RESEARCH REVIEW

INTERNATIONAL MICROBIOLOGY

International Microbiology (2012) 15:157-170
DOI: 10.2436/20.1501.01.169 ISSN 1139-6709 **www.im.microbios.org**

Genotypic and phenotypic diversity in the noncapsulated *Haemophilus influenzae*: adaptation and pathogenesis in the human airways

Junkal Garmendia,[1,2]* Pau Martí-Lliteras,[2,3] Javier Moleres,[1] Carmen Puig,[2,4] José A. Bengoechea[2,3,5]

[1]Institute for Agrobiotechnology, CSIC-Public University of Navarra-Government of Navarra, Mutilva, Spain. [2]Biomedical Research Network for Respiratory Diseases (CIBERES), Bunyola, Spain. [3]Laboratory of Microbial Pathogenesis, Foundation Health Balearic Islands, Bunyola, Spain. [4]Microbiology Department, University Hospital Bellvitge, IDIBELL, University of Barcelona, Barcelona, Spain. [5]National Research Council (CSIC), Spain

Received 30 October 2012 · Accepted 15 November 2012

Summary. The human respiratory tract contains a highly adapted microbiota including commensal and opportunistic pathogens. Noncapsulated or nontypable *Haemophilus influenzae* (NTHi) is a human-restricted member of the normal airway microbiota in healthy carriers and an opportunistic pathogen in immunocompromised individuals. The duality of NTHi as a colonizer and as a symptomatic infectious agent is closely related to its adaptation to the host, which in turn greatly relies on the genetic plasticity of the bacterium and is facilitated by its condition as a natural competent. The variable genotype of NTHi accounts for its heterogeneous gene expression and variable phenotype, leading to differential host-pathogen interplay among isolates. Here we review our current knowledge of NTHi diversity in terms of genotype, gene expression, antigenic variation, and the phenotypes associated with colonization and pathogenesis. The potential benefits of NTHi diversity studies discussed herein include the unraveling of pathogenicity clues, the generation of tools to predict virulence from genomic data, and the exploitation of a unique natural system for the continuous monitoring of long-term bacterial evolution in human airways exposed to noxious agents. Finally, we highlight the challenge of monitoring both the pathogen and the host in longitudinal studies, and of applying comparative genomics to clarify the meaning of the vast NTHi genetic diversity and its translation to virulence phenotypes. [**Int Microbiol** 2012; 15(4): 157-170]

Keywords: *Haemophilus influenzae* · noncapsulated/nontypable *Haemophilus influenzae* (NTHi) · pathogen-host interplay · genetic diversity · virulence phenotype

Introduction

The human upper respiratory tract contains a characteristic and highly adapted microbiota encompassing commensal microorganisms and opportunistic pathogens. The fine-tuned balance of the microbial-airway interplay underlies normal lung function, but it can be altered by host genetic factors or immunological status, by host exposure to external factors such as radiation, infectious agents, chemical contaminants, and environmental pollutants, as well as by diet, lifestyle (e.g., tobacco or alcohol use), occupation, and medical interventions [70]. Regardless of their origin, the changing conditions often allow existing or newly acquired

***Corresponding author:** J. Garmendia
Instituto de Agrobiotecnología
CSIC-Universidad Pública de Navarra-Gobierno de Navarra
31192 Mutilva, Navarra, Spain
Tel. +34-948168484. Fax +34-948232191
E-mail: juncal.garmendia@unavarra.es

opportunistic pathogens to modify their status as colonizers, becoming the cause of a symptomatic infection. Moreover, opportunistic pathogens often display high genetic plasticity as a strategy to drive continuous evolution, thereby facilitating the evasion of host immunity, carrier state colonization, or symptomatic infection. In this review, we focus on a member of the human airway microbiota, the opportunist pathogen nontypable *Haemophilus influenzae* (NTHi). We review the most recent knowledge on its genetic diversity and highlight questions and challenges for its future study with respect to heterogeneity, evolution, and host interplay. Although tailored to *H. influenzae*, our discussion is applicable to almost any other opportunistic pathogen.

General features of the bacterial respiratory pathogen *Haemophilus influenzae*

Haemophilus influenzae is a gram-negative coccobacillus whose environmental niche is primarily restricted to the human respiratory tract. It is classified on the basis of its production of a polysaccharide capsule; strain types a–f produce antigenically distinct capsules while nontypable strains do not. The use of *H. influenzae* type b (Hib) conjugate vaccines has nearly eliminated invasive strains in places where the vaccines have been administered, but they have also promoted the emergence of NTHi strains as the most predominant of this pathogen species [1]. NTHi is a member of the human respiratory microbiota in most healthy individuals beginning in early life. Colonization by several different NTHi strains is often simultaneous [18], continuously renovated, and actively modulates colonization by other opportunistic pathogens such as *Streptococcus pneumoniae* [41,66]. In addition to colonizing the nasopharynx of healthy individuals, NTHi is an opportunistic pathogen. Colonization of the upper airways is also the first step in the pathogenesis of NTHi infection, facilitated by contiguous spread of the bacteria and its migration from the nasopharynx to adjacent structures, including the sinuses, middle ear, trachea, and lower airways. Clinical manifestations of NTHi infection are: (i) upper respiratory tract involvement such as otitis media (OM) in children, as well as sinusitis, and conjunctivitis; (ii) exacerbations of conditions involving the lower respiratory tract (LRT) in adults suffering chronic obstructive pulmonary disease (COPD), as well as pneumonia and infections in cystic fibrosis (CF); and (iii) invasive disease, with bacteremia and meningitis as the most common presentations [1].

The notion that NTHi is highly adapted to the host is supported by the fact that this bacterium is: (i) human host-restricted; (ii) successful at establishing a niche in the human airway as a colonizer; and (iii) provided with virulence factors facilitating the pathogen's ability to take advantage of the host condition and cause a symptomatic infection. The adaptation of NTHi is manifested by wide variations in the DNA material among isolates. While encapsulated serotype type b invasive strains form a clonal group, there is enormous genetic heterogeneity among NTHi strains [25].

In general, existing evidence indicates that bacterial strains belonging to the same species vary considerably in gene content, and that the genetic repertoire of a given species is much larger than the gene content of individual strains. This has important consequences for our understanding of bacterial evolution, adaptation, and population structure, as well as for the identification of virulence genes, vaccine design, etc. Bacterial species are currently described by their gene pools (pan-genomes or supra-genomes), which include a core genome containing genes present in all strains and an accessory or adaptive genome consisting of partially shared and strain-specific genes [47]. Available information on genotyping systems and genome sequencing of NTHi strains indicates that the pan-genome of this bacterial species is large [25,35]. The sources of selective pressure driving genetic diversity among populations of *H. influenzae* are likely related to the bacterial necessity to attach to host cells or surfaces for colonization, to evade host innate and adaptive immunity and persist in the host, to obtain iron and other nutrients essential for replication, and to disseminate or spread.

Genetic mechanisms that modify *H. influenzae* gene content are: (i) the lateral transfer of DNA sequences between different bacterial cells, facilitated by the fact that NTHi is a naturally competent DNA acceptor [56]; (ii) genetic polymorphisms, encompassing gene point mutations, insertions, deletions, or duplications [25]; (iii) phase variation, a slipped-strand mispairing mediated by short DNA repeats (SSR, simple sequence repeats) in the coding or the upstream promoter regions of certain genes such that a spontaneous gain or loss of repeat units in these unstable regions either results in a translational frameshift or alters the distance spanned by the promoter, thus modifying gene expression [48]; and (iv) hypermutation [54].

Genetic variability is likely to have fundamental consequences in NTHi infection, favoring heterogeneous gene expression as well as phenotypic and antigenic diversity while providing this pathogen with strategies to evade host immunity and overcome antimicrobial treatment (Fig. 1).

