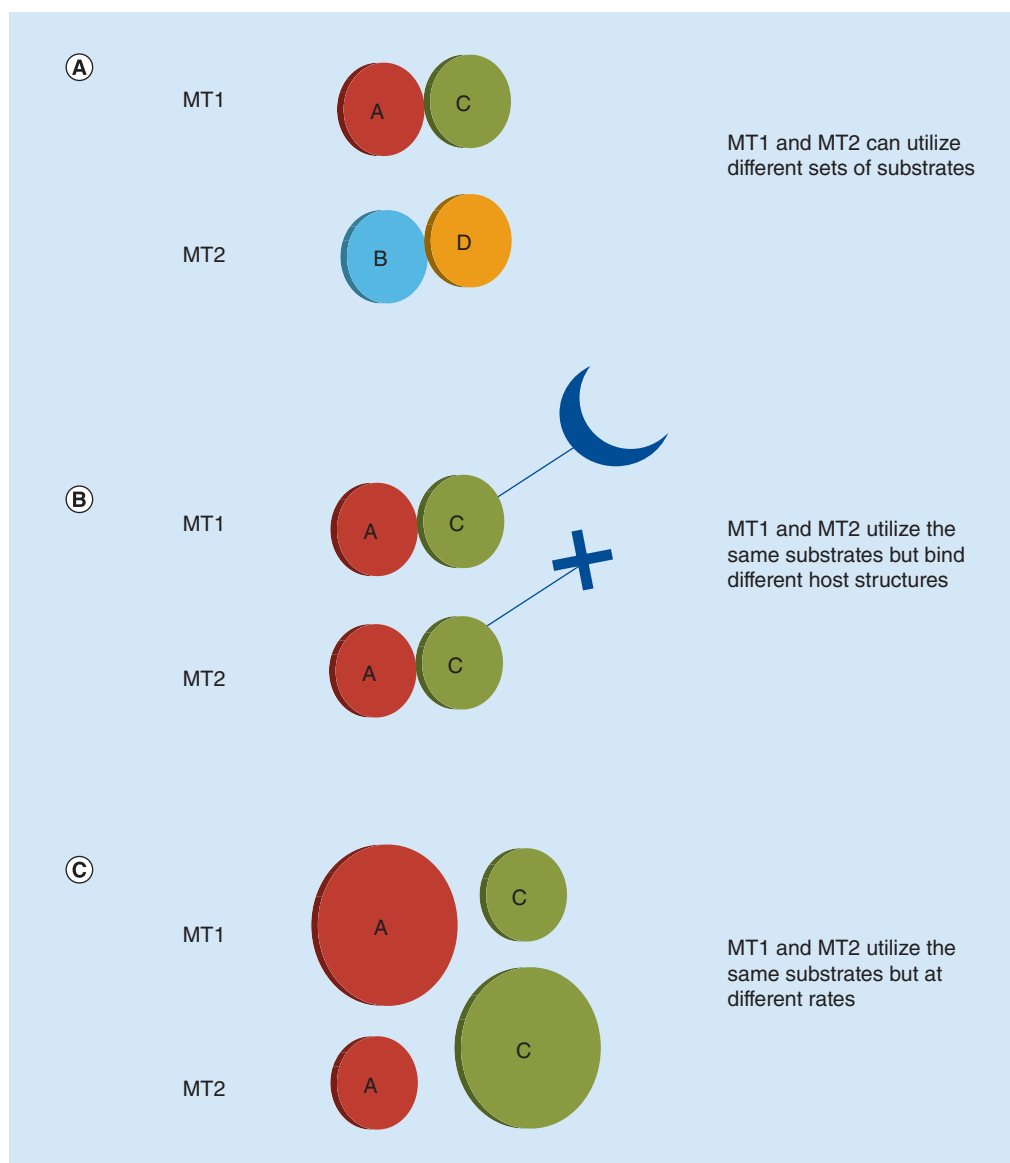


Figure 1. Ways through which bacterial strains with different metabolic profiles, or metabolic types (here MT1 and MT2) are able to reduce ecological competition and co-infect a host. (A) Bacterial strains MT1 and MT2 reduce metabolic competition through targeting different sets of resources. (B) MT1 and MT2 utilise the same substrates, but access these resources via distinct host binding sites. (C) MT1 and MT2 have different rates of uptake and/or utilisation for the same substrates.

and immunity-mediated competition was investigated within a multilocus framework by Buckee *et al.* [9], using an agent-based model. A key finding was that, at high levels of competition, high frequency strains in bacterial populations should show nonoverlapping associations between alleles encoding antigen-encoding and metabolic genes; in other words, a given metabolic profile should primarily associate with a unique antigenic profile. Additionally, such associations should be

stable over long time periods in spite of frequent HGT. The rationale behind these nonoverlapping associations is that the successful and widespread strains have been segregated by selection into different antigenic and metabolic niches in order to minimize the effects of immunological competition (for new hosts) and metabolic competition (for nutritional resources within the host). Such segregation is consistent with one of the earliest axioms of community ecology, Gause's Principle,

which dictates that no two species can occupy the same niche as the superior competitor species would exclude the other [62]. These predictions were upheld in a series of deterministic models [59] in which the strengths of immunological and metabolic competition were also varied over a large range of parameter space; strong nonoverlapping associations were found at relatively low levels of competition, suggesting that even weak levels of competition may be important in structuring bacterial pathogen populations.

Associations between antigenic type and MLST-defined ST/clonal complex, or MLEE-defined ET are indeed found across a range of bacterial pathogens (Table 1). Intriguingly, these associations between antigenic type and MLST-type or MLEE-type appear to be stable over long time periods in a number of pathogens. Among *N. meningitidis* for example, identical associations between clonal complex and several outer membrane antigens (including a number of vaccine candidates, such as PorA, PorB, FetA, fHbp, NHBA and NadA) have been recorded over several decades (the longest period recorded being 74 years) [5,14,63–65]. Identical associations between serotype and ST in *S. pneumoniae* have also been observed over many years (Table 2).

MLST loci comprise fragments of genes that encode metabolic processes and are located around the chromosome, but can variation at these seven loci alone define a metabolic niche? To answer this question, associations between metabolic genes and antigen-encoding genes were explored at the whole genome level by Watkins *et al.* in 616 genomes of *S. pneumoniae* [59]. The metabolic profiles of pneumococcal strains, comprising all metabolic and transport loci identified in the genome, were highly consistent within a capsular serotype, and significantly different between serotypes. They also found that alleles of metabolic and transport genes were highly consistent within each MLST-defined ST, suggesting that STs are a proxy for the metabolic type among pneumococci.

There are several alternative theories which can be put forward to explain the observed associations between antigenic and metabolic types. These, along with the corresponding counter arguments, are presented in Table 3. None of these theories individually are able to account for the observations presented here; however, the processes and selection pressures discussed are not mutually exclusive, and it is likely that a combination of neutral, selective

and mechanistic factors play important roles in the maintenance and diversity of bacterial pathogen structure. For example, the propagation of persistent multilocus associations over time and space is a hallmark of clonal populations, in which HGT is too rare to break the prevalent pattern of clonal population structure. Clonal descent will undoubtedly have contributed to the observations of associations between antigenic and metabolic types. Such an explanation is especially pertinent to those pathogens which do not experience frequent HGT. Thus, although STs and OmpC antigen of *Borrelia lusitanae* show nonoverlapping associations, this may be linked to the fact that HGT across the chromosome is relatively rare in this species [101]. However, all the bacterial species presented in Table 1 have been shown in previous studies to manifest frequent rates of HGT, which should homogenize the associations between antigenic type and metabolic type.

