Glycoconj J (2013) 30:561–576
DOI 10.1007/s10719-012-9455-5

Haemophilus parainfluenzae has a limited core lipopolysaccharide repertoire with no phase variation

Rosanna E. B. Young · Derek W. Hood

Received: 16 September 2012 / Revised: 14 October 2012 / Accepted: 15 October 2012 / Published online: 24 October 2012

Abstract Cell surface lipopolysaccharide (LPS) is a well characterized virulence determinant for the human pathogen *Haemophilus influenzae*, so an investigation of LPS in the less pathogenic *Haemophilus parainfluenzae* could yield important insights. Using a panel of 18 commensal *H. parainfluenzae* isolates we demonstrate that the set of genes for inner core LPS biosynthesis largely resembles that of *H. influenzae*, with an additional heptosyltransferase I gene similar to *waaC* from *Pasteurella multocida*. Inner core LPS structure is therefore likely to be largely conserved across the two *Haemophilus* species. Outer core LPS biosynthetic genes are much less prevalent in *H. parainfluenzae*, although homologues of the *H. influenzae* LPS genes *lpsB,* non-phase variable *lic2A* and *lgtC,* and *losA1*, *losB1* and *lic2C* are found in certain isolates. Immunoblotting using antibodies directed against selected LPS epitopes was consistent with these data. We found no evidence for tetranucleotide repeat-mediated phase variation in *H. parainfluenzae*. Phosphocholine, a phase variable *H. influenzae* LPS epitope that has been implicated in disease, was absent in *H. parainfluenzae* LPS as were the respective (*lic1*) biosynthetic genes. The introduction of the *lic1* genes into *H. parainfluenzae* led to the phase variable incorporation of phosphocholine into its LPS. Differences in LPS structure between *Haemophilus* species could affect interactions at the bacterial-host interface and therefore the pathogenic potential of these bacteria.

Keywords *Haemophilus parainfluenzae* · *Haemophilus influenzae* · Lipopolysaccharide · Phase variation · Phosphocholine

Electronic supplementary material The online version of this article (doi:10.1007/s10719-012-9455-5) contains supplementary material, which is available to authorized users.

R. E. B. Young (✉)
Department of Paediatrics, University of Oxford,
Oxford, UK
e-mail: r.young@ucl.ac.uk

D. W. Hood
Nuffield Department of Medicine, University of Oxford,
Oxford, UK

Present Address:
R. E. B. Young
Department of Structural and Molecular Biology,
University College London,
Darwin Building, Gower Street,
London WC1E 6BT, UK

Introduction

Haemophilus parainfluenzae and *Haemophilus influenzae* are closely related Gram-negative bacteria residing primarily in the human upper respiratory tract. They were originally distinguished on the basis that *H. influenzae* requires both nicotinamide adenine dinucleotide (NAD) and haemin for growth, whilst *H. parainfluenzae* requires NAD alone [1]. The inclusion of *H. parainfluenzae* as a separate species within the genus *Haemophilus* has since been justified on a phylogenetic basis [2].

Non-typeable *H. influenzae* (NT*Hi*) commonly causes otitis media (OM) and upper and lower respiratory tract infections, whilst encapsulated strains can cause invasive disease including meningitis and septicaemia. Conversely *H. parainfluenzae* disease is considered rare enough to merit the publication of individual case reports of childhood meningitis or septicaemia [3], endocarditis [4], thoracic empyema [5] and biliary tract infections [6]. Many authors include OM in a list of putative *H. parainfluenzae*-related diseases, but there are no reliable primary sources of information to confirm this. The low incidence of disease associated with *H. parainfluenzae*

is not a reflection of low carriage rates, as colonisation is found in up to 100 % of individuals tested [7, 8].

The virulence of a *Haemophilus* species or strain that has colonised the upper respiratory tract depends on its ability to reach a new niche (*e.g*. the middle ear or central nervous system) and upon the subsequent host-microbe interactions. These properties are determined in part by cell surface molecules, including lipopolysaccharide (LPS). The LPS structures of many *H. influenzae* strains have been elucidated and most of the genes required for their assembly have been identified [9–13]. *H. influenzae* LPS comprises four regions: lipid A, 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo), a conserved triheptosyl inner core containing L-*glycero*-D-*manno*-heptose (Hep), and an outer core. It lacks the repetitive polysaccharide chain (O-antigen) that forms the basis of serotyping schemes in many other species. *H. influenzae* LPS structures can differ in the presence and location of outer core substituents such as sialic acid, phosphoethanolamine (PEtn), phosphocholine (PCho), O-acetyl groups (OAc), heptose, D-glucose (Glc) and D-galactose (Gal). These vary between strains due to different complements of sugar transferase genes, but also within strains because many of the genes contain simple sequence repeats (SSR) of tetranucleotides within the reading frame that mediate phase variable expression [14, 15]. SSR have a high mutation rate due to slipped-strand mispairing during DNA replication, resulting in a change in the repeat number and a frameshift, leading to the loss of correct translation. This mechanism allows *H. influenzae* to vary its LPS structure in a stochastic, high frequency and reversible manner, enabling it to adapt quickly through selection for individuals expressing the optimum combination of phase variable substituents for a particular host environment.

Comparing the LPS structure of *H. influenzae* and *H. parainfluenzae* could provide valuable insights into which parts of the molecule are likely to contribute to virulence behaviour. Very few studies have examined *H. parainfluenzae* LPS to date. SDS-PAGE analysis of the LPS profiles from 25 *H. parainfluenzae* isolates [16] indicated that most contained only low molecular mass glycoforms, but a minority had ladder-like LPS profiles indicative of molecules containing O-antigens of varying length. During our research, Pollard *et al.* [17] determined the structure of the carbohydrate portion of LPS from two *H. parainfluenzae* isolates from patients with chronic obstructive pulmonary disease. In both strains a single KdoP was found attached to a heptose trisaccharide, with PEtn extending from HepII and Glc extending from HepI; this correlates with the conserved *H. influenzae* LPS inner core structure. The Glc residue on HepI was augmented with Glc(β1-4)DDHep(α1- in strain 4201 and Gal(β1-6)Glc(β1-4)DDHep(α1- in strain 4282, where DDHep is D-*glycero*-D-*manno*-heptose. No other outer core components or O-antigen-like structures were detected.

Here we report the first genetic analysis of LPS biosynthesis in *H. parainfluenzae*, and use this information together with immunoblotting data to infer likely core LPS structures for a set of eighteen commensal *H. parainfluenzae* isolates from the UK and Gambia. We also include two so-called 'hybrid' isolates, Hy6 and Hy11, for comparison; these strains have characteristics of both *H. influenzae* and *H. parainfluenzae* and cannot be clearly categorised using 16S rRNA and MLST data (Abdel Elamin, personal communication). We have recently confirmed our predictions for one isolate, *H. parainfluenzae* 20, by structural analysis of the LPS [18].

Results

Electrophoretic analysis of *H. parainfluenzae* LPS

The core LPS profiles of our *H. parainfluenzae* study panel, visualised by tricine SDS-PAGE, cover approximately the same size range as those of *H. influenzae* but generally contain fewer bands, indicative of fewer glycoforms (Fig. 1). An *H. influenzae* mutant of known LPS structure was included for comparison (Fig. 1b lane C): the lower glycoform contains lipid A, Kdo and inner core LPS only, whilst the upper band contains an additional four-sugar unit known as high molecular mass glycoform (HMG). Many of the *H. parainfluenzae* glycoforms migrate at approximately the same position as these two bands. The striking simplicity of the *H. parainfluenzae* core LPS profiles indicates that they lack the significant heterogeneity found in *H. influenzae* LPS that results from differences in the repertoire of outer core genes and their phase variable expression. This prompted us to examine the genetics of LPS biosynthesis in our *H. parainfluenzae* strains.