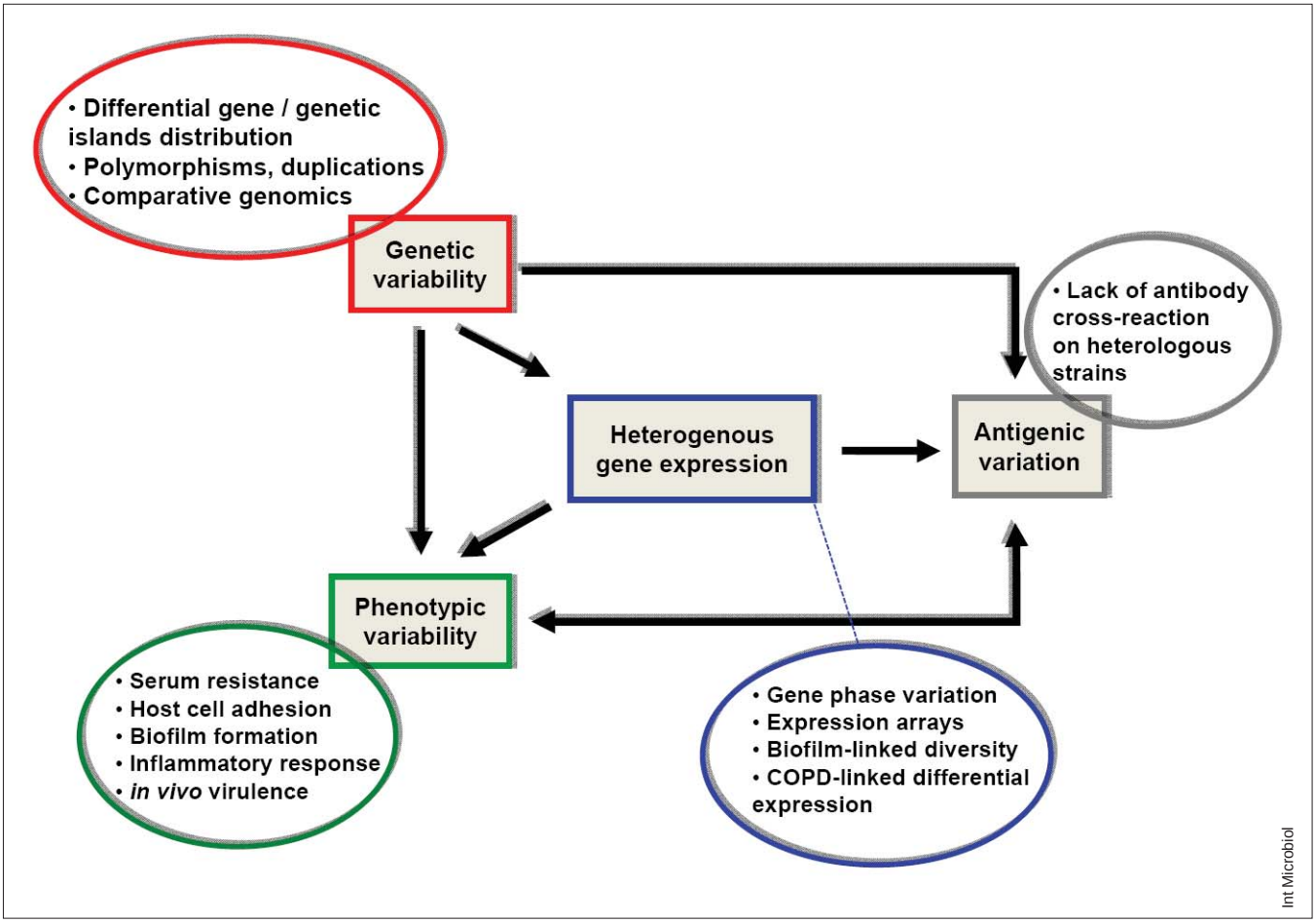

Fig. 1. Diagram of three different domains conferring diversity in *Haemophilus influenzae*: genetic variability, heterogeneous gene expression and phenotypic variability. The three diversity domains are intimately cross-related and lead to the generation of antigenic variation. Non-exhaustive examples of the available experimental evidence for each of these domains are provided.

NTHi genetic diversity: gene distribution/sequence conservation and genome-sequencing-based approaches

Haemophilus influenzae strain Rd KW20 was the first free-living organism from which a complete genome sequence was obtained, and the resulting information provided an excellent scaffold to assess *H. influenzae* diversity [19]. Nontypable strains of *H. influenzae* were long considered as colonizing bacteria whose virulence potential largely reflected alterations in host defenses. However, growing evidence based on NTHi recovered from disease states suggests that these bacteria are genotypically different, both in terms of disease state and compared to strains harvested from healthy carriers [25,52,55,73]. These observations raise several questions: Can we identify genes/genomic regions important for NTHi virulence by comparing the genetic makeup of strains recovered from disease with strains isolated from healthy carriers? Would this virulence-associated genetic material allow strain stratification or the development of tools to predict NTHi virulence? The answers have been sought mainly by analyzing the differential distribution of limited numbers of genes or genetic traits among NTHi isolates, and more recently, by comparative genomics of sequenced strains.

Gene distribution/sequence conservation among NTHi isolates. Explorations of NTHi genetic diversity have mainly been carried out using a reductionist approach, based on the survey of selected genes or genetic islands on isolate panels. The aim of these studies has been the identification of virulence factors, genetic markers for NTHi differentiation from other bacteria, and useful epitopes as vaccine candidates. Gene distribution assessment has focused on

Table 1. Sources of genetic variability among noncapsulated *Haemophilus influenzae* strains

Gene	Variable distribution	Phase variation	Allelic polymorphisms
lic2C	Yes	No	ND*
lic2B	Yes	No	ND
losAB	Yes	Yes, 5′-CGAGCATA in *losA*	ND
lic3A	No	Yes, 5′-CAAT	ND
lic3B	Yes	Yes, 5′-CAAT	ND
lic1A	Yes	Yes, 5′-CAAT	ND
lic1D	Yes	No	Yes
lic2A	No	Yes, 5′-CAAT	ND
lgtC	No	Yes, 5′-GACA	ND
oafA	No	Yes, 5′-GCAA	ND
lex2A	Yes	Yes, 5′-GCAA	ND
lex2B	Yes	No	Yes
hmw1A	Yes	Yes, 5′-ATCTTTC	Yes
hmw2A	Yes	Yes, 5′-ATCTTTC	Yes
hia	Yes	No	Yes
hifABCDE	Yes	Yes, 5′-TA	Yes
hap	No	No	Yes
ompP5	No	No	Yes
oapA	No	No	Yes
igaB	Yes	Predicted in strain 2019, 5′-AAATTCA	Yes

*ND, not determined.

genes encoding NTHi surface molecules, including lipooligosaccharid (LOS) as well as adhesive and immunomodulatory molecules. Table 1 provides a list of genes that are variable on NTHi strains, and their sources of variability.

The NTHi LOS is a glycolipid comprising a membrane-anchoring lipid A linked by a single 2-keto-3-deoxyoctulosonic acid (Kdo) to a heterogeneous oligosaccharide (OS) composed of neutral heptose (Hep) and hexose (Hex) sugars, lacking an O antigen [60]. Each Hep of a conserved trisaccharide (HepI to HepIII) inner core can serve as a point for Hex addition and further chain extensions, the degree and pattern of which vary among strains [60]; a fourth heptose (HepIV) may be present on the OS extension from HepI [37] (Fig. 2). Several genes involved in LOS biosynthesis are variably present among *H. influenzae* strains. This is the case for *li2BC* and *losAB* [16,17,36]. The *lic2C* and *lic2B* genes encode glycosyltransferases responsible for initiating sugar extension from HepII [36] and for adding the second sugar (Glc or Gal) to the Glc on HepII, respectively [65]. The *losB* gene encodes a heptosyltransferase responsible for adding HepIV to the OS on HepI, and *losA* encodes another glycosyltransferase [37]. When present, *lic2C* is located in a genetic island flanked by *infA* and *ksgA*. The *infA-ksgA* island can be absent, with the *infA* and *ksgA* adjacent to each other, or present, containing