The thousands of combinations of antigenic and ST found among bacterial pathogens reflects the large amounts of diversity generated by HGT in these populations [5]. Out of the large pool of genetic variants only a subset of successful genotypes with nonoverlapping repertoires of these alleles are able to propagate to high frequencies (Figure 2). The frameworks of Buckee *et al.* and Watkins *et al.* assume that all strain combinations are continuously generated in the population at low levels, but the prevalent strains which emerge manifest a nonoverlapping pattern. It is this fitness advantage, of minimized competition for hosts and metabolic resources, which may permit them to persist for many years and sometimes disseminate on a global scale [5]. The low frequency variants generated by HGT which show overlapping metabolic and antigenic variants are therefore at a fitness disadvantage with the dominant nonoverlapping variants. There are a few pathogens, such as *Leptospira* species and *Legionella pneumophila*, for which relevant typing data are available, that do not seem to manifest nonoverlapping associations between antigenic and metabolic type [106,107]. Other pathogens, such as *Pseudomonas aeruginosa*, and *S. agalactiae*, show inconsistent evidence of associations [108–110]. Perhaps there is not sufficient host immune selection pressure created by the antigens characterized in the studies to structure the population into distinct antigenic groups, or sufficiently intense competition at the metabolic genes. It is also possible that although immune

Table 1. Associations among antigen-encoding and metabolic genes for 15 bacterial pathogens, for which frequent horizontal genetic transfer has been documented.

| Pathogen | Common disease(s) | Antigenic type | Metabolic type | Details | isolates | Ref. |
|-------------------------------|---|---|----------------------|--|-------------|----------|
| <i>Campylobacter jejuni</i> | Gastroenteritis | PorA (outer membrane protein) | Clonal complex | PorA variants had a nonrandom association with clonal complex ($p < 0.01$). Nine PorA variants (each occurring ten-times or more) were associated predominantly with 11 clonal complexes [66] | 584 | [67] |
| <i>Clostridium difficile</i> | Gastroenteritis | SlpA | ST | Groupings by MLST and slpA had a high concordance [68] | 42 | [69] |
| <i>Enterococcus faecium</i> | Gastroenteritis | Ace, salA, Isa (collagen/laminin adhesin; a cell wall-associated antigen; putative ABC transporter) | ST | 98% (49 out of 50) of isolates showed concordance between a trilocus typing scheme (based on two antigens and one antibiotic resistance locus) and traditional MLST [70] | 50 | [71] |
| <i>Escherichia coli</i> | Gastroenteritis | O antigen (lipopolysaccharide) | ET | Good correlation between O group and ET; the majority of isolates of an ET expressed a common O antigen [72] | 187 | [73] |
| <i>Haemophilus influenzae</i> | Bacteremia, meningitis, cellulitis, epiglottitis, septic arthritis, pneumonia | Serotype Serotype | ET Clonal complex | No ETs were shared among isolates of different serotypes [74] On a minimum evolution tree of concatenated MLST nucleotide data, monophyletic groups were formed according to several serotypes [76] | 2209 131 | [75] |
| <i>Klebsiella pneumoniae</i> | Opportunistic infections (e.g., pneumonia, osteomyelitis, meningitis) | Serotype (C pattern) | ST | High concordance between molecular serotype and STs [77] | 63 | [78] |
| <i>Listeria monocytogenes</i> | Listeriosis | Serotype (somatic [O] antigen and flagellar [H] antigen) | ET | MLEE data divided the isolates into two divisions corresponding to antigenic groups I and II [79] | 175 | [80] |
| <i>Neisseria gonorrhoeae</i> | Gonorrhea | Opa (opacity related protein), at 11 loci | ST | Distinctly related STs manifested unique Opa profiles. Minimal Opa variation was found among isolates of the same ST [81] | 14 | [81, 82] |
| <i>Neisseria lactamica</i> | Asymptomatic colonization of nasopharynx in young children | FetA VR (outer membrane protein) | ST | Particular VRs were associated with STs [83] | 275 | [84] |
| <i>Neisseria meningitidis</i> | Meningitis, septicemia | PorA VR1, VR2 (outer membrane protein) | Clonal complex | Nonoverlapping PorA VR1, VR2 combinations were associated with clonal complexes [9] | 977 | [8] |

PubMed [99] was searched for studies which included typing data from metabolic and antigen-encoding genes. MTs refer to: allelic variants of metabolic genes, or sequence types, clonal complexes or ETs from multilocus typing schemes (MLST and MLEE) which are based on genes (or fragments thereof) coding for enzymes involved in metabolic activity. Bacteria which had been typed using MLST schemes based on noncoding DNA, such as ribosomal intergenic spacers, were not included. The AT is represented by: variants of single antigens which generate distinct immunogenic responses, or multilocus variants exhibiting unique combinations of alleles at a number of antigen-encoding loci.

AT: Antigenic type; ClfA & ClfB: Clumping factor A & B; eBurst Group: ET: Electrophoretic type; LPS: Lipopolysaccharide; MLEE: Multilocus enzyme electrophoresis; MLST: Multilocus sequence typing; MT: Metabolic types; OMP: Outer-membrane protein; SlpA: Surface-associated protein A; ST: Sequence type; VR: Variable region.

| Table 1. Associations among antigen-encoding and metabolic genes for 15 bacterial pathogens, for which frequent horizontal genetic transfer has been documented (cont.). | | | | | |
|--|---|---|----------------|---|---------------|
| Pathogen | Common disease(s) | Antigenic type | Metabolic type | Details | isolates Ref. |
| <i>Neisseria meningitidis</i> (cont.) | | PorA VR1, VR2; FetA VR | Clonal complex | Nonoverlapping PorA VR1, VR2: FetA VR combinations were associated with clonal complexes [5] | 3760 |
| | | fHbp (factor H binding protein) | Clonal complex | Subfamily/variant fHbp alleles clustered with particular clonal complexes [85] | 107 |
| | | Opa (opacity related protein), at 3–4 loci | Clonal complex | For each clonal complex examined, particular Opa alleles were consistently observed at each locus [86] | 77 |
| <i>Pasteurella trehalosi</i> | Systemic infections in sheep | Serotype, outer membrane protein, LPS | ET | Isolates of the same ET were generally associated with the same combination of serotype, LPS type and outer-membrane protein type [87] | 60 [88] |
| <i>Salmonella enterica</i> | Gastroenteritis | Serovar (O and H antigens) | eBGs | Strong correlation between eBGs and serovars (except those in lineage 3 – where many eBGs are associated with >1 serovar) [89] | 4257 [90] |
| <i>Staphylococcus aureus</i> | Opportunistic infections of skin | Spa types (staphylococcal protein A), ClfA & ClfB | Clonal complex | 97% (98 out of 101) of spa types were associated with a single clonal complex. ClfA and ClfB sequence variants were closely associated with clonal complex [91] | 224 [92] |
| | | Spa types | ST | Predominant clonal types had nonoverlapping associations between spa type and ST [93] | 182 |
| <i>Streptococcus pneumoniae</i> | Pneumonia, meningitis, septicaemia | Serotype (polysaccharide capsule) | ST | 76% (26 out of 34) of STs were associated with a single serotype [94] | 295 [95] |
| <i>Streptococcus pyogenes</i> | Opportunistic infections (including necrotizing fasciitis, streptococcal toxic shock syndrome, tonsillitis, impetigo) | Emm types (M protein) | ST | 95% (208 out of 220) of STs were associated with a single emm type [96] | 495 [97] |
| | | Emm types (M protein) | ST | 97% (97/100) of STs were associated with a single emm type [98] | 212 |

PubMed [99] was searched for studies which included typing data from metabolic and antigen-encoding genes. MTs refer to: allelic variants of metabolic genes, or sequence types, clonal complexes or ETs from multilocus typing schemes (MLST and MLEE) which are based on genes (or fragments thereof) coding for enzymes involved in metabolic activity. Bacteria which had been typed using MLST schemes based on noncoding DNA, such as ribosomal intergenic spacers, were not included. The AT is represented by: variants of single antigens which generate distinct immunogenic responses, or multilocus variants exhibiting unique combinations of alleles at a number of antigen-encoding loci.

AT: Antigenic type; ClfA & ClfB: Clumping factor A & B; eBG: eBurst Group; ET: Electrophoretic type; LPS: Lipopolysaccharide; MLEE: Multilocus enzyme electrophoresis; MLST: Multilocus sequence typing; MT: Metabolic types; OMP: Outer-membrane protein; SfpA: Surface-associated protein A; ST: Sequence type; VR: Variable region.

| Table 2. Number of years over which combinations of serotype and multilocus sequence typing-defined sequence type have been recorded for <i>Streptococcus pneumoniae</i> , according to the PubMLST website. | | | |
|--|----------|------------------|--------------------------------|
| Sequence type | Serotype | Timespan (years) | Frequency (number of isolates) |
| 15 | 14 | 37 | 84 |
| 53 | 8 | 25 | 48 |
| 63 | 14 | 23 | 137 |
| 81 | 19F | 22 | 80 |
| 81 | 23F | 30 | 338 |
| 90 | 6B | 28 | 126 |
| 113 | 18C | 32 | 32 |
| 156 | 9V | 24 | 136 |
| 172 | 19A | 34 | 67 |
| 247 | 4 | 31 | 25 |
| 447 | 37 | 11 | 31 |
| Owing to the nature of the database, these figures represent the minimum length of time over which such isolates have existed. Data were taken from the PubMLST website [100] on 22 July 2015. | | | |

and metabolic competition may be exerting an influence on the population structure of these pathogens, highly frequent HGT is eroding the associations between antigenic and metabolic types that may arise.