Identification of lipid A, Kdo and triheptose biosynthesis genes in *H. parainfluenzae*, including an additional HepI transferase gene

We combined two approaches to study the complement of putative LPS-related genes in *H. parainfluenzae*. Firstly, the conceptually translated genome sequence of *H. parainfluenzae* strain T3T1 was searched for possible homologues of known LPS biosynthesis enzymes from other bacteria, principally *H. influenzae*, using TBLASTN. This was the first *H. parainfluenzae* genome to be fully sequenced (P. Power and Wellcome Trust Sanger Institute, manuscript in preparation). Secondly, oligonucleotide primers were designed that would amplify 0.5–1 kb fragments of key LPS genes from *H. influenzae* or *H. parainfluenzae* T3T1, then genomic DNA (gDNA) from our 20 study isolates (including T3T1) was subjected to PCR and Southern analyses to shed light on the potential level of LPS variation between strains.

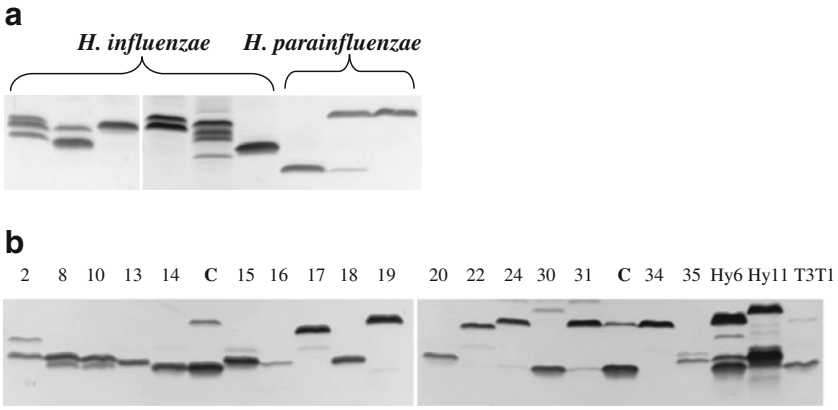

Fig. 1 Core region LPS profiles. Proteinase K treated cell lysates were fractionated by tricine SDS-PAGE and silver-stained. **a** Profiles from representative *Haemophilus* strains. From left to right: *H. influenzae* strains Rd, Eagan, 176, 1231, 1268 and 1292; *H. parainfluenzae* strains 14, 24 and 34. **b** Profiles from all of the *H. parainfluenzae* and 'hybrid' study strains. Strain names are given above each lane. An *H. influenzae* Rd *lpsA lic1* mutant is included for comparison (lane C)

The lipid A moiety of LPS anchors the molecule in the outer membrane of LPS-containing Gram-negative bacteria. In *E. coli*, the transfer of two Kdo residues to lipid A is carried out by KdtA [19], whilst in *H. influenzae* KdtA adds only a single Kdo, which is then phosphorylated by the kinase KdkA [20]. Most species elaborate the lipid A-Kdo moiety with two or three residues of L-*glycero*-D-*manno*-heptose (Hep) to complete the LPS inner core. The entire lipid A, Kdo and triheptose biosynthesis pathways [21] were found to be well conserved between *H. influenzae* and *H. parainfluenzae* T3T1 (Table 1). This is consistent with recent structural evidence that lipid A from *H. parainfluenzae* T3T1 comprises a diglucosamine headgroup and six acyl chains, with an identical configuration to that of *H. influenzae* (Schweda *et al.*, manuscript in preparation). Primers designed to the proposed *H. parainfluenzae* T3T1 Kdo transferase and Kdo kinase genes amplified PCR products of the expected size from the gDNA of 18/18 true *H. parainfluenzae* strains, whilst Southern analysis of restriction digested gDNA confirmed that a potential *orfH* homologue is present in all 20 test strains (Table 2 and Online Resource Fig. S1). *OrfH* encodes the last enzyme required in LPS triheptosyl core synthesis: its presence suggests that this structure is conserved across *H. parainfluenzae* strains as it is in *H. influenzae*. Indeed, a triheptosyl structure has been identified in the LPS of the four *H. parainfluenzae* strains analysed so far, including our panel strains T3T1 and 20 (Schweda *et al.*, manuscript in preparation) [17, 18].

H. influenzae LPS contains lipid A-KdoP instead of the more common lipid A-Kdo_2, so *H. influenzae* OpsX is unusual in acceptor specificity and sequence when compared to Hep I transferases in other bacteria [22]. A TBLASTN search with the *E. coli* Hep I transferase WaaC (GenBank: ACT27183.1) revealed an *H. parainfluenzae* T3T1 gene, *PARA_03820*, encoding a protein with 58 % aa identity. Proteins with high sequence similarity to PARA_03820 are present in several other Pasteurellaceae species including *Histophilus somni* (72 % aa identity), *Pasteurella multocida* (69 %) and *Actinobacillus pleuropneumoniae* (61 %), but not *H. influenzae*. This raises the possibility that *H. parainfluenzae* is capable of synthesising both lipid A-KdoP and lipid A-Kdo_2, and can extend these with HepI using different substrate-specific transferases encoded by *opsX* and *waaC*. Primers designed to *H. parainfluenzae* T3T1 *waaC* amplified a PCR product of the expected size from all 18 true *H. parainfluenzae* strains in the study panel but not from the two hybrid strains (Table 2), providing strong evidence that this gene is conserved across *H. parainfluenzae*.

The LPS inner core of all *H. influenzae* strains contains PEtn linked *via* the 6 position on HepII [23]. The *H. parainfluenzae* T3T1 homologue of the respective transferase (Lpt6) is likely to be PARA_02360, which shares 71 % aa identity with *H. influenzae* Rd Lpt6 (HI0275). Southern analysis using a fragment of the *H. parainfluenzae* T3T1 *lpt6* gene as a probe indicated that the gene was present in all 18 true *H. parainfluenzae* strains in our collection (Table 2 and Online Resource Fig. S1).

H. parainfluenzae contains putative homologues of *H. influenzae* LPS genes that initiate oligosaccharide extensions

The translated *H. parainfluenzae* T3T1 genome was searched using TBLASTN for possible homologues of 32 additional *H. influenzae* proteins with a known or suspected role in core LPS assembly (Online Resource Table S1). Key proteins that initiate oligosaccharide extensions from the inner core of *H. influenzae* LPS are LgtF (adds Glc to HepI), Lic2C (adds Glc to HepII) and LpsA (adds Glc or Gal to HepIII); the presence and sequence of the genes encoding these proteins are major determinants of LPS heterogeneity

Table 1 Identification in the *H. parainfluenzae* T3T1 genome of genes required for the synthesis of lipid A, Kdo and triheptosyl inner core LPS. The genes required for each process in *H. influenzae* are listed along with their protein functions. Results of TBLASTN searches using *H. influenzae* Rd amino acid sequences against the *H. parainfluenzae* T3T1 genome are shown in the last three columns: the closest match (*i.e.* the translated ORF representing the highest scoring HSP) is specified alongside the E-value and % amino acid identity (%ID) for the match