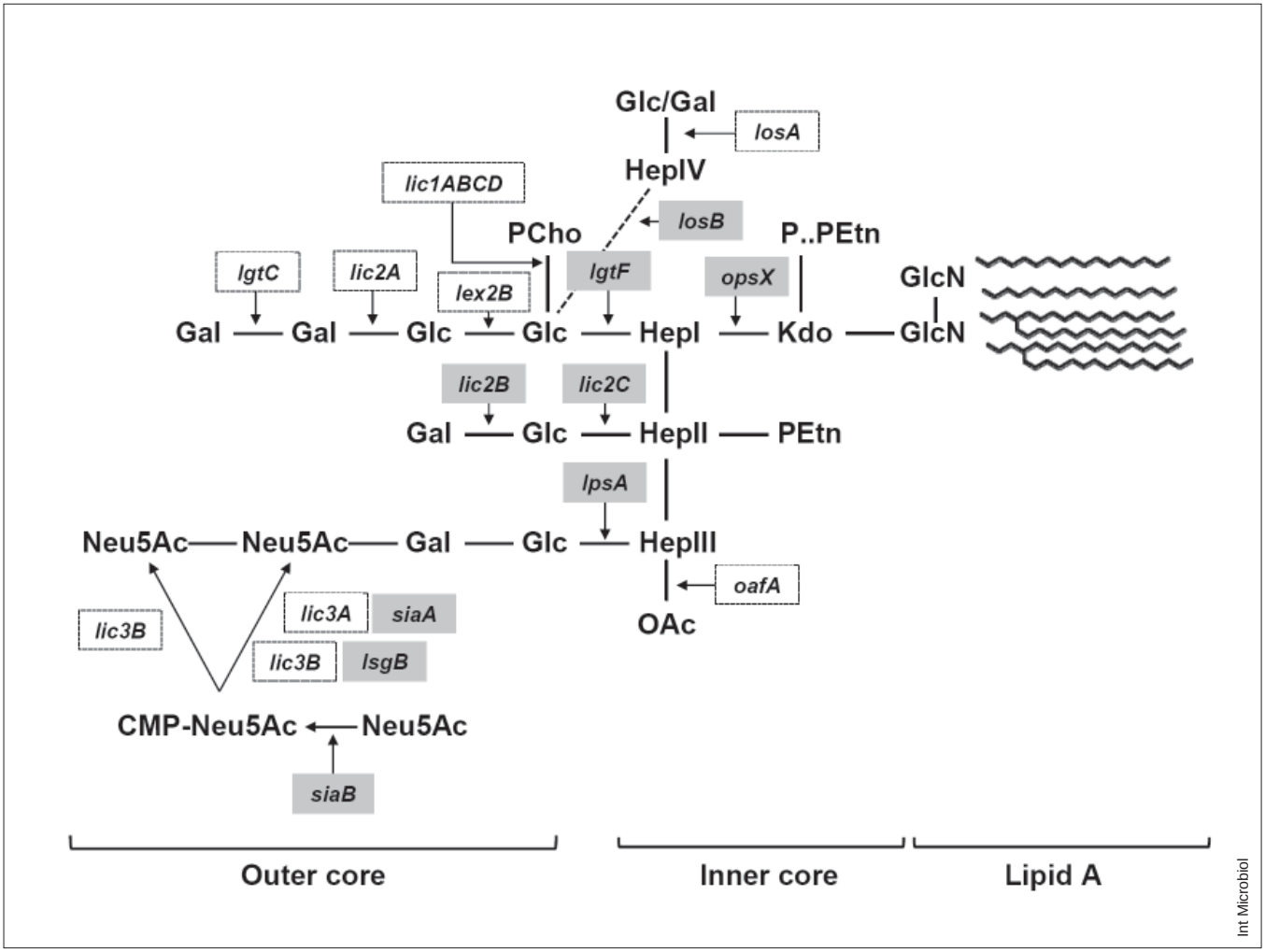

Fig. 2. Model structure of NTHi lipooligosaccharide (LOS). A repertoire of modifications, whose presence and location are variable among strains, is shown. GlcN, glucosamine; Kdo, 2-keto-3-deoxyoctulosonic; PEtn, phosphoethanolamine; Hep, heptose; Glc, glucose; Gal, galactose; Neu5Ac, sialic acid; PCho, phosphorylcholine; OAc, O-acetyl group. Genes encoding enzymes responsible for the biosynthesis of the LOS molecule are indicated. Phase-variable genes are shown in white; non-phase-variable genes are shown in gray.

(i) *lic2C*, (ii) *lic2B* and *lic2C*, or (iii) *losA* and *losB* [17]. A comparison between invasive NTHi isolates obtained from the host middle ear and nasopharynx/throat revealed that this island is present in most nasopharyngeal and OM isolates but absent from 40 % of invasive isolates [17].

A survey of *lic2C* from a collection of NTHi inner ear-OM isolates showed the presence of the gene in approximately half of the analyzed strains [36]. A later study from our laboratory on a panel of non-isogenic NTHi isolates of different pathological origin showed a 95 % prevalence of *lic2C*, suggesting that it encodes a molecular feature conferring bacterial fitness during infectious processes [44]. Support for this observation comes from an analysis of *lic2C* distribution in a strain panel encompassing 54 NTHi strains collected from 20 adults suffering an underlying chronic respiratory disease. The patients were seen at a tertiary reference center (University Hospital Bellvitge, Spain) between two and five times from 1996 to 2007. Strain molecular typing by pulse-field gel electrophoresis (PFGE) indicated a high diversity (45 PFGE different profiles). Patients were classified as follows: Group A, consisting of 14 patients in whom each of the collected strains differed from the others with respect to the PFGE profile; and Group B, consisting of six patients, among whom at least two of the strains collected per patient displayed the same PFGE profile.

Collectively, *lic2C* was detected in 63 % of the isolates. Additional data from our laboratory suggested that *lic2C* is not necessarily linked to virulence, but, more generally,

to bacterial host adaptation. Evidence for this notion was obtained in an analysis of *lic2C* in a panel of 42 isolates encompassing 25 NTHi strains collected from 25 pediatric patients with OM (University Hospital Germans Trias i Pujol, Spain) and 17 NTHi nasopharyngeal isolates from 17 healthy children (University Hospital Bellvitge), in which the gene had a prevalence of 76 % and 94 %, respectively.

The linkage of *lic2B* with *lic2C* has been reported [17], with several studies addressing *lic2B* distribution and the gene's prevailing presence in middle ear-OM isolates [55,71,73]. Our data on *lic2B* distribution within the panel of 42 NTHi pediatric strains described above confirms an association between *lic2B* and *lic2C*, given that *lic2B* was only detected in *lic2C*-positive strains. Among OM patients and healthy carriers, the prevalence of *lic2B* was 56 % and 47 % , respectively; among the *lic2C*-positive isolates, the prevalence of *lic2B* was 73 % and 50 %, respectively. These data slightly differ from those previously reported, as the prevalence of *lic2B* among healthy carrier isolates was somewhat higher, which could be due to the origin, size, or nature of the strain panels. Nonetheless, they suggest the general involvement of *lic2BC* in NTHi-host interplay, rather than its exclusive role in virulence. Unlike *lic2BC*, the presence of *losAB* seems to be scattered, based on the gene's detection in only three *lic2BC*-negative strains in the same panel of 42 pediatric isolates. Similarly, a previous evaluation of *losAB* in two collections of NTHi clinical isolates yielded a prevalence of 16 % and 18 %, respectively [16,17]. The phase variation of *losA* is an additional source of variability [16].

Sialylation, catalyzed by the sialyltransferases Lic3A, Lic3B, SiaA, and LsgB, is another variable modification of NTHi LOS. Although the *lic3A* gene seems to be universally present, a survey of *lic3B* on a collection of NTHi inner-ear isolates identified *lic3B* in 60 % of the strains [20]. However, a later study from our laboratory on a panel of non-isogenic NTHi isolates of different pathological origin showed the 100 % prevalence of *lic3B* [44], and an assessment of the gene on the above-discussed panel of 42 pediatric isolates found a 72 % and 100 % prevalence of *lic3B* among OM and healthy carriers, respectively. An additional source of variation in LOS sialylation is *lic3A* and *lic3B* phase variation [20].

The *lic1* locus, encompassing the *lic1ABCD* operon, is responsible for the addition of phosphorylcholine (PCho) to LOS [68]. A survey of a collection of NTHi isolates detected *lic1A* in 96 % of the strains [45]; a later study from our laboratory on a panel of non-isogenic NTHi isolates of different pathological origin found a 100 % prevalence for *lic1D* [44]. PCho substitutions may occur on OSs extending from any Hep, depending on the *lic1D* allele (*lic1D*$_I$, *lic1D*$_{III}$, *lic1D*$_{IV}$), which encodes a diphosphonucleoside choline transferase [43,45]. Moreover, although most strains have a single *lic1D* gene, a survey of NTHi strains collected from the middle ear found that 16 % of them had two *lic1D* alleles, each in a separate, phase-variable *lic1* locus, which together could result in two PCho substitutions in the LOS of the respective strain [21].