The number of species presented in **Table 1** is limited for several reasons. First, the examples presented are constrained by the specific typing methods used to characterize pathogens; only those studies which include typing information on both metabolic and antigen-encoding genes were included. Second, there are a number of pathogens which show very little genetic variation among isolates as they evolved only a short time ago or have recently passed through population bottlenecks. Such monomorphic pathogens, including *M. tuberculosis*, *Treponema pallidum*, *Yersinia pestis*, *Bacillus anthracis* and *Chlamydophila pneumoniae*, may not manifest antigenic and/or metabolic variation within populations for this reason [111–115]. Low antigenic diversity may also result if the primary antigens are conserved across the population and exert a strong immune response [10].

Implications for clinical interventions & disease control

The accumulating evidence for the essential roles of variants in metabolic genes in virulence and pathogenesis supports the need for further study of this area. As well as enhancing our understanding of the disease process, the identification of metabolic-associated proteins which play essential roles in virulence and pathogenesis could be exploited to design antimicrobial drugs which inhibit virulence-associated characteristics. Such

an approach is therapeutically attractive as the drug would not kill the pathogen outright, and consequently would be less likely to select for resistance. For example, bacterial ureases are a target for the development of novel antimicrobial and could lead to new therapeutics for urinary tract and gastric infections [33].

The notion that strains of bacterial pathogens occupy distinct metabolic niches within the host, and that differences in metabolism and transport genes lead to differences in transmission and fitness, has a number of important implications for clinical interventions. Watkins *et al.* used a mathematical model combining antigen-encoding and metabolic loci to explore the effects of metabolic competition on serotype-targeted vaccination [59]. They found that vaccinating against particular serotypes can cause their metabolic components to transfer through HGT to nonvaccine serotypes, an effect referred to as vaccine-induced metabolic shift (**Figure 3**). The authors extended the model to include virulence-associated loci and found that vaccination resulted in the transfer of virulence genes in addition to metabolic genes, from vaccine to nonvaccine strains.

The pneumococcal conjugate vaccines that are currently available target only a subset of the 90 different serotypes (7, 10 or 13 serotypes). This has resulted in the increase of nonvaccine serotypes following vaccination in a number of countries worldwide [116]. The predictions of vaccine-induced metabolic shift are consistent with a number of changes observed in pneumococcal population structure following mass vaccination. A number of nonvaccine serotypes have become associated with MLST profiles of

vaccine serotypes since the introduction of the heptavalent pneumococcal conjugate vaccine (PCV-7). In North America, for example, there has been a significant increase in the prevalence of the nonvaccine serotype 19A: initially this was due to clonal expansion of the pre-existing genotype ST199, but ST320 – previously associated with the vaccine serotype 19F – has now replaced ST199 as the most common genotype in 19A invasive disease and carriage in several US regions and in Canada [117–119]. Increases have also occurred in the prevalence of the ST695^{19A} strain since the introduction of PCV-7, in

which the vaccine serotype 4 capsule has been switched for a 19A [120,121]. Furthermore, the model predictions helped to explain changes in genomic population structure observed in a dataset of 616 genomes of *S. pneumoniae* from the USA [122], in which metabolic profiles were observed to shift from vaccine to nonvaccine serotypes following vaccination [59]. Such predictions were also consistent with the increase in virulence-associated pili genes observed among pneumococcal nonvaccine strains following vaccination [123]. Importantly, the observation of vaccine-induced metabolic shift in the

Table 3. Alternative theories to explain associations between antigenic and metabolic types in bacterial pathogens.

| Theory | Summary | Limitations |
|--|--|---|
| Neutral model [7] | Bacterial population structures are maintained by neutral mutational drift | Unable to account for the population structures of <i>S. aureus</i> , <i>N. meningitidis</i> and <i>S. pneumoniae</i> [7] |
| CLONAL Epidemic model [6] | The rapid spread of epidemic clones can result in temporary linkage disequilibrium between strains | Unable to account for the longevity of sequence types of <i>N. meningitidis</i> [8,9] |
| Neutral microepidemic model [7] | Bacterial population structures are maintained by neutral mutational drift, modulated by HGT and epidemic transmission chains | Unable to account for the diversity and longevity of sequence types of <i>N. meningitidis</i> [8,9] |
| Restriction modification systems [102] | Whole genome clades are associated with different RM systems, which prevent the exchange of DNA between clades despite frequent HGT | Associations between clades and RM systems could be interpreted as a consequence of the diversification processes caused by other mechanisms (e.g., metabolic and immunological competition), as opposed to the cause of the observed structure |
| Biased sampling [6] | Linkage disequilibrium in recombining populations can result superficially from biased sampling of hyperinvasive clones which are prevalent in invasive disease isolates, and thus not representative of the natural population as a whole | Nonoverlapping associations are evident in a number of population samples obtained solely from carriage. The studies in Table 1 and data in Figure 3 comprise both invasive and carriage collections |
| Epistasis [103] | Nonoverlapping associations between metabolic and antigenic types have arisen as a result of epistatic interactions between antigenic and metabolic areas of the genome, resulting in heightened fitness. Specialized metabolic machinery may be required for the uptake and synthesis of particular antigens (e.g., polysaccharide sugars) Strong epistasis would result in consistent associations of antigenic and metabolic types being observed across different time periods and geographical locations. Consistent associations between serotype and sequence type are frequently reported for <i>S. pneumoniae</i> | Croucher <i>et al.</i> [103] used knockouts of the capsule locus in <i>S. pneumoniae</i> and found that serogroup-specific adaptations may not be responsible for the associations. They also examined the genomic distribution of carbohydrate uptake genes to assess if limitations to acquiring the necessary carbohydrates to synthesize a given serogroup accounts for the associations, but results did not generally support this hypothesis The theoretical frameworks predict that associations between antigenic and metabolic type are arbitrary (i.e., an antigenic type may associate with one metabolic type in one location/time period, and with another in a different location/period). Indeed, one antigenic type associates with multiple sequence types/clonal complexes (and <i>vice versa</i>) across different time periods and regions in <i>N. meningitidis</i> [5,9,14] For <i>S. pneumoniae</i> , there are still examples of multiple sequence types linked to the same serotype in distinct locations [94,104–105] |

HGT: Horizontal genetic transfer; RM: Restriction modification.