Process	Gene name	Protein function	*Hi* Rd	Predicted *Hp* T3T1 homologue		
			Protein	Protein	E-value	% ID
Lipid A synthesis	*lpxA*	UDP-*N*-acetylglucosamine acyltransferase	HI_1061	PARA_10200	1.2E-127	92
	lpxC	UDP-3-*O*-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	HI_1144	PARA_12580	4.1E-147	92
	lpxD	UDP-3-*O*-[3-hydroxymyristoyl] glucosamine N-acyltransferase	HI_0915	PARA_10180	1.7E-154	87
	lpxH	UDP-2,3-diacylglucosamine hydrolase	HI_0735	PARA_08690	8.9E-087	69
	lpxB	lipid-A-disaccharide synthase	HI_1060	PARA_10210	2.8E-162	81
	lpxK	tetraacyldisaccharide 4′-kinase	HI_0059	PARA_13450	1.1E-140	80
	lpxL (*htrB*)	laurate transferase	HI_1527	PARA_18870	9.0E-147	84
	lpxM (*msbB*)	myristate transferase	HI_0199	PARA_19060	2.3E-143	81
Kdo synthesis and transfer	*kdsD*	arabinose-5-phosphate isomerase	HI_1678	PARA_14340	7.9E-148	91
	kdsA	Kdo 8-phosphate synthase	HI_1557	PARA_14490	2.9E-142	94
	kdsC (*yrbI*)	Kdo 8-phosphate phosphatase	HI_1679	PARA_14330	1.7E-074	82
	kdsB	Kdo cytidylyltransferase	HI_0058	PARA_13470	1.0E-115	89
	kdtA (*waaA*)	Kdo transferase	HI_0652	PARA_04700	8.2E-161	71
	kdkA (*orfZ*)	Kdo kinase	HI_0260.1	PARA_02570	1.7E-092	70
Regulation	*ftsH* (*hflB*)	ATP-dependent metallopeptidase/cell division protease	HI_1335	PARA_17160	5.1E-311	89
Triheptosyl core synthesis and transfer	*gmhA*	phosphoheptose isomerase	HI_1181	PARA_12250	1.7E-092	90
	hldE (*rfaE*)	bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase	HI_1526	PARA_18860	1.0E-230	92
	gmhB	D,D-heptose 1,7-bisphosphate phosphatase	HI_0621.1	PARA_08320	5.8E-083	83
	hldD (*rfaD*)	ADP-L-*glycero*-D-*manno*heptose 6-epimerase	HI_1114	PARA_03860	3.6E-154	91
	opsX	ADP-heptose:LPS heptosyltransferase I	HI_0261	PARA_02550	6.2E-148	79
	rfaF	ADP-heptose:LPS heptosyltransferase II	HI_1105	PARA_03850	7.5E-154	83
	orfH	ADP-heptose:LPS heptosyltransferase III	HI_0523	PARA_08430	9.4E-156	81

between strains [24]. The *H. parainfluenzae* T3T1 gene *PARA_04710* was found to encode a protein with 86 % aa identity to LgtF from *H. influenzae* Rd (HI0653); in both species the gene is adjacent to that which encodes KdtA. Southern analysis confirmed that *lgtF* was present in all 18 *H. parainfluenzae* test strains (Table 2 and Online Resource Fig. S1). It is therefore likely that the addition of Glc to HepI and the addition of PEtn to HepII (above) are conserved features of the *H. parainfluenzae* LPS inner core. Current structural data are consistent with this hypothesis [17, 18].

In *H. influenzae* the enzyme responsible for hexose addition to HepIII is encoded by one of four allelic variants of the *lpsA* gene; LpsA (directs the addition of Glc *via* a β1-2 linkage), LpsB (Gal β1-2), LpsC (Glc β1-3) and LpsD (Gal β1-3) [10]. Although the *H. parainfluenzae* T3T1 genome does not encode homologues of these enzymes, when some contigs of the *H. parainfluenzae* strain 20 genome sequence (made available through an ongoing collaboration with the Wellcome Trust Sanger Institute in Cambridge) were searched for homologues, one contig was found to contain a fragment of a likely *lpsB* homologue at one end. This was followed by an ORF that encoded 95 % aa identity to PARA_08860 in strain T3T1. To check the gene arrangement in this region, primers 5′-GATATGGGCTTAGTGAATACTG-3′ and 5′-ATTCAATGGCTTGCTCTTG-3′ were designed to bridge the genes *PARA_08850* (*menB*) and *PARA_08860* (*menC*) and used for PCR analysis of the 20 study strains. For thirteen strains a 0.5 kb product was amplified, indicating that the genes were adjacent as in *H. parainfluenzae* T3T1. This is likely to be the ancestral gene arrangement, as MenB and MenC are involved in the same biosynthetic pathway and the genes are also contiguous in *E. coli* and *H. influenzae*. For the remaining five strains (*H. parainfluenzae* 8, 10, 15, 20 and 35) the amplified products were 1.4 kb in

Table 2 Experimental detection of selected inner and outer core LPS biosynthesis genes in the *H. parainfluenzae* study strains. One of three methods was used per gene, as specified by the 'Expt' column. The primers used are listed in Online Resource Table S3. *H. influenzae* positive controls were included where necessary but are not shown. 'S' indicates Southern analysis of gDNA digests using a 0.5–1 kb probe which was amplified from the template gDNA listed; the number of hybridising bands is stated. 'P' indicates the attempted amplification of an internal gene fragment using primers designed to the strain listed under 'Template'; '+' means that a product of the expected size was amplified, whilst '–' means that no product was amplified. Note that *lpsA* and *losA1* were later sequenced also from *H. parainfluenzae* strain 35. (+) indicates that no product was detected in PCR but that an ORF encoding >90 % amino acid identity was identified from DNA sequence fragments that were available for these four strains only. 'L' (locus PCR) indicates that the *H. influenzae* location of this gene was amplified using primers designed to the two flanking genes; *H. parainfluenzae* strains with short PCR products (*i.e.* no intervening gene) were scored as negative, whilst strains with longer products were scored as positive and confirmed by sequencing. GT stands for glycosyltransferase

Function	Gene	Template	Expt	*H. parainfluenzae* strain number																		
				2	8	10	13	14	15	16	17	18	19	20	22	24	30	31	34	35	Hy6	Hy11	T3T1
Kdo transferase	*kdtA*	*Hp* T3T1	P	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	+
Kdo kinase	*kdkA*	*Hp* T3T1	P	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	+
Hep transferases	*opsX*	*Hp* T3T1	P	(+)	(+)	–	+	–	(+)	+	+	+	+	(+)	+	+	–	+	+	+	–	–	+
	rfaF	*Hp* T3T1	P	(+)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	–	+
	orfH	*Hp* T3T1	S	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1
	waaC	*Hp* T3T1	P	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–	–	+
Initial GTs	*lgtF*	*Hp* T3T1	S	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	0	0	1
	lic2C	*Hi* Rd	L	–	–	–	+	–	–	+	–	+	–	–	–	+	–	+	–	–	–	–	–
	lpsA (*Hi*)	*Hi* Rd	P	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	lpsA (*Hp*)	*Hp* 20	P	–	+	+	–	–	+	–	–	–	–	+	–	–	–	–	–	–	–	–	–
Other GTs	*lex2A/B*	*Hi* Eagan	S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	lic2A	*Hp* T3T1	L	–	–	–	–	–	–	–	+	–	+	–	–	–	–	–	+	–	–	–	–
	losA1	*Hp* 17	P	–	–	–	–	–	–	–	+	–	+	–	–	–	–	–	+	–	–	–	–
Sialyltransferases	*lic3A or B*	*Hi* Rd	P	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	lic3B	*Hp* T3T1	L	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
PEtn transferases	*lpt6*	*Hp* T3T1	S	1	1	1	1	1	1	1	1	1	3	1	1	1	1	3	1	1	0	0	2
	lpt3	*Hi* 86-028NP	P	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Choline transport	*lic1B*	*Hi* Rd	S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O-acetylase	*oafA*	*Hi* Rd	S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HepIV transferase	*losB1*	*Hi* R2846	P	–	–	–	–	–	–	–	+	–	+	–	–	–	–	–	+	+	–	–	–

length, and DNA sequencing revealed that intact *lpsB* homologues were present in this genomic location in all five strains. The five *H. parainfluenzae lpsB* sequences have been submitted to GenBank; upon translation they share 73–75 % aa identity with *H. influenzae* Rd LpsA (HI0765).