Available information on the heterogeneous distribution of additional genes involved in NTHi OS extensions, such as *lpsA*, *lic2A*, *lgtC*, and *oafA*, suggests that, although extensively present in NTHi strains [15,22,36,44], these genes are not necessarily conserved; for example, *lic2A*, *lgtC* and *oafA* are phase variable [15,22,32]. Moreover, allelic polymorphisms have been found in *lpsA*. This gene encodes a glycosyltransferase responsible for the addition of a Hex to HepIII; the added Hex can be either Glc or Gal, and Hep linkage can be either β 1-2 or β 1-3. Each *H. influenzae* strain produces only one of the four possible combinations of linked sugars in its LOS, due to a specific allelic variant of *lpsA* directing both linkage and the added Hex, Glc, or Gal [10]. Variable distribution, allelic polymorphisms, and phase-variable expression also characterize the *lex2* locus. The *lex2A* gene contains a variable number of 5′-GCAA repeats; *lex2B* encodes the glucosyltransferase that adds the second Hex during the extension of LOS by HepI [28]. Allelic polymorphisms are assumed for *lex2B*, based on the alteration of a single amino acid in Lex2B, which correlates with the addition of Glc or Gal to the OS extension from HepI [9].

Variable distribution has also been evaluated on genes encoding adhesive molecules. Thus, the distribution of the adhesin-encoding genes *hmw* and *hia* in a panel of 59 non-capsulated strains showed that 47 strains contained *hmw1* and *hmw2* while nine strains contained *hia*, but no strain harbored both *hmw* and *hia* [63]. Based on the available evidence: (i) all strains having *hmw* genes contain two *hmw* loci in conserved unlinked physical locations on the chromosome [5]; (ii) *hmw* genes occur in different allelic versions among strains [5,13]; and (iii) *hmw* genes are more prevalent in isolates associated with acute OM than in the throat isolates of healthy children [14,39,73]. An additional source of diversity is the phase variation of both *hmw1A* and *hmw2A* [8]. Although it has not been formally analyzed, *hia* may present polymorphisms, since its PCR amplification in two panels of clinical strains rendered variable size products [17,59].

The prevalence of the phase-variable *hifABCDE* gene cluster, responsible for the biosynthesis of the hemagglutinating pili, seems to be generally low [3,24], with a higher

prevalence among Hib than among NTHi isolates [14]. Discrepancies among independent studies do not allow a clear association between anatomic isolation site (throat or middle ear) and *hifABCDE* distribution [14,64].

According to current information, the adhesin-encoding genes *hap*, *ompP5*, and *oapA* are universally distributed among noncapsulated isolates, but they display variation. Thus *hap*, encoding a self-associating autotransporter involved in intercellular aggregation [62], has a stop codon in strain Rd KW20, and its PCR amplification results in products of different sizes among clinical isolates (B. Euba, personal communication). The *ompP5* gene, encoding an outer membrane protein involved in bacterial adhesion to host cell surfaces [34], is highly variable among strains [12,49]. Although its amplification product was size-invariable among non-isogenic strains of different pathological origin, variability was detected in the five extracellular loop domains predicted for P5 by PRED-TMBB analysis [44]. Despite the heterogeneity of *ompP5*, a P5 sequence comparison in two separate isolate panels containing sets of identical strains recovered from patients with a chronic respiratory disease who were seen in independent medical visits showed no differences among identical strains ([49], A. López-Gómez, personal communication), pointing to the relative stability of P5 during NTHi persistence in the host. Conversely, the *oapA* amplification product is size variable, due to insertions/deletions in the gene region encoding the protein segment starting at amino acid 195 [44].

The *iga* gene, encoding an antigenically variable IgA1 protease, is extensively distributed among strains [42]. However, compared to strains from other clinical sources, genomes of isolates from adults with COPD have a higher likelihood of also having *igaB*, encoding a second IgA1 protease [52]. A sequence analysis of *igaB* showed minor sequence changes among isolates [52].

Collectively, variability studies based on a limited number of genes may facilitate associations between genes/gene groups and disease manifestation or bacterial anatomic location, which in turn could reveal virulence factors and provide tools to predict virulence. However, gene selection, the number of selected genes, and the nature and size of the strain collections, are critical limiting factors that must be considered to obtain useful information. Comparative analysis of panels of whole-genome sequenced strains is a powerful approach that may contribute significantly to overcome these limitations.

Whole-genome multiple-strain sequencing. Sequenced strain Rd KW20 was useful in understanding the basic biology of *H. influenzae*, but it did not provide significant insight into disease because is a rough derivative of *H. influenzae* serotype d, which is rarely disease-associated [31]. Nonetheless, the elucidation of differences between the genomes of strains isolated from disease states and the genome of strain Rd KW20 may yield insight into NTHi pathogenicity. Thus, an analysis of NTHi strain 86-028NP, isolated from a patient with chronic OM, revealed large rearrangements in its genome architecture compared to strain Rd KW20, in addition to the presence of 280 ORFs not present in the latter strain [30]. Since then, further studies have provided increasing information on the *H. influenzae* core- and pan-genome. A comparative genomic study of strain Rd KW20 and 12 NTHi clinical isolates identified 2786 genes, of which 1461 were common to all strains. That study allowed the development of a finite supra-genome model in which a NTHi supra-genome containing between 4425 and 6052 genes was predicted [35]. A recent study sought to identify bacterial genetic elements with increased prevalence among strains isolated from COPD patients, compared to those from healthy carriers. Two NTHi strains recovered from the airways of two COPD patients and two strains from a healthy individual were sequenced. Seven genetic islands were defined, with their distribution among a panel of 421 strains of both disease and commensal origins revealing that four of these islands were more prevalent in COPD than in colonizing strains [73]. Whole-genome sequencing on *H. influenzae* has also been applied to study the impact of transformation-mediated homologous recombination in inter-strain exchange of DNA [46,57]. Indeed, *H. influenzae* rendered the first genome-wide analysis of chromosomes directly transformed with DNA from a divergent genotype [46].

Heterogeneity in gene expression and its contribution to NTHi strain stratification

The presence or absence of a gene is not necessarily indicative of the infection outcome, as the same gene may be found in asymptomatically carried strains but with slight genetic changes or differences in expression. NTHi differential gene expression has been mainly explored in phase-variable genes. The *lic1ABCD* operon is phase variably expressed due to a 5′-CAAT repeat within the *lic1A* reading frame [68]. Differential PCho expression has been reported among NTHi isolates [44] and may vary depending on the anatomic location in the host. In fact, *H. influenzae* variable PCho expression may correlate with the ability of the bacterium to persist on the mucosal

surface ($PCho^+$ phenotype), and to cause invasive infection by evading innate immunity mediated by acute-phase C-reactive protein ($PCho^-$ phenotype) [67]. The *losA* gene is phase variably expressed due to a 5′-CGAGCATA repeat within the reading frame. Of 30 NTHi strains containing *losA*, 24 had two tandem copies of the SSR, allowing full-length translation of *losA* (on), and six had 3, 4, 6, or 10 tandem copies (*losA* off). The expression of *losA*, which is determined by the variations in its repeats, has been shown to affect NTHi resistance to serum-mediated killing [16].

Similarly, *lic3A* and *lic3B*, encoding two sialyltransferases, are phase variably expressed due to a 5′-CAAT repeat within their reading frames. The number of repeated motifs in 25 NTHi isolates was found to vary from 14 to 41 in *lic3A* and from 12 to 28 in *lic3B*; for both genes, two of the three possible reading frames were predicted to allow translation of full-length gene products from alternative initiation codons upstream of the repeats [20]. The *lic2A* galactosyltransferase-encoding gene is variably expressed due to a 5′-CAAT repeat within its reading frame [32]. The number of repeated motifs within *lic2A* varied between 7 and 33 in a group of 19 NTHi isolates [44]. The repeated tract of *lic2A* is preceded by four putative initiation codons in two reading frames [11]. Fifteen of those 19 isolates contained an in-frame *lic2A* gene [44]. Independently, in an SSR analysis of *lic2A* using the above-described panel of *H. influenzae* isolates collected from adult patients suffering an underlying chronic respiratory disease, the number of repeated motifs within *lic2A* in 28 of those isolates varied between 7 and 28. Sequence comparison from sets of identical strains recovered from the above-described group B patients demonstrated diversity in the number of *lic2A* repeats among identical strains over time. Digalactose has been linked to NTHi resistance to serum-mediated killing [15] and to virulence [27]. Evaluation of *hmw1A* and *hmw2A* gene expression in three NTHi invasive isolates and in the prototype strain 12 showed that increased numbers of 5′-ATCTTTC repeats within the *hmwA* promoters correlate with decreased amounts of transcript [26]. In agreement with this finding, an analysis of HMW1 and HMW2 adhesins in isolates collected serially from COPD patients revealed that the expression of both proteins by a given strain decreased over time in the majority of patients, reflecting a progressive increase in the numbers of 7-bp repeats [7].