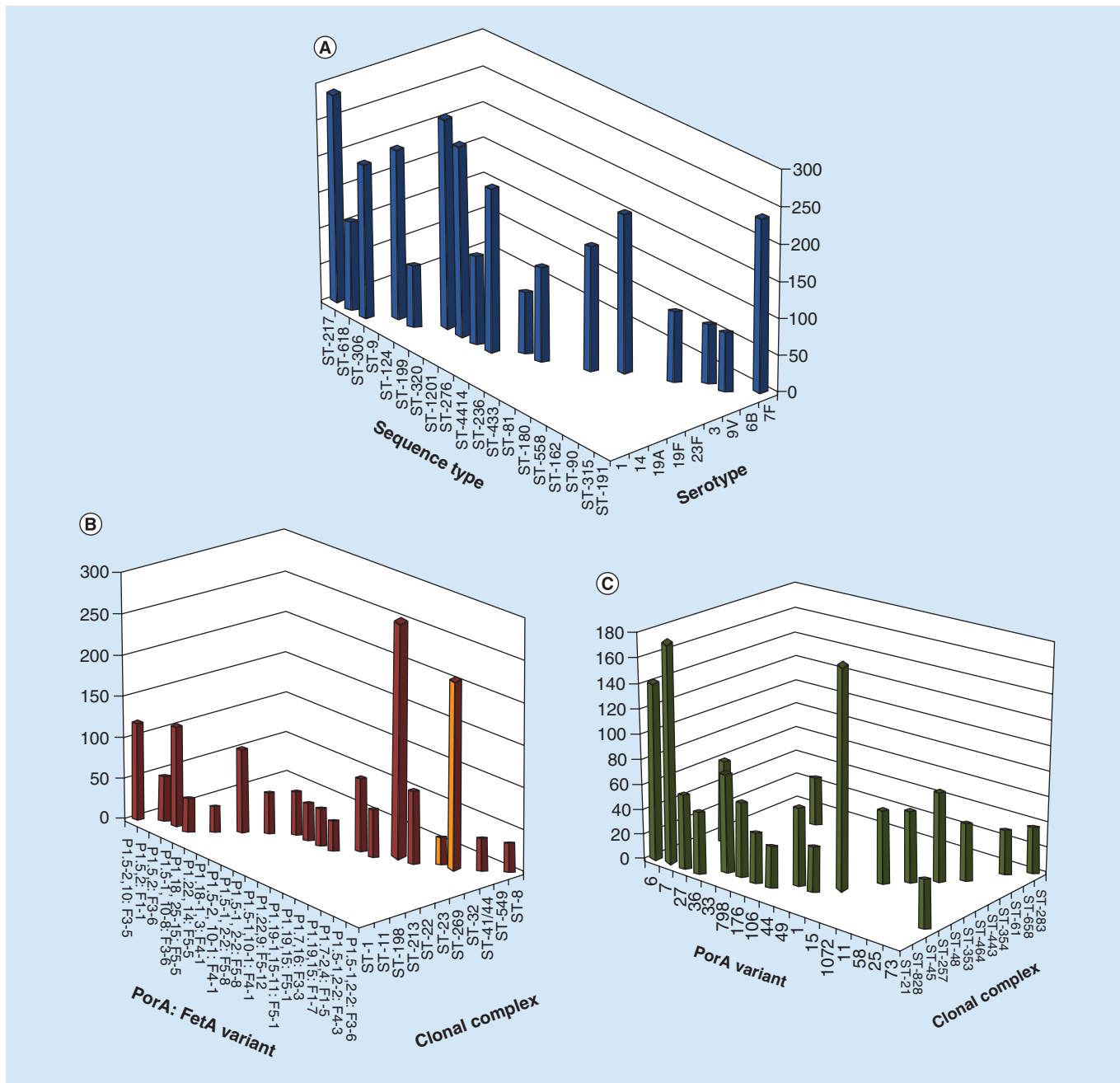


Figure 2. Nonoverlapping associations between alleles at antigen-encoding and multilocus sequence typing-defined metabolic loci in bacterial pathogen populations. Allelic combinations among (A) serotype and sequence type (with a frequency >80 isolates per combination) in *Streptococcus pneumoniae*. (B) PorA; FetA variant and clonal complex (with a frequency >20) in *Neisseria meningitidis*. (C) PorA variant and clonal complex (with a frequency >30) in *Campylobacter jejuni*. All data were extracted from the PubMLST and multilocus sequence typing websites as described in [Figure 2](#).

USA suggests that neutral processes alone may not account for the dynamics observed within pneumococcal populations.

The metabolic shifts predicted by the model also have implications for resistance to antimicrobial drugs. The alleles associated with drug

resistance may also shift to nonvaccine serotypes as a result of the removal of competition at these loci. Indeed, many postvaccine variants possessing metabolic and virulence factors previously associated with vaccine strains show increased resistance to antimicrobial drugs [118].

It is also possible that the increased transmission fitness of such strains may help to alleviate the fitness costs associated with antibiotic resistance.

Finally, understanding the distinct metabolic niches of virulent pathogen strains could ultimately help in the design of probiotic treatments. As *E. coli* strains have slightly different sugar requirements, each using approximately seven sugars in total and including at least one sugar not used by the others *in vivo*, it would be possible in theory to prevent invasion by virulent strains through precolonizing the host with specific combinations of strains which together fill the nutritional niche of the invasive strains [26]. Indeed, inoculating mice with specific mixtures of commensal *E. coli* strains is able to prevent invasion by the virulent strain O157:H7 [52,124]. As different virulent strains (or pathotypes) have distinct nutritional requirements, no single combination of commensal strains will be able to prevent invasion against all virulent strains – but

it may be feasible to artificially engineer such strains that do [42].

Conclusion

Populations of bacterial pathogens are paradoxically structured into distinct strains despite frequent HGT. There are a number of theories and conceptual frameworks which attempt to account for these observations (Table 3), but few have focused on bacterial metabolism. There is an increasing evidence base which suggests that bacterial strains compete for metabolic resources within the host; it is possible that this metabolic competition may be sufficiently strong so as to structure pathogen populations into distinct metabolic types, which overlap less in their resource requirements (Figure 1). A series of mathematical models in which bacterial strains are defined by antigenic and metabolic genes [9,59] predicted that strong competition would result in bacterial populations in which the predominant strains manifest nonoverlapping associations between

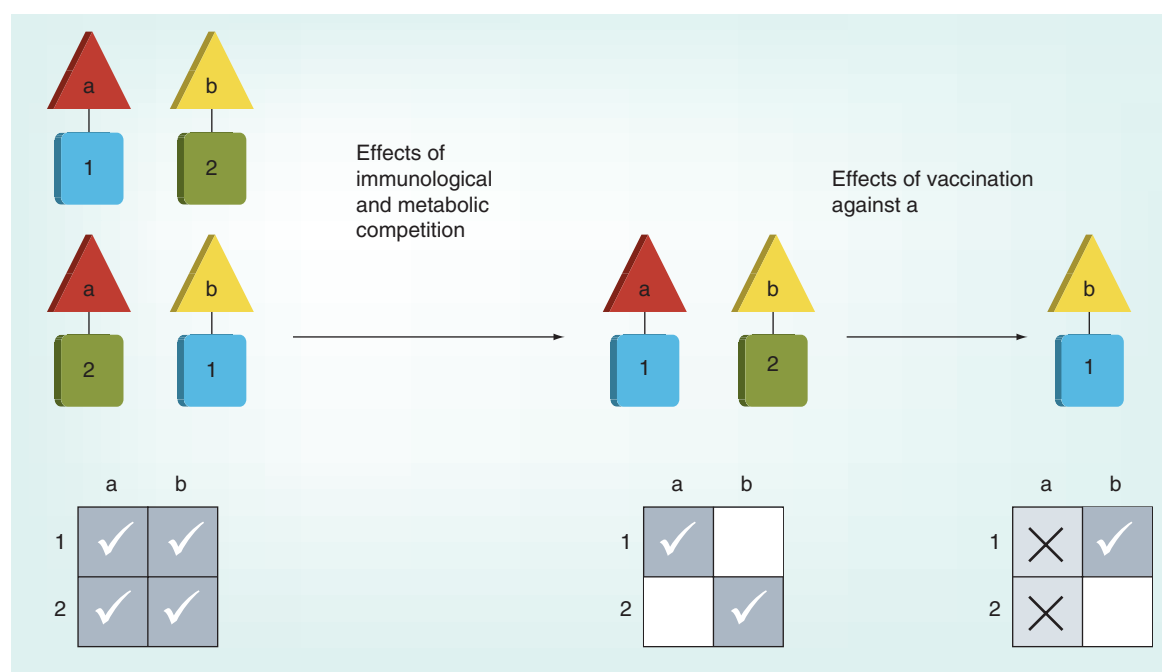


Figure 3. Vaccine-induced metabolic shift, following vaccination against antigenic type a. In a population comprising two antigenic types (a and b) and two metabolic types (1 and 2; where 1 has a slightly higher transmission fitness than 2), there are four possible strains if full recombination is assumed (a1, b1, a2 and b2). As a result of immunological and metabolic competition, the population falls into nonoverlapping associations, with the surviving strains showing minimal overlap in antigenic and metabolic alleles (a1 and b2). Despite strain b1 having a higher transmission efficiency than b2, it is suppressed by metabolic competition with a1. Following vaccination against serotype a, the previously suppressed strain b1 expands and competitively excludes strain b2 owing to its greater transmission efficiency.