In *H. influenzae* the amino acid at position 151 of the different LpsA alleles determines whether the enzyme adds Gal (threonine) or Glc (cysteine, alanine or methionine) to the distal heptose of the LPS core, whilst the C-terminal region determines the linkage [10]. All five of the *H. parainfluenzae* homologues encode threonine at position 151 and have a generally Rd-like (β1-2) rather than strain 486-like (β1-3) C-terminal region (Online Resource Fig. S2), suggesting that they add Gal *via* a β1-2 linkage to the third heptose. Structural analysis has recently confirmed that this is indeed the case for strain 20 [18].

Monoclonal antibodies of known binding specificity can be useful tools to detect certain LPS epitopes (Fig. 2a–c). MAb L6A9 detects the Gal(β1-2)HepIII epitope on *Haemophilus* LPS [24]. Dot immunoblots using lysates showed that only *H. parainfluenzae* strains 8, 10, 15, 20 and 35 reacted with this antibody (Fig. 2a), corresponding to the same five strains identified above as containing an *lpsB* homologue. The intense signal obtained with the mAb suggests that the epitope is unlikely to be further substituted.

Many *H. influenzae* strains contain non-phase variable outer core LPS genes, including *lic2C*, located between *ksgA* and *infA* (Fig. 3a). No similar genes were found in the T3T1 genome sequence but we investigated the organisation of this locus in other *H. parainfluenzae* strains by PCR and DNA sequencing (Fig. 3b, PCR 1). The PCR1 primer pair (Online Resource Table S3) amplified one of three different sized *ksgA-infA* PCR products from each study strain. For eleven strains including T3T1 and the two hybrids, a 700 bp product indicated that no extra genes were present. For five strains (*H. parainfluenzae* 13, 16, 18, 24 and 31) a 2 kb product was amplified: sequencing of the *ksgA-infA* region from strain 13 revealed a single ORF, which encoded a protein with 94 % aa identity to Lic2C from *H. influenzae* strain RM7004. The strain 13 *lic2C* sequence is available in GenBank. Further PCR analysis confirmed that this gene arrangement was conserved in the other four *H. parainfluenzae* strains listed above, *i.e.* they contain only *lic2C* at this locus (data not shown). These strains thus have a potential to add Glc to HepII of their LPS core.

For the remaining four *H. parainfluenzae* isolates (strains 17, 19, 34 and 35) the PCR1 primer pair amplified a 2.6 kb product. Upon DNA sequencing, the *ksgA-infA* intergenic region from strain 17 was found to encode two proteins with similarity to LosB1 and LosA1 (92 % and 81 % identity respectively) from *H. influenzae* strain R2846. In *H. influenzae* these enzymes add a Glc-DDHep- epitope to the Glc on the proximal heptose of the inner core [25]. This exact extension was observed in the two *H. parainfluenzae* isolates studied by Pollard *et al.* [17]. Further analysis confirmed that our strains 19, 34 and 35 also contained *losB1* and *losA1* at this locus. The proportion of strains with these genes is therefore similar between *H. parainfluenzae* (4/18; 22 %) and NT*Hi* (21 %).

H. parainfluenzae strain 17 contains repeatless *lic2A* and *lgtC* homologues for digalactoside addition

A digalactoside epitope, formed by the sequential action of two phase variable galactosyltransferases Lic2A then LgtC [26], can be part of oligosaccharides extending from any Hep of the *H. influenzae* LPS inner core. *Lic2A* is present in all NT*Hi* strains between the *apaH* and *ksgA* genes, *i.e.* adjacent to the *ksgA-infA* locus discussed above (Fig. 3a). To examine whether *H. parainfluenzae* can also encode Lic2A in this location, primer pair PCR2 was designed to

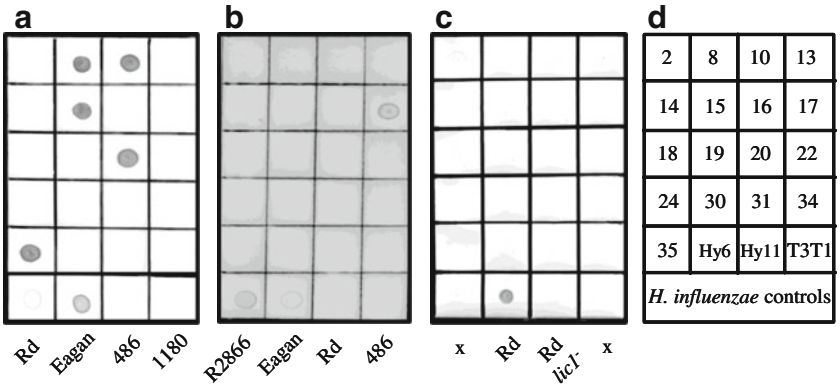

Fig. 2 Dot immunoblots to detect Gal, digalactoside and PCho epitopes on *H. parainfluenzae* LPS. Each *square* contains lysed cells from one study strain, with *H. influenzae* control strains in the *bottom row* as indicated. A key showing the layout of *H. parainfluenzae* strains in all panels is shown on the right, in panel **d**. **a** Blot using mAb L6A9 to detect Gal(β1-2)HepIII. The four *H. influenzae* strains contain, from left to right, *lpsA*, *lpsB*, *lpsC* and *lpsD*. **b** Blot using mAb 4C4 to detect digalactoside. *H. influenzae* R2866, Eagan and Rd express this epitope in a phase variable manner. *H. influenzae* 486 LPS does not contain digalactoside. **c** Blot using mAb TEPC15 to detect PCho