Microarray studies comparing gene expression among isolates have provided evidence for a conserved core of genes preferentially expressed during *H. influenzae* growth in iron/heme-restricted condition [69]. Differential expression of surface molecules between bacteria grown planktonically or forming biofilms demonstrated a greater abundance of peroxi–redoxin-glutaredoxin in *H. influenzae* biofilms than in planktonically grown bacteria. This molecule is involved in biofilm formation by *H. influenzae* and the degree of its involvement varies among strains; note that peroxiredoxin-glutaredoxin is recognized by the human immune system *in vivo*, which suggests its expression by *H. influenzae* during human respiratory tract infection [51]. LRT isolates associated with COPD exacerbation are more resistant to the bactericidal effect of serum than colonizing isolates from the upper airway, with the expression of *vacJ* and *yrb* positively correlating with serum resistance. The *vacJ* gene functions with an ABC transporter encoded by *yrb* in the retrograde trafficking of phospholipids from the outer to the inner leaflet of the cell envelope, suggesting that NTHi adapts to inflammation encountered during LRT infection by modulating its outer leaflet through the increased expression of *vacJ* and *yrb*, thereby minimizing recognition by bactericidal anti-OS antibodies [53].

Collectively, existing data reinforce the notion that the heterogeneous expression of genes involved in NTHi virulence should be considered and integrated in studies of bacterial diversity, as this may be a useful basis for stratifying the virulence potential of clinical isolates and/or identifying potential therapeutic targets.

Variable phenotypes among NTHi isolates and differential bacterial interplay with host immunity

Genetic traits may be ultimately of little interest unless they can be associated with virulence. However, a clear-cut relationship between virulence-linked genotype and phenotype remains elusive for NTHi. This gap could be due to: (i) the absence of clearly defined phenotypes that can differentiate among NTHi strains with and without virulence potential; (ii) the absence of systematic comparative phenotypic studies using a significant number of isolates recovered from different disease states and from healthy carriers; and (iii) the lack of studies in which both genotypic and phenotypic traits are simultaneously analyzed in wide strain panels.

An assessment of the phenotypic diversity of NTHi pointed out the differential interplay of host immunity elements and the various isolates. Variable resistance to serum-mediated killing among panels of NTHi isolates recovered from the pediatric inner ear and of non-isogenic NTHi isolates from different pathological origin suggested an association between LOS sialylation and NTHi resistance to

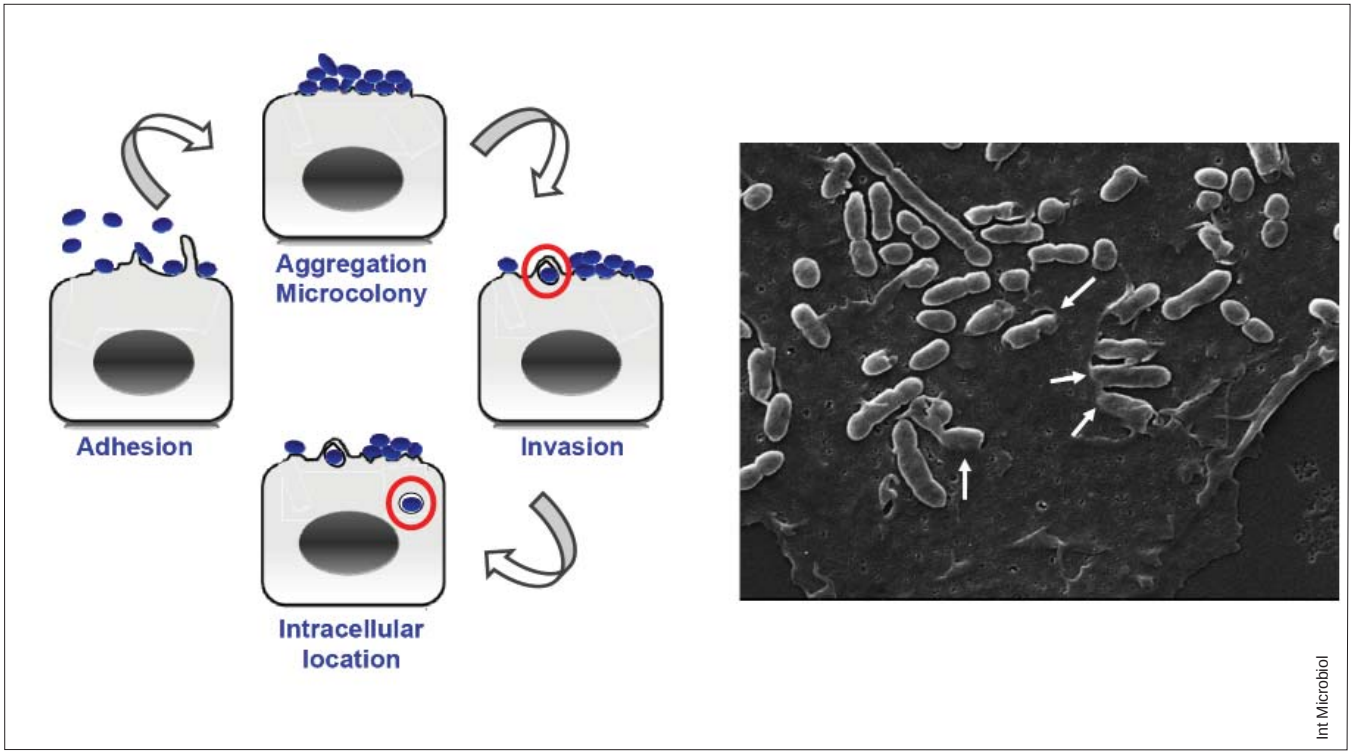

Fig. 3. Model presenting key features of epithelial cell infection by *Haemophilus influenzae*. Bacteria adhere to the cell surface. Once bacteria have adhered, inter-bacterial interactions lead to microcolony formation. Microcolony formation facilitates bacterial invasion into epithelial cells, potentially providing a protected niche and allowing bacterial evasion of host immunity. Rigth: a scanning electron micrograph shows NTHi infection of A549 immortalized human type II pneumocytes; the white arrows point at the attachment of bacteria to the host cell surface. Image courtesy of Dr. José Ramos Vivas, Fundación Marqués de Valdecilla, Santander, Spain.

complement. This finding was supported by the high serum-susceptibility displayed by a mutant strain lacking the CMP-synthetase *siaB* gene [38,44]. In addition, serum resistance of *losAB*-containing strains has been correlated with an on-vs. off-state of *losA* [16]. However, an attempt to establish serum resistance as a virulence trait potentially shared by invasive noncapsulated *H. influenzae* strains did not render conclusive results [17]. Similarly, there was no clear difference in serum resistance or binding to complement inhibitors between invasive NTHi isolates obtained from patients with sepsis and nasopharyngeal strains obtained from patients with upper respiratory tract infection [29], although a significant correlation between disease severity and serum resistance was identified in cases of NTHi invasive disease [29].

Evidence points out that *H. influenzae* interplay with the respiratory epithelium involves bacterial adherence to epithelial cells and inter-bacterial interactions leading to microcolony formation. Microcolony formation may lead to the establishment of a biofilm resistant to host immune factors. Attachment promotes bacterial invasion into epithelial cells, potentially providing a protected niche that may allow bacterial evasion from local immune mechanisms (Fig. 3). Adhesion to epithelial host cell surfaces [7,44] and biofilm formation [50] are also heterogeneous features of NTHi. Of note, significant phenotypic differences between NTHi strains from COPD exacerbation and colonizers have been reported, with the former strains having greater adherence to airway epithelial cells and inducing more severe airway inflammation [6]. Another variable phenotypic trait is the antigenic variability of surface-exposed epitopes, evidenced by the development of new highly strain-specific bactericidal antibodies after exacerbation; these antibodies show low bactericidal activity for heterologous strains [61].