Adapted with permission from [59].

antigenic and metabolic alleles. Associations between MLST-defined metabolic genes and outer membrane antigens have been well-documented in *N. meningitidis*, a subset of which were associated with invasive disease and have persisted for several decades [5,9,85,86]. More recent allele-based analyses of 616 pneumococcal whole genomes [59] (originally published by Croucher *et al.* [122]) suggested that such nonoverlapping associations were also evident at the whole-genome level. After reviewing the literature, we find that such associations are found between antigenic and metabolic genes in a variety of bacterial pathogens despite frequent HGT (Table 1 & Figure 2).

Future perspective

As Rohmer recently argued, bacterial metabolism is one of the most fundamental aspects of virulence and transmission of bacterial pathogens, yet our understanding of it remains limited [33]. It is important that we understand the mechanisms which underlie the evolution and maintenance of metabolic diversity both within and among bacterial species – whether they are shaped by neutral processes or by selection and competition. Many of the studies on bacterial metabolism to date have been based on transcriptional data obtained during infection or colonization models, comparative genomics and deletion mutants grown *in vitro*. It is of concern that some of the data obtained from these studies are inconsistent with each other, depending on the model system used, the type of infection and the route of inoculation [33]. This stresses the importance of making use of model systems which mimic the host environment as closely as possible. Indeed, much remains to be understood about resource-based metabolic competition mediated through biofilms, quorum sensing and through interactions with other species present in the natural host microbiome. Such interactions may prove vital to the way in which bacterial strain compete for nutrients in the host, but are relatively poorly understood; for example, it is likely that *E. coli* obtains several of its required sugars locally from specific anaerobes in mixed biofilms [26].

A number of novel technologies, in addition to improvements to current techniques, have emerged in recent years which should further our understanding of bacterial pathogen metabolism. Flow-cell technologies, in which bacteria are supplied with a constant flow of nutrients and oxygen within a cell, allow researchers to

better understand the formation of biofilm communities and the physiological and competitive processes of the bacteria therein. A variety of such methods are now available, enabling the investigation of nutrient gradients, spatio-temporal analyses of metabolism and growth in a controlled *in vitro* environment [125]. An alternative approach to metabolism research, which makes use of isotopically labeled carbon atoms in metabolic substrates, has recently been used to investigate the metabolic diversity of a number of bacteria in soil communities [126]. In conjunction with nanometer-scale secondary ion MS, which measures the isotopic composition of single cells, the rate at which single cells assimilate isotopically labeled substrates into their biomass can be determined [127]. Zimmerman and colleagues recently made use of flow cell technologies in combination with nanometer-scale secondary ion MS to gain a detailed understanding of nitrogen and carbon dioxide fixation at the cellular level in the green sulfur bacterium *Chlorobium phaeobacteroides*. Perhaps such technologies could be utilized to investigate the metabolic heterogeneities among pathogenic strains. These detailed cellular approaches would complement the data obtained from whole-genome flux metabolic models, in which genetic information on primary metabolic pathways obtained from whole genomes are verified through flux-balance analysis in chemostat cultures [128].

In order to further our understanding of the importance of within-host metabolism in pathogenesis, it would be valuable for both existing and novel techniques to be used in experimental models which provide a representation of the natural disease-causing process, which is as realistic as possible. Studies which employ *in vivo* transcriptomics in infection and colonization models continue to reveal a wealth of information on metabolic control of pathogenesis (reviewed by [129]), as they provide a comprehensive picture of the genes which are upregulated during pathogenesis. Such studies are in turn complemented by gene knock-out experiments to verify the role of various metabolic genes in the control of pathogenesis. For example, Jorth *et al.* used RNA sequencing to analyze the genes which were involved in the pathogenesis of the opportunistic human pathogen *Aggregatibacter actinomycetemcomitans*. They found that genes involved in fermentative metabolism and anaerobic respiration were

important, and used mutational analyses to verify the importance of particular enzymes in the process [130].

With the current rate of accumulation of whole genome sequence data, the increasing number of functional genomics studies and growing interest in metabolic modeling, in addition to new technologies, it is likely in the next 5–10 years that we will gain large increases in metabolic genomic and functional data. Mathematical models in which strains can be divided into different genomic functional components may act as useful conceptual tools with which to analyze the wealth of accumulating whole genome data and interpret the patterns observed. Such models allow a number of predictions to be tested, both experimentally and *in silico* through the analysis of whole genome data. The observation of metabolic shift following vaccination in pneumococcal populations in the USA provides just one example of how one can test the hypothesis of metabolic competition. As well as vaccine-induced metabolic shift,

differences between metabolic genes may also be found between strains of successive outbreaks or epidemics. There are a number of examples of newly introduced and highly successful virulent strains which out-compete the previously dominant strains, but with all strains bearing the same principal antigens. The newly invading strain may have superior metabolic characteristics which allow it to out-compete the resident strains, even with an identical antigenic repertoire.

Much work remains in order to further test the hypothesis that strains compete metabolically within the host, and it is through an interdisciplinary approach of experimental work, genomic and functional genomic analysis and mathematical modeling that we can fully understand the role of metabolic competition as a driver of bacterial population structure.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a

EXECUTIVE SUMMARY

Paradoxical population structures of bacterial pathogens

- Many bacterial pathogens are composed of distinct strains despite frequent horizontal genetic transfer (HGT).
- The prevalent strains which arise in many bacterial populations show nonoverlapping associations between alleles of metabolic and antigen-encoding genes, which can persist for many years despite high rates of HGT.
- A number of frameworks have been proposed to account for the population structures of bacterial pathogens, but few can account for the high rates of HGT and long-term persistence of strains, as well as the nonoverlapping patterns observed.

Potential importance of the metabolic genes

- With the advent of next-generation sequencing techniques and high-throughput functional assays, metabolic genes have been demonstrated to play important roles in transmission and virulence in recent years. The notion that pathogen strains occupy distinct metabolic niches has gained increasing support from studies with enteric and respiratory pathogens.
- Most frameworks of population structure have modeled only antigenic variants, or do not distinguish explicitly between different types of genes.
- Buckee *et al.* and Watkins *et al.* investigated the effects of competition operating at polymorphic/variable metabolic and antigen-encoding loci in a series of mathematical models. The results showed stable associations between alleles at metabolic and antigen-encoding loci, suggesting that both immune-mediated and metabolic competition may be important phenomena in the structuring of bacterial pathogens.

Future studies

- Metabolic competition among pathogen strains has important implications for clinical interventions including vaccines, antibiotic resistance and probiotic treatments.
- A growing number of studies are focusing on the importance of the metabolic genes in transmission, competition and virulence; conceptual frameworks that account explicitly for the metabolic genes may form a useful basis for analyzing the data accumulated.

financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Open access

This article is distributed under the terms of the Creative Commons Attribution License 4.0 which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited. To view a copy of the license, visit <http://creativecommons.org/licenses/by/4.0/>

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest.