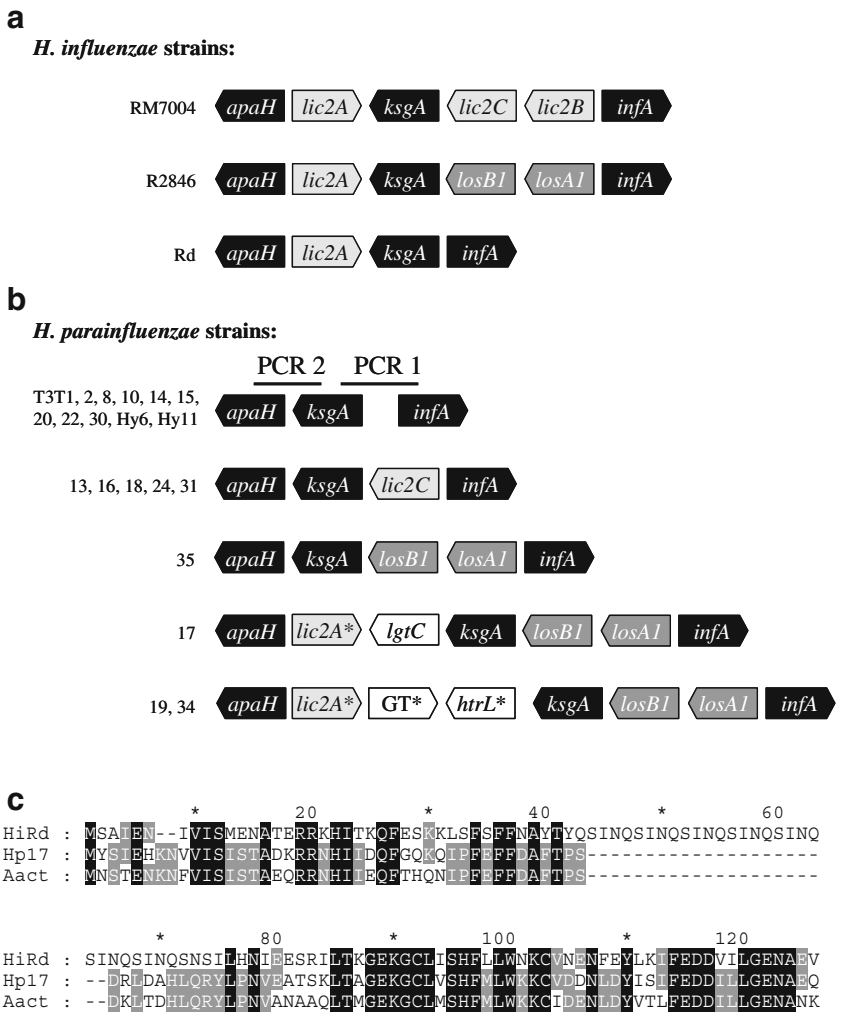

Fig. 3 The *lic2* locus of *H. influenzae* and *H. parainfluenzae*. Panels **a** and **b** show the invariable flanking genes in black and genes related to LPS synthesis in grey (typical *H. influenzae lic2* locus genes) or white (genes not previously found at this locus). Not drawn to scale. **a** The three known layouts of the locus in *H. influenzae*, represented here by strains RM7004, R2846 and Rd. **b** The organisation of the *lic2* locus in all 20 *H. parainfluenzae* or hybrid study strains as deduced by PCR amplification and sequencing. Genes with closer homologues in *Aggregatibacter actinomycetemcomitans* than in *H. influenzae* are indicated by asterisks. **c** Alignment of N-terminal Lic2A sequences. HiRd = *H. influenzae* strain Rd (protein HI0550); Hp17 = *H. parainfluenzae* strain 17; Aact = *A. actinomycetemcomitans* strain D11S-1 (protein D11S_0429). The *H. influenzae lic2A* gene contains CAAT repeats, which translate as $(SINQ)_n$ and mediate phase variation of the enzyme

the strain T3T1 homologues of *apaH* and *ksgA* and used to amplify the locus from the study strains (Fig. 3b, PCR 2; Online Resource Table S3). As with *H. parainfluenzae* T3T1, there was no ORF between these two flanking genes in most of the study strains. Only *H. parainfluenzae* strains 17, 19 and 34 yielded larger PCR products.

The *apaH-ksgA* region from strain 17 was fully sequenced and those from strains 19 and 34 were partially sequenced; the deduced loci are shown in Fig. 3b. The loci from strains 19 and 34 each contained three genes with similarity to D11S_0429, D11S_0428 and D11S_0427 from *Aggregatibacter actinomycetemcomitans*; these three contiguous genes have likely transferred between species *en bloc*. They are part of the 6 kb 'Genomic island G–LPS biosynthesis enzymes' region of the *A. actinomycetemcomitans* D11S-1 genome [27]. Precise functions have not been ascribed for these ORFs, but D11S_0427 belongs to the HtrL superfamily of unknown function (often found in LPS loci) and D11S_0428 contains a glycosyltransferase DXD motif and is related to capsule synthesis proteins.

The *H. parainfluenzae* strain 17 *apaH-ksgA* locus contains two ORFs. The first encodes a protein with 75 % aa identity to D11S_0429 from *A. actinomycetemcomitans* (as above) and 56 % aa identity to HI0550 (Lic2A) from *H. influenzae*. The second ORF encodes a protein with 50 % aa identity to HI0258 (LgtC) from *H. influenzae*. The two enzymes are likely to be functional and to add sequential Gal residues of the digalactoside epitope as their *H. influenzae* counterparts do, because a mAb specific for this epitope bound only to an *H. parainfluenzae* strain 17 lysate (Fig. 2b). The discovery of convergent *lic2A* and *lgtC* homologues in *H. parainfluenzae* strain 17 was striking, as they are always found at separate genomic locations in *H. influenzae*. Given that the three *H. parainfluenzae* strains with a *lic2A* homologue correspond to a subset of the four strains that contain *losA1* and *losB1*, it is possible that Lic2A adds a Gal residue to the Glc-DDHep extension in these isolates.

Remarkably, the *lic2A* and *lgtC* genes in *H. parainfluenzae* strain 17 both lack the tetranucleotide repeat tracts found in their counterparts in *H. influenzae* that normally enable phase variable expression of the galactosyltransferases. In the *H. influenzae* Rd *lic2A* gene, the $(CAAT)_n$ tract begins 120 bp downstream of the second potential start codon. The

H. parainfluenzae lic2A homologue contains a shorter, non-repetitive sequence in place of this repeat tract (Fig. 3c), suggesting that the expression of the enzyme is not phase variable. Similarly, the *H. parainfluenzae* strain 17 homologue of *lgtC* contains neither the $(GACA)_n$ tract of the *H. influenzae* version nor the homopolymeric repeat tract of the *Neisseria meningitidis* version (data not shown).

Identification of LPS core biosynthesis genes aids interpretation of the observed *H. parainfluenzae* core LPS profiles

The core LPS structures that we would predict for the 18 true *H. parainfluenzae* strains investigated, based on our genetic, immunological and initial structural analyses, are summarised in Fig. 4. The presence of outer core glycosyltransferases in *H. parainfluenzae* shows some correlation with the observed core LPS profiles (Fig. 1b). For example, no outer core genes were found in strain 14, whose LPS consists of a single glycoform of a similar molecular mass as the lipid A plus inner core LPS control (lower band in lane C); strains 13 and 15 contain putative *lic2C* and *lpsB* homologues respectively and their core LPS glycoforms appear to be one sugar larger than that of strain 14, as would be expected. Four putative outer core glycosyltransferases were identified in strains 19 and 34, which each display a single core glycoform migrating at the same level as the upper band in lane C (corresponding to inner core plus four outer core sugars).

The highest molecular mass band in the core LPS profiles of *H. parainfluenzae* strains T3T1, 22, 24 and 31 also migrates to approximately the same position as the higher molecular mass glycoform in lane C, but cannot be accounted for by their identified outer core glycosyltransferases. However, preliminary analysis of the T3T1 genome reveals a potential O-antigen biosynthesis locus that we would predict encodes the addition of a polymerised four-sugar O-unit to the LPS; this offers a potential explanation for the observed glycoforms. Like Roberts *et al.* [16], we observed that some of our *H.*