While a significant correlation between disease phenotype and global comparative genomic data would facilitate the stratification of isolates and our ability to predict disease manifestations, this goal remains elusive. In an *in vivo* chinchilla model of OM aimed at characterizing the local and systemic virulence patterns of ten genomically analyzed NTHi isolates from children with chronic OM with effusion or with otorrhea, strain stratification was indeed possible, but global comparative genomics of the same strains did not cluster them

by clinical phenotype [4]. Although several reasons could explain this inaccuracy, the wide genetic diversity among strains is particularly probable, given that genome sequence comparison has shown that the mean number of gene differences among each of the possible strain pairs are >350, but the number of genes associated with each parameter of clinical virulence may be a small fraction thereof [4].

In summary, the wide genetic and phenotypic variability among NTHi strains highlights the need to explore alternative approaches to facilitate the association of genotype with phenotype.

Current questions and challenges for future studies on the diversity of non-capsulated *Haemophilus influenzae*

Pathogenicity is the result of the relationship between a bacterium and its host, specifically, between bacterial virulence factors, including how and when they are expressed, and the host immune status. The latter is determined by genetic factors, age, lifestyle, co-infections, and exposure to external agents, all of which can modulate host physiology and the ability to fight infection.

Host factors in the dynamics of NTHi infection. Defining the role of host immunity in disease outcome is crucial; indeed, pathogen diversity studies should ideally be conducted in parallel with immunological studies on the respective host. This aspect may be particularly crucial for highly adapted and very flexible opportunistic pathogens such as *H. influenzae*, for which host immunological status is a strong determinant in the ability of a pathogen to cause symptomatic disease in a previously asymptomatic healthy carrier.

An example of this notion is the association between NTHi infection and the progression of COPD. Patients with COPD sufffer from chronic bronchitis, emphysema, or both. In these diseases, the airways become narrowed, which leads to an irreversible limitation of airflow to and from the lungs, causing shortness of breath [2]. COPD is caused by airway exposure to noxious particles or gas, most commonly from tobacco smoking, which triggers an abnormal inflammatory response in the lung. These deleterious agents impair normal respiratory function and alter the host's response to infection by opportunistic pathogens such as NTHi, which colonizes the upper airways, causes chronic LRT infection, and is frequently isolated in disease exacerbation [23]. Prospective comparative genotype and phenotype analyses of multiple NTHi isolates serially recovered from the upper and lower airways of COPD patients in stable and acute condition, together with detailed clinical, inflammatory, and patho-physiological information obtained from those patients at the time of each microbial isolation, would provide invaluable biological material and information with which to assess microbial evolution. It would also faciliate the design of tools to predict disease severity, the virulence potential of a bacterial strain, and the outcome of the host-pathogen encounter.

Virulence vs. niche factors and NTHi adaptation vs. infection. Given that NTHi is highly adapted to the human respiratory microbiota, it is likely to be equipped with evasion strategies allowing the bacterium to endlessly colonize the host. Evidence demonstrating differential gene distribution between strains isolated from different body locations and/or disease states supports the existence of genetic traits associated with disease [73]. However, an increase in the number, size, and clinical and geographical diversity of the strain panels screened may dilute the relevance of those proposed genetic virulence traits due to their extensive presence in healthy carrier isolates. Instead, they may prompt us to consider the fine line between virulence, adaptation, and genetic fitness for NTHi. This consideration should not limit the potential of currently identified genetic virulence traits, which could well be involved in both the colonization of healthy hosts and the symptomatic infection of immunocompromised individuals. In fact, many structures and strategies playing important roles in establishing and maintaining infection have been discovered and characterized in pathogens. However, these virulence factors can also be shared by commensals because they are required for their existence in the host, thus suggesting their re-consideration as niche factors [33].

Our current understanding of the role of NTHi virulence factors is in part based on lack-of-function mutant strains generated in the laboratory, when assayed for phenotypes linked to virulence. This approach, essential for gene–function associations, nonetheless has certain risks that must be taken into account in any discussion of the resulting data, given that: (i) there is often a reliance on reference strains that can be mutated under laboratory conditions, which can generate strain-dependent bias; (ii) functional redundancy is frequently not considered, although it could be a source of bias in the form of single-mutant strain-dependent data. Moreover, the relevance of so-called virulence phenotypes in the refined adaptation and colonization of the human host by NTHi cannot be excluded. Indeed, for NTHi, the precise

definition of phenotypes that clearly differentiate virulent and colonizing strains may be risky, as the difference may actually depend on host status. Further experimental evidence is required to address these issues. The widest possible repertoire of virulence phenotypes should be systematically assayed on vast collections of genotypically characterized NTHi strains, recovered from healthy carriers and from different disease states, in order to cluster phenotypes into categories and to define virulence and/or adaptation indexes.

The challenges of genomic information in the study of NTHi diversity. In general, comparative genomics of microbial pathogens aims to predict the virulence potential of a bacterial strain from its genome sequence [58]. Sequencing can identify which virulence factor-encoding genes are present in a genome. However, the presence of these genes in itself is not indicative of disease outcome, as the same gene might well be found in asymptomatically carried strains. Therefore, without an understanding of the regulatory and epistatic processes controlling gene expression, the contribution of a list of genes to virulence cannot be quantified. A systems biology approach based on a comprehensive understanding of the combinations of genetic backgrounds, regulatory networks, and virulence factors that produce virulent strains has been proposed to help researchers determine the propensity of a particular strain to cause disease.

The goals of the proposed framework are: (i) to define phenotypes that differentiate virulent and avirulent strains; (ii) to characterize how the relevant phenotypes are encoded, using expression arrays to construct models of the gene-regulatory networks as well as process diagrams informed by the underlying genetics; (iii) to develop models that predict the gene combinations leading to specific virulence phenotypes; and (iv) to test and refine the models with sets of strains independent from those used to build the model [58]. Although tailored to *Staphylococcus aureus*, mounting information on *H. influenzae* diversity may provide the necessary conditions to apply this type of framework to the prediction of virulence phenotypes using *H. influenzae* genome sequences.

Laboratory experiments have led to important findings relating organism adaptation to genomic evolution. Continuous monitoring of long-term evolution in natural systems is expanding our knowledge of these processes *in situ*. We highlight here two exemples. Thus, the evolutionary dynamics of a lineage of *Pseudomonas aeruginosa* as it adapted to the airways of several individual CF patients over 200,000 bacterial generations has been reported. In contrast to predictions based on *in vitro* evolution experiments, the evolving lineage showed limited diversification, in which an initial period of rapid adaptation caused by a small number of mutations with pleiotropic effects was followed by a period of genetic drift with limited phenotypic change and a genomic signature of negative selection. This pattern suggests that the evolving lineage reached a major adaptive peak in the fitness landscape [72]. Independently, in a retrospective study of a *Burkholderia dolosa* outbreak among CF patients, the genomes of 112 isolates collected from 14 individuals over 16 years were sequenced. Seventeen of the bacterial genes had acquired non-synonymous mutations that were detected in multiple individuals, indicating parallel adaptive evolution. Importantly, these mutations shed light on the genetic basis of pathogenic phenotypes [40]. NTHi acute and chronic infection has been associated with the progression of cigarette-smoke-related diseases such as COPD, which suggests the ability of this pathogen to adapt to a human niche rich in free radicals and other aromatic compounds present in smoke. This type of disease state offers a unique natural system for continuous monitoring of the long-term evolution of *H. influenzae* in the upper and lower airways of humans.

Final remarks

Its relatively small genome size and wide genetic plasticity, together with its asymptomatic colonizer–virulence duality and prominent association with chronic respiratory diseases make noncapsulated *H. influenzae* a unique bacterial system for studies of microbial adaptation, pathogenesis, and long-term microbial evolution in human hosts exposed to external deleterious agents. Access to comprehensive strain panels and detailed clinical data from the respective hosts, when combined with extensive whole-genome sequencing and systematic phenotypic analysis in large number of isolates, will provide extensive insights into NTHi pathogenesis as well as both the tools to predict virulence and information on bacterial evolution and adaptation. Now that microbial whole-genome sequencing is becoming routine in diagnostic and public-health microbiology, this may be the right time to tackle detailed studies of the opportunistic pathogen nontypable *H. influenzae*.

Acknowledgements. We thank Drs. Cristina Prat (Germans Trias i Pujol Hospital) and Josefina Liñares (University Hospital Bellvitge) for providing strains, Dr. Laura Calatayud for helping with PFGE and clinical data, and Dr. Pau Morey for helpful reading of the manuscript. This work has been funded by grants from the Health Institute Carlos III (ISCIII), grant PI09/00130, and from the Health Department of the Government of Navarra, Spain (Call 2011) to J.G., and by grant PI09/01904 (ISCIII) to J. Liñares. CIBERES is an initiative from ISCIII, Spain.