- Chewapreecha C, Harris SR, Croucher NJ *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* 46(3), 305–309 (2014).
- Whittam TS, Ochman H, Selander RK. Multilocus genetic-structure in natural-populations of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 80, 1751–1755 (1983).
- Maiden MCJ, Bygraves JA, Feil E *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl Acad. Sci. USA* 95, 3140–3145 (1998).
- Caugant DA, Frøholm LO, Bøvre K *et al.* Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease. *Proc. Natl Acad. Sci. USA* 83, 4927–4931 (1986).
- Watkins ER, Maiden MCJ. Persistence of hyperinvasive meningococcal strain types during global spread as recorded in the PubMLST database. *PLoS ONE* 7(9), e45349 (2012).
- Smith JM, Smith NH, O'Rourke M, Spratt BG. How clonal are bacteria. *Proc. Natl Acad. Sci. USA* 90, 4384–4388 (1993).
- Fraser C, Hanage WP, Spratt BG. Neutral microepidemic evolution of bacterial pathogens. *Proc. Natl Acad. Sci. USA* 102, 1968–1973 (2005).
- Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MCJ. The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Mol. Biol. Evol.* 22, 562–569 (2005).
- Buckee CO, Jolley KA, Recker M *et al.* Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. *Proc. Natl Acad. Sci. USA* 105, 15082–15087 (2008).
- Gupta S, Maiden MCJ, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. *Nat. Med.* 2, 437–442 (1996).
- Cobey S, Lipsitch M. Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science* 335(6074), 1376–1380 (2012).
- Flasche S, Edmunds WJ, Miller E, Goldblatt D, Robertson C, Choi YH. The impact of specific and non-specific immunity on the ecology of *Streptococcus pneumoniae* and the implications for vaccination. *Proc. Biol. Sci.* 280, 20131939 (2013).
- Gupta S, Ferguson N, Anderson R. Chaos, persistence, and evolution of strain structure in antigenically diverse infectious agents. *Science* 280(5365), 912–915 (1998).
- Buckee CO, Gupta S, Kriz P, Maiden MCJ, Jolley KA. Long-term evolution of antigen repertoires among carried meningococci. *Proc. Biol. Sci.* 277, 1635–1641 (2010).
- Bjune G, Høiby EA, Grønnesby JK *et al.* Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* 338(8775), 1093–1096 (1991).
- Oster P, Lennon D, O'Hallahan J, Mulholland K, Reid S, Martin D. McNZB: a safe and highly immunogenic tailor-made vaccine against the New Zealand *Neisseria meningitidis* serogroup B disease epidemic strain. *Vaccine* 23(17–18), 2191–2196 (2005).
- Snape MD, Saroey P, John TM *et al.* Persistence of bactericidal antibodies following early infant vaccination with a serogroup B meningococcal vaccine and immunogenicity of a preschool booster dose. *CMAJ* 185(15), E715–E724 (2013).
- Callaghan MJ, Buckee CO, Jolley KA, Kriz P, Maiden MCJ, Gupta S. The effect of immune selection on the structure of the meningococcal Opa protein repertoire. *PLoS Pathog.* 4(3), e1000020 (2008).
- Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B. Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and emm-gene type with T-protein agglutination pattern and serum opacity factor. *J. Med. Microbiol.* 55, 157–164 (2006).
- Areschoug T, Ståhlhammar-Carlemalm M, Larsson C, Lindahl G. Group B streptococcal surface proteins as targets for protective antibodies: Identification of two novel proteins in strains of serotype V. *Infect. Immun.* 67(12), 6350–6357 (1999).
- Futse JE, Brayton KA, Dark MJ, Knowles DP, Palmer GH. Superinfection as a driver of genomic diversification in antigenically variant pathogens. *Proc. Natl Acad. Sci. USA* 105(6), 2123–2127 (2008).
- Auranen K, Mehtälä J, Tanskanen A, S Kaltoft M. Between-strain competition in acquisition and clearance of pneumococcal carriage – epidemiologic evidence from a longitudinal study of day-care children. *Am. J. Epidemiol.* 171, 169–176 (2010).
- Lipsitch M, Abdullahi O, D'Amour A *et al.* Estimating rates of carriage acquisition and clearance and competitive ability for pneumococcal serotypes in Kenya with a Markov transition model. *Epidemiology* 23, 510–519 (2012).
- Lipsitch M, Dykes JK, Johnson SE *et al.* Competition among *Streptococcus pneumoniae* for intranasal colonization in a mouse model. *Vaccine* 18, 2895–2901 (2000).
- Tong H, Chen W, Merritt J, Qi F, Shi W, Dong X. *Streptococcus oligofermentans* inhibits *Streptococcus mutans* through conversion of lactic acid into inhibitory H₂O₂: a possible counteroffensive strategy for interspecies competition. *Mol. Microbiol.* 63(3), 872–880 (2007).
- Conway T, Cohen PS. Commensal and pathogenic *Escherichia coli* metabolism in the gut. *Microbiol. Spectr.* (3), 1–15 (2015).
- Highly comprehensive review on the research of *Escherichia coli* metabolism and competition within the intestine.
- Lawley TD, Walker AW. Intestinal colonization resistance. *Immunology* 138(1), 1–11 (2013).
- Van der Waaij D, Van der Waaij BD. The colonization resistance of the digestive tract in

- different animal species and in man; a comparative study. *Epidemiol. Infect.* 105(2), 237–243 (1990).
- 29 Tibayrenc M. *Genetics and Evolution of Infectious Diseases*. Science Direct, Elsevier, Oxford (2011).
 - 30 Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MCJ. A gene-by-gene population genomics platform: *de novo* assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics* 15, 1138 (2014).
 - 31 Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT. The evolution of metabolic networks of *E. coli*. *BMC Syst. Biol.* 5, 182 (2011).
 - 32 Joseph B, Schwarz RF, Linke B *et al.* Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome. *PLoS ONE* 6(4), e18441 (2011).
 - 33 Rohmer L, Hocquet D, Miller SI. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol.* 19(7), 341–348 (2011).
 - Uses a number of detailed examples of different bacterial pathogens to explore the idea that metabolism plays a key role in pathogenesis.
 - 34 Helling RB. Speed versus efficiency in microbial growth and the role of parallel pathways. *J. Bacteriol.* 184(4), 1041–1045 (2002).
 - 35 Sabarly V, Bouvet O, Glodt J *et al.* The decoupling between genetic structure and metabolic phenotypes in *Escherichia coli* leads to continuous phenotypic diversity. *J. Evol. Biol.* 24(7), 1559–1571 (2011).
 - 36 Schoen C, Kischkies L, Elias J, Ampattu BJ. Metabolism and virulence in *Neisseria meningitidis*. *Front. Cell. Infect. Microbiol.* 4, 114 (2014).
 - 37 Shelburne SA, Davenport MT, Keith DB, Musser JM. The role of complex carbohydrate catabolism in the pathogenesis of invasive streptococci. *Trends Microbiol.* 16, 318–325 (2008).
 - 38 Buckwalter CM, King SJ. Pneumococcal carbohydrate transport: food for thought. *Trends Microbiol.* 20, 517–522 (2012).
 - 39 Abu Kwaik Y, Bumann D. Microbial quest for food *in vivo*: “nutritional virulence” as an emerging paradigm. *Cell Microbiol.* 15, 882–890 (2013).
 - 40 Bryan JD, Liles R, Cvek U, Trutschl M, Shelper D. Global transcriptional profiling reveals *Streptococcus agalactiae* genes controlled by the MtaR transcription factor. *BMC Genomics* 9, 607 (2008).
 - 41 Mahdi LK, Deihimi T, Zamansani F *et al.* A functional genomics catalogue of activated transcription factors during pathogenesis of pneumococcal disease. *BMC Genomics* 15, 769 (2014).
 - 42 Meador JP, Caldwell ME, Cohen PS, Conway T. *Escherichia coli* pathotypes occupy distinct niches in the mouse intestine. *Infect. Immun.* 82(5), 1931–1938 (2014).
 - 43 Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8(1), 15–25 (2010).
 - 44 Philips BJ, Meguer JX, Redman J, Baker EH. Factors determining the appearance of glucose in upper and lower respiratory tract secretions. *Intensive Care Med.* 29, 2204–2210 (2003).
 - 45 Armstrong SK. Bacterial metabolism in the host environment: pathogen growth and nutrient assimilation in the mammalian upper respiratory tract. *Microbiol. Spectr.* 3(3), doi: 10.1128/microbiolspec.MBP-0007-2014 (2015).
 - 46 Marion C, Limoli DH, Bobulsky GS, Abraham JL, Burnaugh AM, King SJ. Identification of a pneumococcal glycosidase that modifies O-linked glycans. *Infect. Immun.* 77(4), 1389–1396 (2009).
 - 47 Blaut M, Clavel T. Metabolic diversity of the intestinal microbiota: implications for health and disease. *J. Nutr.* 137(3 Suppl. 2), S751–S755 (2007).
 - 48 Fuchs TM. From food to cell: nutrient exploitation strategies of enteropathogens. *Microbiology* 160, 1020–1039 (2014).
 - 49 Bidossi A, Mulas L, Decorosi F *et al.* A functional genomics approach to establish the complement of carbohydrate transporters in *Streptococcus pneumoniae*. *PLoS ONE* 7(3), e33320 (2012).
 - 50 Linke CM, Woodiga SA, Meyers DJ, Buckwalter CM, Salhi HE, King SJ. The ABC transporter encoded at the pneumococcal fructooligosaccharide utilization locus determines the ability to utilize long- and short-chain fructooligosaccharides. *J. Bacteriol.* 195, 1031–1041 (2013).
 - 51 Fabich AJ, Jones SA, Chowdhury FZ *et al.* Comparison of carbon nutrition for pathogenic and commensal *Escherichia coli* strains in the mouse intestine. *Infect. Immun.* 76(3), 1143–1152 (2008).
 - 52 Maltby R, Leatham-Jensen MP, Gibson T, Cohen PS, Conway T. Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. *PLoS ONE* 8(1), e53957 (2013).
 - 53 Higgins MA, Whitworth GE, El Warry N *et al.* Differential recognition and hydrolysis of host carbohydrate antigens by *Streptococcus pneumoniae* family 98 glycoside hydrolases. *J. Biol. Chem.* 284, 26161–26173 (2009).
 - 54 Harrison OB, Evans NJ, Blair JM *et al.* Epidemiological evidence for the role of the hemoglobin receptor, HmbR, in meningococcal virulence. *J. Infect. Dis.* 200, 94–98 (2009).
 - 55 Marsh JW, O’Leary MM, Shutt KA, Harrison LH. Deletion of *fetA* gene sequences in serogroup B and C *Neisseria meningitidis* isolates. *J. Clin. Microbiol.* 45, 1333–1335 (2007).
 - 56 Pagliarulo C, Salvatore P, De Vitis LR *et al.* Regulation and differential expression of *gdhA* encoding NADP-specific glutamate dehydrogenase in *Neisseria meningitidis* clinical isolates. *Mol. Microbiol.* 51, 1757–1772 (2004).
 - 57 Freter F. Mechanisms that control the microflora in the large intestine. In: *Human Intestinal Microflora In Health And Disease*. Academic Press, Inc., New York, NY, USA, 33–54 (1983).
 - 58 Fabich AJ, Leatham MP, Grissom JE *et al.* Genotype and phenotypes of an intestine-adapted *Escherichia coli* K-12 mutant selected by animal passage for superior colonization. *Infect. Immun.* 79(6), 2430–2439 (2011).
 - 59 Watkins ER, Penman BS, Lourenço J, Buckee CO, Maiden MCJ, Gupta S. Vaccination drives changes in metabolic and virulence profiles of *Streptococcus pneumoniae*. *PLoS Pathog.* 11(7), e1005034 (2015).
 - 60 Lipsitch M. Vaccination against colonizing bacteria with multiple serotypes. *Proc. Natl Acad. Sci. USA* 94, 6571–6576 (1997).
 - 61 Zhang Y, Auranen K, Eichner M. The influence of competition and vaccination on the coexistence of two pneumococcal serotypes. *Epidemiol. Infect.* 132(6), 1073–1081 (2004).
 - 62 Gause GF. Experimental analysis of vito volterra’s mathematical theory of the struggle for existence. *Science* 79, 16–17 (1934).
 - 63 Urwin R, Russell JE, Thompson EAL, Holmes EC, Feavers IM, Maiden MCJ. Distribution of surface protein variants among hyperinvasive meningococci: Implications for vaccine design. *Infect. Immun.* 72, 5955–5962 (2004).