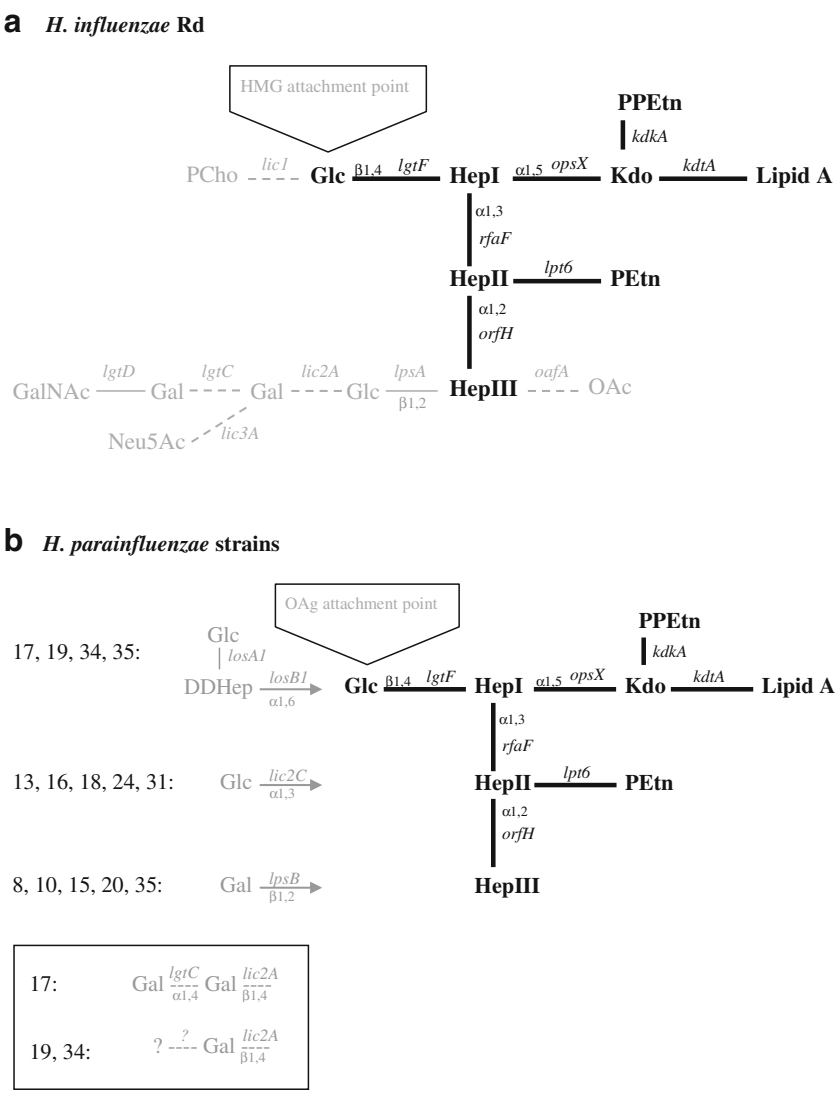

Fig. 4 LPS structure of *H. influenzae* Rd and putative core LPS structures of the *H. parainfluenzae* study strains. *Bold* type represents the inner core region, which is conserved across all *H. influenzae* strains and which genetic analysis suggests is also conserved across *H. parainfluenzae*. **a** LPS structure of *H. influenzae* strain Rd for comparison [59, 60]. Outer core LPS is shown in *grey*, with phase variable additions indicated by *dotted lines*. The gene encoding each transferase is given in italics. **b** Putative core LPS structures of the 18 true *H. parainfluenzae* study strains based on genetic and immunoblot evidence. Outer core sugars and the genes encoding their transferases are shown in grey alongside their predicted points of addition; strains with homologues to each gene are listed to the left. The position of the boxed outer core additions cannot be predicted. No outer core genes were detected for *H. parainfluenzae* strains T3T1, 2, 14, 22 or 30. All 18 true *H. parainfluenzae* study strains also contain a putative O-antigen biosynthesis locus, which will be described in a separate report

parainfluenzae LPS profiles contained evenly-spaced ladders of bands, typical of O-antigens, in the upper part of the gel (not shown). We are currently examining the genetic potential of these strains to synthesise O-antigens.

H. parainfluenzae T3T1 does not use tetranucleotide repeat-mediated phase variation to alter its LPS structure

A significant part of LPS structural heterogeneity within and between *H. influenzae* strains is realized through multiple phase variable biosynthesis genes. Our identification of repeatless *lic2A* and *lgtC* homologues in an *H. parainfluenzae* strain led us to consider whether this species ever uses tetranucleotide repeats to mediate the phase variation of cell surface components. In a study of 16 *H. influenzae* genomes, our laboratory previously used an algorithm called *Hi*SSR-finder to demonstrate that each strain contained 8–20 tetranucleotide repeat tracts of at least 20 bp in length [14]. In contrast, the algorithm shows that the *H. parainfluenzae* T3T1 genome sequence does not contain any tetranucleotide repeat tracts longer than 12 bp (Power *et al.*, manuscript in preparation). This finding suggests a major difference in the mechanisms of cell surface variation between the two species.

When the twelve phase variable enzymes that have proven or putative roles in *H. influenzae* outer core LPS biosynthesis were searched against the translated *H. parainfluenzae* T3T1 genome sequence, no close matches were found (Online Resource Table S1). The epitopes added by these transferases in *H. influenzae* include hexose (*lex2A*, *lgtC*, *lic2A* and other genes), sialic acid (*lic3A* and *lic3B*), phosphoethanolamine (*lpt3*), phosphocholine (*lic1A* locus) and O-acetyl groups (*oafA*). However, our study has already shown that the repertoire of LPS biosynthesis genes varies between *H. parainfluenzae* strains. To confirm that there are no potential LPS genes in this organism that contain particular tetranucleotide repeats, Southern analysis was carried out on restriction digested gDNA from the 20 study strains using four different $(NNNN)_5$ oligonucleotide probes which between them represent 91 % of the 199 tetranucleotide repeat tracts identified across 16 *H. influenzae* genomes by Power *et al.* [14]. No hybridising bands were found for any *H. parainfluenzae* strain using the probes $(CAAT)_5$, $(GACA)_5$ or $(CAAC)_5$, contrasting with the strong signals obtained for all three *H. influenzae* control strains. Figure 5 shows representative results. For the fourth probe, $(GCAA)_5$, very weak bands were found for some *H. parainfluenzae* strains (data not shown). In *H. influenzae*, GCAA repeat tracts are found in two phase variable outer core LPS-related genes, *lex2A* and *oafA*. Repeatless fragments of these genes were used to reprobe the same Southern blot but did not hybridise to the gDNA of any of the 18 true *H. parainfluenzae* strains, suggesting that the weak signals observed with the $(GCAA)_5$ probe represent short and/or imperfect GCAA repeat tracts that are not within *lex2A* or *oafA* homologues. Whilst we cannot rule out the possibility that other tetranucleotide permutations may be present as SSR in some *H. parainfluenzae* study strains, the Southern analysis provides evidence that the four tetranucleotide repeat sequences that are most commonly used by *H. influenzae* to mediate phase variation are not prevalent in *H. parainfluenzae*.

H. parainfluenzae does not naturally express phosphocholine on its LPS but can do so when a transgenic *lic1* locus is introduced

Most of the epitopes added to *H. influenzae* LPS in a phase variable manner have proven or inferred roles in pathogenesis. It is not clear if the absence of these epitopes from *H. parainfluenzae* LPS is due to some instability of the respective SSR-containing genes. In *H. influenzae* the uptake of choline and its incorporation into the LPS as PCho requires the four genes (*lic1A-lic1D*) of the *lic1* locus, which is found in most NT*Hi* and capsular strains [13, 28]. *Lic1A* contains a $(CAAT)_n$ SSR that underlies phase variation of Lic1A and PCho expression [15].

Several lines of evidence suggest that *H. parainfluenzae* does not contain the *lic1A-lic1D* genes. Firstly, no homologues of the four Lic1 proteins were found in TBLASTN searches of the translated *H. parainfluenzae* T3T1 genome sequence. Secondly, when an 800 bp fragment of the *H. influenzae* Rd *lic1B* gene was used as a probe in Southern analysis against gDNA digests from the *H. parainfluenzae* study strains, no hybridising bands were detected (Fig. 6a). Thirdly, previous Southern analysis using a probe comprising the *lic1A* SSR sequence $(CAAT)_5$ was also negative in all 20 study strains (Fig. 5a). Fourthly, none of the *H. parainfluenzae* strains reacted with the PCho-specific mAb TEPC15 on a dot immunoblot (Fig. 2c), consistent with their LPS lacking PCho.

Given the supposed high level of natural transformation amongst *Haemophilus* species, we sought to explore why *H. parainfluenzae* does not naturally support PCho incorporation into its LPS. A plasmid construct (pRY29) was used to introduce the *lic1* locus from *H. influenzae* strain Rd into the *H. parainfluenzae* genome (Fig. 6b). A tetracycline resistance gene cluster (*tet*) found in a number of the study strains was chosen as the target site for recombination as it is non-essential and its disruption would result in a defined phenotypic change. The insert of pRY29 was sequenced to ensure that its *lic1A* gene contained a number of CAAT repeats that was permissive for expression. *H. parainfluenzae* strain 19 was successfully transformed with the *lic1* locus; lysates from two $lic1^+$ transformants, 19.4 and 19.5, were immunoblotted and found to bind mAb TEPC15 (Fig. 6c). The *lic1* locus was also introduced into, and