Competing interest. None declared.

References

1. Agrawal A, Murphy TF (2011) *Haemophilus influenzae* infections in the *H. influenzae* type b conjugate vaccine era. J Clin Microbiol 49:3728-3732
2. Barnes PJ (2004) Mediators of chronic obstructive pulmonary disease. Pharmacol Rev 56:515-548
3. Bayliss CD, Sweetman WA, Moxon ER (2004) Mutations in *Haemophilus influenzae* mismatch repair genes increase mutation rates of dinucleotide repeat tracts but not dinucleotide repeat-driven pilin phase variation rates. J Bacteriol 186:2928-2935
4. Buchinsky FJ, Forbes ML, Hayes JD, Shen K, Ezzo S, Compliment J, Hogg J, Hiller NL, Hu FZ, Post JC, Ehrlich GD (2007) Virulence phenotypes of low-passage clinical isolates of nontypeable *Haemophilus influenzae* assessed using the chinchilla laniger model of otitis media. BMC Microbiol 7:56
5. Buscher AZ, Burmeister K, Barenkamp SJ, St Geme JW 3rd (2004) Evolutionary and functional relationships among the nontypeable *Haemophilus influenzae* HMW family of adhesins. J Bacteriol 186:4209-4217
6. Chin CL, Manzel LJ, Lehman EE, Humlicek AL, Shi L, Starner TD, Denning GM, Murphy TF, Sethi S, Look DC (2005) *Haemophilus influenzae* from patients with chronic obstructive pulmonary disease exacerbation induce more inflammation than colonizers. Am J Respir Crit Care Med 172:85-91
7. Cholon DM, Cutter D, Richardson SK, Sethi S, Murphy TF, Look DC, St Geme JW 3rd (2008) Serial isolates of persistent *Haemophilus influenzae* in patients with chronic obstructive pulmonary disease express diminishing quantities of the HMW1 and HMW2 adhesins. Infect Immun 76:4463-4468
8. Dawid S, Barenkamp SJ, St Geme JW 3rd (1999) Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. Proc Natl Acad Sci USA 96:1077-1082
9. Deadman ME, Hermant P, Engskog M, Makepeace K, Moxon ER, Schweda EK, Hood DW (2009) Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. Infect Immun 77:2376-2384
10. Deadman ME, Lundstrom SL, Schweda EK, Moxon ER, Hood DW (2006) Specific amino acids of the glycosyltransferase LpsA direct the addition of glucose or galactose to the terminal inner core heptose of *Haemophilus influenzae* lipopolysaccharide via alternative linkages. J Biol Chem 281:29455-29467
11. Dixon K, Bayliss CD, Makepeace K, Moxon ER, Hood DW (2007) Identification of the functional initiation codons of a phase-variable gene of *Haemophilus influenzae*, *lic2A*, with the potential for differential expression. J Bacteriol 189:511-521
12. Duim B, Bowler LD, Eijk PP, Jansen HM, Dankert J, van Alphen L (1997) Molecular variation in the major outer membrane protein P5 gene of nonencapsulated *Haemophilus influenzae* during chronic infections. Infect Immun 65:1351-1356
13. Ecevit IZ, McCrea KW, Marrs CF, Gilsdorf JR (2005) Identification of new *hmwA* alleles from nontypeable *Haemophilus influenzae*. Infect Immun 73:1221-1225
14. Ecevit IZ, McCrea KW, Pettigrew MM, Sen A, Marrs CF, Gilsdorf JR (2004) Prevalence of the *hifBC*, *hmw1A*, *hmw2A*, *hmwC*, and *hia* genes in *Haemophilus influenzae* isolates. J Clin Microbiol 42:3065-3072
15. Erwin AL, Allen S, Ho DK, Bonthuis PJ, Jarisch J, Nelson KL, Tsao DL, Unrath WC, Watson ME Jr., Gibson BW, Apicella MA, Smith AL (2006) Role of *lgtC* in resistance of nontypeable *Haemophilus influenzae* strain R2866 to human serum. Infect Immun 74:6226-6235
16. Erwin AL, Bonthuis PJ, Geelhood JL, Nelson KL, McCrea KW, Gilsdorf JR, Smith AL (2006) Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide biosynthetic gene *losA* affects serum resistance. Infect Immun 74:3408-3414
17. Erwin AL, Nelson KL, Mhlanga-Mutangadura T, et al. (2005) Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. Infect Immun 73:5853-5863
18. Farjo RS, Foxman B, Patel MJ, Zhang L, Pettigrew MM, McCoy SI, Marrs CF, Gilsdorf JR (2004) Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. Pediatr Infect Dis J 23:41-46
19. Fleischmann RD, Adams MD, White O, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496-512
20. Fox KL, Cox AD, Gilbert M, Wakarchuk WW, Li J, Makepeace K, Richards JC, Moxon ER, Hood DW (2006) Identification of a bifunctional lipopolysaccharide sialyltransferase in *Haemophilus influenzae*: incorporation of disialic acid. J Biol Chem 281:40024-40032
21. Fox KL, Li J, Schweda EK, Vitiazeva V, Makepeace K, Jennings MP, Moxon ER, Hood DW (2008) Duplicate copies of *lic1* direct the addition of multiple phosphocholine residues in the lipopolysaccharide of *Haemophilus influenzae*. Infect Immun 76:588-600
22. Fox KL, Yildirim HH, Deadman ME, Schweda EK, Moxon ER, Hood DW (2005) Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in *Haemophilus influenzae*, identification of a gene for adding O-acetyl groups. Mol Microbiol 58:207-216
23. Garmendia J, Morey P, Bengoechea JA (2012) Impact of cigarette smoke exposure on host-bacterial pathogen interactions. Eur Respir J 39:467-477
24. Geluk F, Eijk PP, van Ham SM, Jansen HM, van Alphen L (1998) The fimbria gene cluster of nonencapsulated *Haemophilus influenzae*. Infect Immun 66:406-417
25. Gilsdorf JR, Marrs CF, Foxman B (2004) *Haemophilus influenzae*: genetic variability and natural selection to identify virulence factors. Infect Immun 72:2457-2461
26. Giufre M, Carattoli A, Cardines R, Mastrantonio P, Cerquetti M (2008) Variation in expression of HMW1 and HMW2 adhesins in invasive nontypeable *Haemophilus influenzae* isolates. BMC Microbiol 8:83
27. Griffin R, Bayliss CD, Herbert MA, Cox AD, Makepeace K, Richards JC, Hood DW, Moxon ER (2005) Digalactoside expression in the lipopolysaccharide of *Haemophilus influenzae* and its role in intravascular survival. Infect Immun 73:7022-7026
28. Griffin R, Cox AD, Makepeace K, Richards JC, Moxon ER, Hood DW (2003) The role of *lex2* in lipopolysaccharide biosynthesis in *Haemophilus influenzae* strains RM7004 and RM153. Microbiology 149:3165-3175
29. Hallstrom T, Resman F, Ristovski M, Riesbeck K (2010) Binding of complement regulators to invasive nontypeable *Haemophilus influenzae* isolates is not increased compared to nasopharyngeal isolates, but serum resistance is linked to disease severity. J Clin Microbiol 48:921-927
30. Harrison A, Dyer DW, Gillaspy A, Ray WC, Mungur R, Carson MB, Zhong H, Gipson J, Gipson M, Johnson LS, Lewis L, Bakaletz LO, Munson RS, Jr. (2005) Genomic sequence of an otitis media isolate of nontypeable *Haemophilus influenzae*: comparative study with *H. influenzae* serotype d, strain KW20. J Bacteriol 187:4627-4636
31. Heath PT, Booy R, Azzopardi HJ, Slack MP, Fogarty J, Moloney AC, Ramsay ME, Moxon ER (2001) Non-type b *Haemophilus influenzae* disease: clinical and epidemiologic characteristics in the *Haemophilus influenzae* type b vaccine era. Pediatr Infect Dis J 20:300-305