- 64 Caugant DA, Mocca LF, Frasch CE, Froholm LO, Zollinger WD, Selander RK. Genetic-structure of neisseria-meningitidis populations in relation to serogroup, serotype, and outer-membrane protein pattern. *J. Bacteriol.* 169, 2781–2792 (1987).
- 65 Bambini S, Piet J, Muzzi A *et al.* An analysis of the sequence variability of meningococcal fHbp, NadA and NHBA over a 50-year period in The Netherlands. *PLoS ONE* 8(5), e65043 (2013).
- 66 Cody AJ, Maiden MJC, Dingle KE. Genetic diversity and stability of the porA allele as a genetic marker in human *Campylobacter* infection. *Microbiology* 155(Pt 12), 4145–4154 (2009).
- 67 Suerbaum S, Lohrengel M, Sonnevend A, Ruberg F, Kist M. Allelic diversity and recombination in *Campylobacter jejuni*. *J. Bacteriol.* 183, 2553–2559 (2001).
- 68 Killgore G, Thompson A, Johnson S *et al.* Comparison of seven techniques for typing international epidemic strains of *Clostridium difficile*: restriction endonuclease analysis, pulsed-field gel electrophoresis, PCR-ribotyping, multilocus variable-number tandem-repeat analysis, amplified fragment length polymorphism, and surface layer protein A gene sequence typing. *J. Clin. Microbiol.* 46(2), 431–437 (2008).
- 69 He M, Sebaihia M, Lawley TD *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl Acad. Sci. USA* 107, 7527–7532 (2010).
- 70 Chowdhury SA, Arias CA, Nallapareddy SR, Reyes J, Willems RJL, Murray BE. A trilocus sequence typing scheme for hospital epidemiology and subspecies differentiation of an important nosocomial pathogen, *Enterococcus faecalis*. *J. Clin. Microbiol.* 47, 2713–2719 (2009).
- 71 de Been M, van Schaik W, Cheng L, Corander J, Willems RJ. Recent recombination events in the core genome are associated with adaptive evolution in *Enterococcus faecium*. *Genome Biol. Evol.* 5(8), 1524–1535 (2013).
- 72 Maslow JN, Whittam TS, Gilks CF *et al.* Clonal relationships among blood-stream isolates of *Escherichia coli*. *Infect. Immun.* 63, 2409–2417 (1995).
- 73 Wirth T, Falush D, Lan RT *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* 60, 1136–1151 (2006).
- 74 Musser JM, Kroll JS, Moxon ER, Selander RK. Evolutionary genetics of the encapsulated strains of *Haemophilus influenzae*. *Proc. Natl Acad. Sci. USA* 85, 7758–7762 (1988).
- 75 Mell JC, Shumilina S, Hall IM, Redfield RJ. Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog.* 7(7), e1002151 (2011).
- 76 Meats E, Feil EJ, Stringer S *et al.* Characterization of encapsulated and noncapsulated *Haemophilus influenzae* and determination of phylogenetic relationships by multilocus sequence typing. *J. Clin. Microbiol.* 41, 1623–1636 (2003).
- 77 Vimont S, Mnif B, Fevre C, Brisse S. Comparison of PFGE and multilocus sequence typing for analysis of *Klebsiella pneumoniae* isolates. *J. Med. Microbiol.* 57, 1308–1310 (2008).
- 78 Brisse S, Fevre C, Passet V *et al.* Virulent clones of *Klebsiella pneumoniae*: identification and evolutionary scenario based on genomic and phenotypic characterization. *PLoS ONE* 4(3), e4982 (2009).
- 79 Piffaretti JC, Kressebuch H, Aeschbacher M *et al.* Genetic-characterization of clones of the bacterium listeria-monocytogenes causing epidemic disease. *Proc. Natl Acad. Sci. USA* 86, 3818–3822 (1989).
- 80 den Bakker HC, Didelot X, Fortes ED, Nightingale KK, Wiedmann M. Lineage specific recombination rates and microevolution in *Listeria monocytogenes*. *BMC Evol. Biol.* 8, 277 (2008).
- 81 Bilek N, Ison CA, Spratt BG. Relative contributions of recombination and mutation to the diversification of the opa gene repertoire of *Neisseria gonorrhoeae*. *J. Bacteriol.* 191, 1878–1890 (2009).
- 82 O'Rourke M, Stevens E. Genetic-structure of neisseria-gonorrhoeae populations – a non-clonal pathogen. *J. Gen. Microbiol.* 139, 2603–2611 (1993).
- 83 Bennett JS, Thompson EAL, Kriz P, Jolley KA, Maiden MCJ. A common gene pool for the Neisseria FetA antigen. *Int. J. Med. Microbiol.* 299, 133–139 (2009).
- 84 Alber D, Oberkotter M, Suerbaum S, Claus H, Frosch M, Vogel U. Genetic diversity of *Neisseria lactamica* strains from epidemiologically defined carriers. *J. Clin. Microbiol.* 39, 1710–1715 (2001).
- 85 Brehony C, Wilson DJ, Maiden MCJ. Variation of the factor H-binding protein of *Neisseria meningitidis*. *Microbiol. SGM* 155, 4155–4169 (2009).
- 86 Callaghan MJ, Jolley KA, Maiden MCJ. Opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. *Infect. Immun.* 74, 5085–5094 (2006).
- 87 Davies RL, Arkinsaw S, Selander RK. Genetic relationships among *Pasteurella trehalosi* isolates based on multilocus enzyme electrophoresis. *Microbiology-UK* 143, 2841–2849 (1997).
- 88 Davies RL, Campbell S, Whittam TS. Mosaic structure and molecular evolution of the leukotoxin operon (lktCABD) in mannheimia (*Pasteurella*) haemolytica, *Mannheimia glucosida*, and *Pasteurella trehalosi*. *J. Bacteriol.* 184, 266–277 (2002).
- 89 Achtman M, Wain J, Weill F-X *et al.* Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 8(6), e100277 (2012).
- 90 Didelot X, Bowden R, Street T *et al.* Recombination and population structure in *Salmonella enterica*. *PLoS Genet.* 7(7), e1002191 (2011).
- 91 Murphy E, Lin SL, Nunez L *et al.* Challenges for the evaluation of *Staphylococcus aureus* protein based vaccines monitoring antigenic diversity. *Hum. Vaccin.* 7, 51–59 (2011).
- 92 Everitt RG, Didelot X, Batty EM *et al.* Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. *Nat. Commun.* 5, 3956 (2014).
- 93 Tavares A, Faria NA, de Lencastre H, Miragaia M. Population structure of methicillin-susceptible *Staphylococcus aureus* (MSSA) in Portugal over a 19-year period (1992–2011). *Eur. J. Clin. Microbiol. Infect.* 33, 423–432 (2014).
- 94 Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 144(Pt 11), 3049–3060 (1998).
- 95 Chewapreecha C, Harris SR, Croucher NJ *et al.* Dense genomic sampling identifies highways of pneumococcal recombination. *Nat. Genet.* 46(3), 305–309 (2014).
- 96 McGregor KF, Spratt BG, Kalia A *et al.* Multilocus sequence typing of *Streptococcus pyogenes* representing most known emm types and distinctions among subpopulation genetic structures. *J. Bacteriol.* 186, 4285–4294 (2004).
- 97 Bessen DE. Population biology of the human restricted pathogen, *Streptococcus pyogenes*. *Infect. Genet. Evol.* 9, 581–593 (2009).
- 98 Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect. Immun.* 69, 2416–2427 (2001).
- 99 PubMed.gov. www.ncbi.nlm.nih.gov/pubmed
- 100 PubMLST. <http://pubmlst.net>