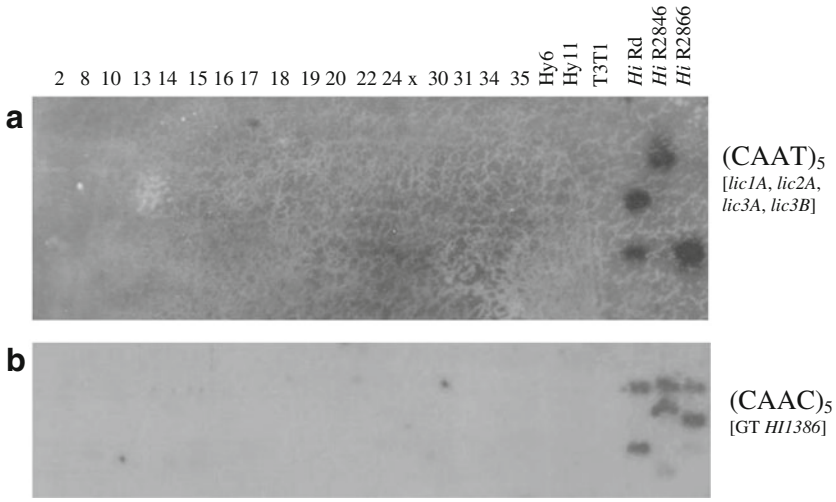

Fig. 5 Southern analysis of MfeI-restricted gDNA hybridised with ^{32}P-labelled tetranucleotide repeat probes. *H. parainfluenzae* strain numbers are listed above each lane; three *H. influenzae* control strains are shown on the right. Probe sequences used for each panel are given to the right of the figure, with *H. influenzae* LPS genes that contain this repeat sequence listed in square brackets. In *H. influenzae* strain Rd, $(CAAC)_n$ is also present in four *hgp* genes, which encode haemoglobin and haemoglobin-haptoglobin binding proteins

successfully expressed in, a strain derived from *H. parainfluenzae* 20 (Fig. 6c). This demonstrates that *H. parainfluenzae* can express functional proteins encoded by the *H. influenzae lic1* locus and that there is no fundamental reason why the species could not include PCho as a constituent of its LPS if it were advantageous to do so. Our *lic1*$^-$ and *lic1*$^+$ strain pairs will be valuable for future investigations into the importance of PCho in epithelial attachment, complement-mediated killing and the ability of *H. parainfluenzae* to cause OM in a murine model.

To investigate whether transgenic *lic1A* SSR in *H. parainfluenzae* can promote the phase variable expression of PCho, we carried out colony immunoblotting. A preferred 'non-clumping' *lic1*$^+$ strain (*H. parainfluenzae* 20.21; see Experimental procedures) was developed for this experiment. When approximately 3000 *H. parainfluenzae* 20.21 colonies were immunoblotted with mAb TEPC15, the majority of colonies were highly reactive; however, six non-reactive colonies were observed (Fig. 6d). The repeat regions of the *lic1* locus from these putative phase variants were sequenced: the $(CAAT)_n$ SSR of all six had increased in length from $n=16$ to $n=17$, making the *lic1A* gene out of frame with either of the potential initiation codons. One of these ON to OFF phase variants, 20.21A, was further immunoblotted; 29/4200 colonies were TEPC15-reactive (Fig. 6e). When the $(CAAT)_n$ SSR was sequenced for six of these reactive colonies, three were found to have gained either one or two repeats ($n=18$ or $n=19$) and three had lost a repeat ($n=16$), resulting in the restoration of Lic1A function.

These experiments provide evidence that *H. parainfluenzae* is capable of repeat-mediated slippage (both loss and gain of repeats) and that PCho can be added to its LPS in a phase variable manner when provided with a transgenic *lic1* locus. Although accurate slippage rates were not calculated, phase variants were detected at a similar frequency in *H. parainfluenzae* 20.21 as in the *H. influenzae* Rd control experiment.

Discussion

Our investigation of the genetic basis of LPS core biosynthesis indicates that *H. parainfluenzae* has a lipid A and inner core that is highly conserved and effectively identical to that found in *H. influenzae*. The striking difference between the two species is the significantly reduced capacity for synthesis of outer core structures in *H. parainfluenzae* and a concomitant lack of phase variation of the LPS molecule. Outer core epitopes expressed in a phase variable manner, including PCho, sialic acid, digalactoside and O-acetyl groups, are all involved with *H. influenzae* pathogenesis, strongly suggesting that the lack of these epitopes in the core LPS of *H. parainfluenzae* may be an important factor in determining its significantly reduced association with disease. Other potential virulence factors expressed by *H. influenzae* include IgA protease [29], for which there appears to be no homologue encoded in the *H. parainfluenzae* T3T1 genome sequence, and adhesins, which are present in *H. parainfluenzae* but differ in size and sequence between the two species [30] (RY, unpublished data).

The release of LPS during bacterial colonisation and lysis induces the production of host proinflammatory cytokines through the recognition of lipid A by a complex of Toll-like receptor 4 (TLR4) and its co-receptor MD2 [31]. Whilst this is an important mechanism for local bacterial clearance, systemic overstimulation of this pathway can result in fatal septic shock. The degree of lipid A phosphorylation and the number, position and length of its acyl chains can affect the binding affinity between lipid A and TLR4-MD2, altering activation of the downstream signalling cascade [32]. Tuyau and Sims [33] found that *H. parainfluenzae* LPS was eight times less potent per microgram than *H. influenzae* LPS in terms of inducing lethal endotoxic shock in mice. We might therefore have expected the lipid A structures of *H. influenzae* and *H. parainfluenzae* to differ, but our genetic and structural analyses indicate that in fact these species