32. High NJ, Jennings MP, Moxon ER (1996) Tandem repeats of the tetramer 5′-CAAT-3′ present in *lic2A* are required for phase variation but not lipopolysaccharide biosynthesis in *Haemophilus influenzae*. Mol Microbiol 20:165-174
33. Hill C (2012) Virulence or niche factors: what's in a name? J Bacteriol 194:5725-5727
34. Hill DJ, Toleman MA, Evans DJ, Villullas S, Van Alphen L, Virji M (2001) The variable P5 proteins of typeable and non-typeable *Haemophilus influenzae* target human CEACAM1. Mol Microbiol 39:850-862
35. Hogg JS, Hu FZ, Janto B, Boissy R, Hayes J, Keefe R, Post JC, Ehrlich GD (2007) Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. Genome Biol 8:R103
36. Hood DW, Deadman ME, Cox AD, Makepeace K, Martin A, Richards JC, Moxon ER (2004) Three genes, *lgtF*, *lic2C* and *lpsA*, have a primary role in determining the pattern of oligosaccharide extension from the inner core of *Haemophilus influenzae* LPS. Microbiology 150:2089-2097
37. Hood DW, Deadman ME, Engskog MK, Vitiazeva V, Makepeace K, Schweda EK, Moxon R (2010) Genes required for the synthesis of heptose-containing oligosaccharide outer core extensions in *Haemophilus influenzae* lipopolysaccharide. Microbiology 156:3421-3431
38. Hood DW, Makepeace K, Deadman ME, Rest RF, Thibault P, Martin A, Richards JC, Moxon ER (1999) Sialic acid in the lipopolysaccharide of *Haemophilus influenzae*: strain distribution, influence on serum resistance and structural characterization. Mol Microbiol 33:679-692
39. Krasan GP, Cutter D, Block SL, St Geme JW 3rd (1999) Adhesin expression in matched nasopharyngeal and middle ear isolates of nontypeable *Haemophilus influenzae* from children with acute otitis media. Infect Immun 67:449-454
40. Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR, Jr., Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP, Kishony R (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat Genet 43:1275-1280
41. Lijek RS, Weiser JN (2012) Co-infection subverts mucosal immunity in the upper respiratory tract. Curr Opin Immunol 24:417-423
42. Lomholt H, van Alphen L, Kilian M (1993) Antigenic variation of immunoglobulin A1 proteases among sequential isolates of *Haemophilus influenzae* from healthy children and patients with chronic obstructive pulmonary disease. Infect Immun 61:4575-4581
43. Lysenko E, Richards JC, Cox AD, Stewart A, Martin A, Kapoor M, Weiser JN (2000) The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. Mol Microbiol 35:234-245
44. Martí-Lliteras P, López-Gómez A, Mauro S, Hood DW, Viadas C, Calatayud L, Morey P, Servin A, Linares J, Oliver A, Bengoechea JA, Garmendia J (2011) Nontypable *Haemophilus influenzae* displays a prevalent surface structure molecular pattern in clinical isolates. PLoS One 6:e21133
45. McCrea KW, Xie J, LaCross N, Patel M, Mukundan D, Murphy TF, Marrs CF, Gilsdorf JR (2008) Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains. J Clin Microbiol 46:406-416
46. Mell JC, Shumilina S, Hall IM, Redfield RJ (2011) Transformation of natural genetic variation into *Haemophilus influenzae* genomes. PLoS Pathog 7:e1002151
47. Mira A, Martin-Cuadrado AB, D'Auria G, Rodriguez-Valera F (2010) The bacterial pan-genome: a new paradigm in microbiology. Int Microbiol 13:45-57
48. Moxon R, Bayliss C, Hood D (2006) Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet 40:307-333
49. Murphy TF, Brauer AL, Schiffmacher AT, Sethi S (2004) Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 170:266-272
50. Murphy TF, Kirkham C (2002) Biofilm formation by nontypeable *Haemophilus influenzae*: strain variability, outer membrane antigen expression and role of pili. BMC Microbiol 2:7
51. Murphy TF, Kirkham C, Sethi S, Lesse AJ (2005) Expression of a peroxiredoxin-glutaredoxin by *Haemophilus influenzae* in biofilms and during human respiratory tract infection. FEMS Immunol Med Microbiol 44:81-89
52. Murphy TF, Lesse AJ, Kirkham C, Zhong H, Sethi S, Munson RS Jr. (2011) A clonal group of nontypeable *Haemophilus influenzae* with two IgA proteases is adapted to infection in chronic obstructive pulmonary disease. PLoS One 6:e25923
53. Nakamura S, Shchepetov M, Dalia AB, Clark SE, Murphy TF, Sethi S, Gilsdorf JR, Smith AL, Weiser JN (2011) Molecular basis of increased serum resistance among pulmonary isolates of non-typeable *Haemophilus influenzae*. PLoS Pathog 7:e1001247
54. Oliver A, Mena A (2010) Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance. Clin Microbiol Infect 16:798-808
55. Pettigrew MM, Foxman B, Marrs CF, Gilsdorf JR (2002) Identification of the lipooligosaccharide biosynthesis gene *lic2B* as a putative virulence factor in strains of nontypeable *Haemophilus influenzae* that cause otitis media. Infect Immun 70:3551-3556
56. Poje G, Redfield RJ (2003) Transformation of *Haemophilus influenzae*. Methods Mol Med 71:57-70
57. Power PM, Bentley SD, Parkhill J, Moxon ER, Hood DW (2012) Investigations into genome diversity of *Haemophilus influenzae* using whole genome sequencing of clinical isolates and laboratory transformants. BMC Microbiol 12:273
58. Priest NK, Rudkin JK, Feil EJ, van den Elsen JM, Cheung A, Peacock SJ, Laabei M, Lucks DA, Recker M, Massey RC (2012) From genotype to phenotype: can systems biology be used to predict *Staphylococcus aureus* virulence? Nat Rev Microbiol 10:791-797
59. Rodriguez CA, Avadhanula V, Buscher A, Smith AL, St Geme JW 3rd, Adderson EE (2003) Prevalence and distribution of adhesins in invasive non-type b encapsulated *Haemophilus influenzae*. Infect Immun 71:1635-1642
60. Schweda EK, Richards JC, Hood DW, Moxon ER (2007) Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: implication in virulence. Int J Med Microbiol 297:297-306
61. Sethi S, Wrona C, Grant BJ, Murphy TF (2004) Strain-specific immune response to *Haemophilus influenzae* in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 169:448-453
62. Spahich NA, St Geme JW 3rd (2011) Structure and function of the *Haemophilus influenzae* autotransporters. Front Cell Infect Microbiol 1:5
63. St Geme JW 3rd, Kumar VV, Cutter D, Barenkamp SJ (1998) Prevalence and distribution of the *hmw* and *hia* genes and the HMW and Hia adhesins among genetically diverse strains of nontypeable *Haemophilus influenzae*. Infect Immun 66:364-368
64. Syed SS, Gilsdorf JR (2007) Prevalence of *hicAB*, *lav*, *traA*, and *hifBC* among *Haemophilus influenzae* middle ear and throat strains. FEMS Microbiol Lett 274:180-183
65. Twelkmeyer B, Deadman ME, Haque E, Li J, Hood DW, Schweda EK (2011) The role of *lic2B* in lipopolysaccharide biosynthesis in *Haemophilus influenzae* strain Eagan. Carbohydr Res 346:1262-1266
66. Weimer KE, Armbruster CE, Juneau RA, Hong W, Pang B, Swords WE (2010) Coinfection with *Haemophilus influenzae* promotes pneumococcal biofilm formation during experimental otitis media and impedes the progression of pneumococcal disease. J Infect Dis 202:1068-1075

67 Weiser JN, Pan N, McGowan KL, Musher D, Martin A, Richards J (1998) Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. J Exp Med 187:631-640

68. Weiser JN, Shchepetov M, Chong ST (1997) Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. Infect Immun 65:943-950

69. Whitby PW, Seale TW, VanWagoner TM, Morton DJ, Stull TL (2009) The iron/heme regulated genes of *Haemophilus influenzae*: comparative transcriptional profiling as a tool to define the species core modulon. BMC Genomics 10:6

70. Wild CP (2012) The exposome: from concept to utility. Int J Epidemiol 41:24-32

71. Xie J, Juliao PC, Gilsdorf JR, Ghosh D, Patel M, Marrs CF (2006) Identification of new genetic regions more prevalent in nontypeable *Haemophilus influenzae* otitis media strains than in throat strains. J Clin Microbiol 44:4316-4325

72. Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Hoiby N, Sommer MO, Molin S (2011) Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci USA 108:7481-7486

73. Zhang L, Xie J, Patel M, Bakhtyar A, Ehrlich GD, Ahmed A, Earl J, Marrs CF, Clemans D, Murphy TF, Gilsdorf JR (2012) Nontypeable *Haemophilus influenzae* genetic islands associated with chronic pulmonary infection. PLoS One 7:e44730