- 101 Victorino LR, Margos G, Feil EJ, Collares-Pereira M, Zé-Zé L, Kurtenbach K. Fine-scale phylogeographic structure of *Borrelia lusitaniae* revealed by multilocus sequence typing. *PLoS ONE* 3(12), e4002 (2008).
- 102 Budroni S, Siena E, Hotopp JCD *et al.* *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proc. Natl Acad. Sci. USA* 108, 4494–4499 (2011).
- 103 Croucher N, Kagedan L, Thompson CM *et al.* Selective and genetic constraints on pneumococcal serotype switching. *PLoS Genet.* 11(3), e1005095 (2015).
- 104 Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, Spratt BG. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J. Infect. Dis.* 187, 1424–1432 (2003).
- 105 Hanage WP, Kaijalainen TH, Syrjanen RK *et al.* Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infect. Immun.* 73, 431–435 (2005).
- 106 Boonsilp S, Thaipadungpanit J, Amornchai P *et al.* A single multilocus sequence typing (MLST) scheme for seven pathogenic leptospira species. *PLoS Negl. Trop. Dis.* 7(1), e1954 (2013).
- 107 Tompkins LS, Troup NJ, Woods T, Bibb W, McKinney RM. Molecular epidemiology of *Legionella* species by restriction endonuclease and alloenzyme analysis. *J. Clin. Microbiol.* 25(10), 1875–1880 (1987).
- 108 Nemec A, Krizova L, Maixnerova M, Musilek M. Multidrug-resistant epidemic clones among bloodstream isolates of *Pseudomonas aeruginosa* in the Czech Republic. *Res. Microbiol.* 161(3), 234–242 (2010).
- 109 Jones N, Bohnsack JF, Takahashi S *et al.* Multilocus sequence typing system for group B *Streptococcus*. *J. Clin. Microbiol.* 41(6), 2530–2536 (2003).
- 110 Van Der Mee-Marquet N, Fourny L, Arnault L *et al.* Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites. *J. Clin. Microbiol.* 46(9), 2906–2911 (2008).
- 111 Achtman M, Zurth K, Morelli G, Torrea G, Guiryoule A, Carniel E. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* 96(24), 14043–14048 (1999).
- 112 Harper KN, Ocampo PS, Steiner BM *et al.* On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl. Trop. Dis.* 2(1), e148 (2008).
- 113 Priest FG, Barker M, Baillie LWJ, Holmes EC, Maiden MCJ. Population structure and evolution of the *Bacillus cereus* group. *J. Bacteriol.* 186(23), 7959–7970 (2004).
- 114 Rattei T, Ott S, Gutacker M *et al.* Genetic diversity of the obligate intracellular bacterium *Chlamydomydia pneumoniae* by genome-wide analysis of single nucleotide polymorphisms: evidence for highly clonal population structure. *BMC Genomics* 8, 355 (2007).
- 115 Sreevatsan S, Pan X, Stockbauer KE *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* 94, 9869–9874 (1997).
- 116 Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. *Lancet* 378, 1962–1973 (2011).
- 117 Beall B, McEllistrem MC, Gertz RE *et al.* Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. *J. Clin. Microbiol.* 44, 999–1017 (2006).
- 118 Pillai DR, Shahinas D, Buzina A *et al.* Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. *BMC Genomics* 10, 642 (2009).
- 119 Yildirim I, Stevenson A, Hsu KK, Pelton SI. Evolving picture of invasive pneumococcal disease in Massachusetts children: a comparison of disease in 2007–2009 with earlier periods. *Pediatr. Infect. Dis. J.* 31, 1016–1021 (2012).
- 120 Golubchik T, Brueggemann AB, Street T *et al.* Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat. Genet.* 44, 352–355 (2012).
- 121 Brueggemann AB, Pai R, Crook DW, Beall B. Vaccine escape recombinants emerge after pneumococcal vaccination in the united states. *PLoS Pathog.* 3, 1628–1636 (2007).
- 122 Croucher NJ, Finkelstein JA, Pelton SI *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology. *Nat. Genet.* 45(6), 656–663 (2013).
- 123 Zahner D, Gudlavalleti A, Stephens DS. Increase in Pilus Islet 2-encoded Pili among *Streptococcus pneumoniae* Isolates, Atlanta, Georgia, USA. *Emerg. Infect. Dis.* 16, 955–962 (2010).
- 124 Leatham MP, Banerjee S, Autieri SM, Mercado-Lubo R, Conway T, Cohen PS. Precolonized human commensal *Escherichia coli* strains serve as a barrier to *E. coli* O157:H7 growth in the streptomycin-treated mouse intestine. *Infect. Immun.* 77(7), 2876–2886 (2009).
- 125 Moormeier DE, Bayles KW. Examination of *Staphylococcus epidermidis* biofilms using flow-cell technology. *Methods Mol. Biol.* 1106, 143–155 (2014).
- 126 Verastegui Y, Cheng J, Engel K *et al.* Multisubstrate isotope labeling and metagenomic analysis of active soil bacterial communities. *MBio* 5(4), e01157–e01114 (2014).
- 127 Zimmermann M, Escriu S, Hübschmann T *et al.* Phenotypic heterogeneity in metabolic traits among single cells of a rare bacterial species in its natural environment quantified with a combination of flow cell sorting and NanoSIMS. *Front. Microbiol.* 6, 243 (2015).
- **Presents novel techniques to quantify phenotypic heterogeneity in the metabolic activities of rare bacteria.**
- 128 Baart GJE, Willemsen M, Khatami E *et al.* Modeling *Neisseria meningitidis* B metabolism at different specific growth rates. *Biotechnol. Bioeng.* 101, 1022–1035 (2008).
- 129 Skvortsov TA, Azhikina TL. [Transcriptome analysis of bacterial pathogens *in vivo*: problems and solutions]. *Bioorg. Khim.* 36(5), 596–606 (2010).
- 130 Jorth P, Trivedi U, Rumbaugh K, Whiteley M. Probing bacterial metabolism during infection using high-resolution transcriptomics. *J. Bacteriol.* 195(22), 4991–4998 (2013).