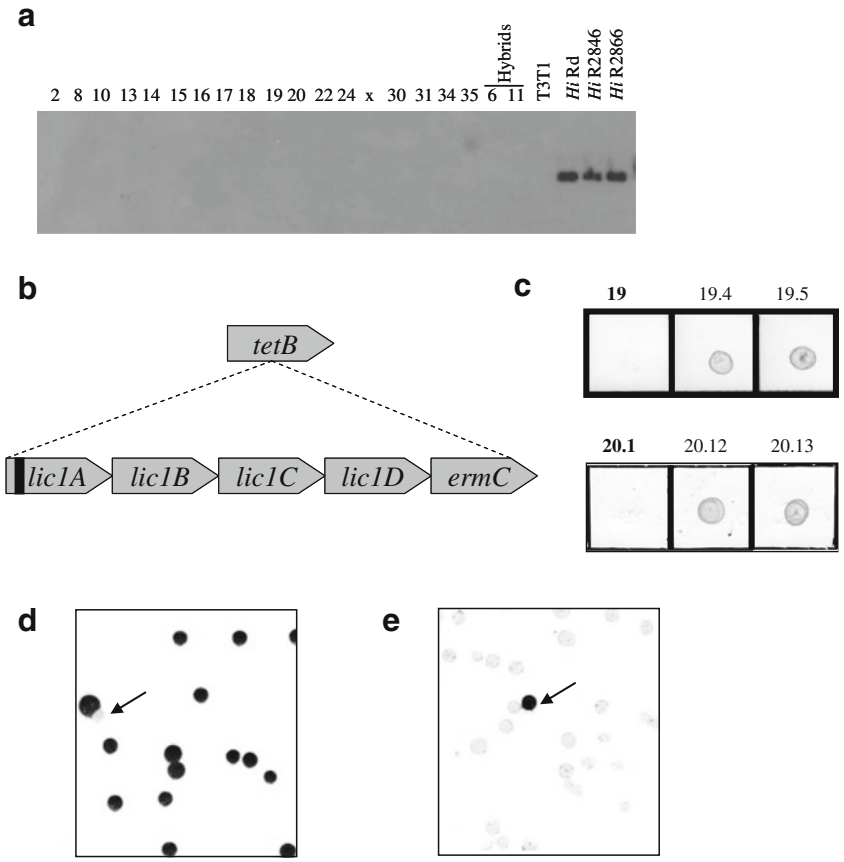

Fig. 6 *H. parainfluenzae* can phase variably express PCho using a transgenic *lic1* locus. **a** Southern analysis of MfeI-digested gDNA hybridised with an *H. influenzae* Rd *lic1B* PCR fragment. *Lic1B* encodes a choline transporter required for PCho synthesis. *H. parainfluenzae* strain numbers are listed above each lane; three *H. influenzae* controls are shown on the right. **b** The *lic1* locus from *H. influenzae* Rd, along with an erythromycin resistance gene (*ermC*), was introduced into *H. parainfluenzae* strains 19 and 20.1 by the homologous recombination of flanking *tetB* fragments. The $(CAAT)_n$ repeat tract in *lic1A* is indicated by a black rectangle. **c** TEPC15 dot immunoblots using lysates from parental *H. parainfluenzae* strains (in *bold*) and their derivative clones containing a transgenic *lic1* locus, demonstrating PCho expression in the mutants. **d**–**e** Detection of repeat-mediated phase variation in a transgenic *H. parainfluenzae* strain. Colonies were grown on BHI agar and immunoblotted using TEPC15 as the primary mAb to detect the presence of PCho. TEPC15-reactive (*dark*) colonies are assumed to have a permissive number of CAAT repeats in the transgenic *lic1* locus, whereas non-reactive (*pale*) colonies are likely to have a non-permissive number of repeats at this locus. Panel **d** shows clone 20.21 (genotype *H. parainfluenzae* 20 Δ*wcfS lic1*$^+$) with phase variant 20.21A indicated by an arrow. Panel **e** shows *H. parainfluenzae* 20.21A with a phase variant colony indicated by an arrow, demonstrating that the loss of TEPC15 reactivity is reversible

synthesise apparently identical lipid A. However, it is becoming clear that the composition of other components of LPS in Gram-negative bacteria can alter the host inflammatory response. PCho on LPS can influence host signalling through binding to the platelet activating factor receptor [34], sialic acid can modulate signalling *via* TLR4 [35] and the presence of O-antigen can dampen the induction of proinflammatory cytokines [36]. We are currently investigating the extent of O-antigen expression in *H. parainfluenzae* and its potential to explain the rarity of *H. parainfluenzae*-related inflammatory disease in the human host.

Preliminary structural analysis of *H. parainfluenzae* T3T1 LPS (Schweda *et al.*, manuscript in preparation) indicates that glycoforms contain a single, phosphorylated Kdo residue, similar to *H. influenzae* LPS. However, the presence of *opsX* and *waaC* homologues in *H. parainfluenzae* suggests that this species has a potential to add a second Kdo residue to a proportion of its LPS glycoforms. *P. multocida*, which also has both Hep I transferase genes, produces heterogeneous LPS molecules of which around 20 % contain Kdo_2 rather than KdoP [37]; the authors speculate that regulation of the KdoP:Kdo_2 ratio may allow the bacteria to survive in different environments. For *H. parainfluenzae* T3T1, LPS analyses followed growth of the bacteria *in vitro*, but it may be that the Kdo_2-containing glycoforms are not optimally synthesized under these growth conditions and are difficult to detect.

Of the Pasteurellaceae species for which genome sequences are publicly available, 13 have putative homologues of both *orfH* and *lgtF* with >60 % aa identity to the *H. parainfluenzae* versions, indicating that LPS in these species

may contain a conserved triheptosyl inner core with a Glc residue linked to HepI. The Glc residue is not essential, as viable *H. parainfluenzae lgtF* mutants have been developed in our laboratory, but it can serve as the attachment point for outer core extensions in *Haemophilus* LPS and the O-antigen in at least some *H. parainfluenzae* isolates. Another substituent of the inner core LPS, PEtn on HepII, is not conserved across all Pasteurellaceae species but in *H. influenzae* and *H. parainfluenzae* at least can be considered a fixed part of the inner core structure.

According to our PCR and sequence analyses of glycosyltransferase genes, most *H. parainfluenzae* strains (12/18) are predicted to add a short non-phase variable *H. influenzae*-like outer core extension to a single inner core sugar through the following linkages: DDHep(α1-6)Glc (on HepI) using LosB1; Glc(α1-3)HepII using Lic2C; or Gal(β1-2)HepIII using LpsB. The latter two extensions have been structurally confirmed in preliminary analysis for *H. parainfluenzae* strains 13 and 20 respectively [18] (Schweda *et al.*, manuscript in preparation). The largely mutually exclusive distribution of the *losB1/lic2C* and *lpsB* loci is intriguing and contrasts with the situation in *H. influenzae*, where a majority of strains are capable of extending from multiple inner core sugars. One possibility is that an *H. parainfluenzae* strain lacking the *losB1/lic2C* locus obtained *lpsB* through horizontal transfer from *H. influenzae*, and that the subsequent spread of this gene was mainly through clonal expansion rather than genetic exchange. Although we cannot rule out the possibility that further LPS biosynthesis genes remain undetected in our *H. parainfluenzae* strains, the general correlation that we observe between the gene patterns and LPS phenotype and structure would suggest that any contribution of uncharacterised genes is unlikely to be significant.

Pettigrew *et al.* [38] found that the *H. influenzae lic2B* gene, which encodes an LPS glucosyltransferase [39], is present more often in OM NT*Hi* isolates than commensal NT*Hi* isolates. Interestingly, whilst five of our *H. parainfluenzae* strains contained a *lic2C* homologue, the *lic2B* gene that flanks it in *H. influenzae* was always absent from the locus (Fig. 3). This could be a contributing factor to the scarcity of *H. parainfluenzae*-related OM.

One of our most striking findings regarding *H. parainfluenzae* was the apparent absence of *H. influenzae*-like phase variably expressed LPS epitopes. Previous work in our laboratory has shown that 25/25 NT*Hi* isolates possess (among others) the phase variable genes *lic3A*, *oafA* and *lic2A/lgtC*, whose products direct the addition of sialic acid, O-acetyl and digalactoside epitopes respectively to the LPS. All three of these epitopes confer some level of resistance to the killing effects of human complement [12]. Sialic acid and digalactoside are also virulence factors in chinchilla OM and rat bacteraemia models respectively [26, 40]. We found no evidence for core sialic acid or O-acetyl expression in our *H. parainfluenzae* isolates, and strain 17 alone is thought to express a (non-phase variable) digalactoside epitope.

PCho is a component of the LPS in several Pasteurellacean species including the bovine respiratory pathogen *Histophilus somni* [41], the human periodontal pathogen *A. actinomycetemcomitans* [42], and *H. influenzae* [43]. $PCho^{+}$ phase variants of *H. influenzae* are associated with greater colonisation of the chinchilla middle ear than $PCho^{-}$ phase variants [44]; and in middle ear NT*Hi* isolates from children with OM, greater PCho expression was related to the prolongation of the disease [45]. Our evidence that *H. parainfluenzae* does not add PCho to its LPS adds weight to the idea that PCho is a key virulence factor for *H. influenzae* OM.

The lack of evidence for any functional SSR in *H. parainfluenzae* suggests that repeat-mediated slipped-strand mispairing of DNA to phase vary the expression of cell surface structures is not a predominant feature of *H. parainfluenzae*. Tetranucleotide SSR mediate the phase variable expression of four main groups of proteins in *H. influenzae*: outer core LPS synthesis enzymes, YadA-like outer membrane proteins (putative adhesins), a type III restriction/modification system methylase (Mod), and the surface expressed haemoglobin and haemoglobin-haptoglobin binding proteins (Hgp). In contrast, the few outer core LPS synthesis genes that appear to be present in *H. parainfluenzae* all lack SSR, the putative adhesin and *hgp* genes in strain T3T1 are not associated with repeat tracts, and there is no T3T1 homologue of *mod*. The successful introduction of a functional *lic1* locus into *H. parainfluenzae* showed that this species can efficiently express *H. influenzae* genes and that the DNA replication machinery in *H. parainfluenzae* allows a transgenic tetranucleotide repeat tract to both shorten and lengthen at a low rate, as is required for a viable phase variation mechanism. The repertoire of DNA replication and repair genes is largely conserved between *H. influenzae* and *H. parainfluenzae* T3T1; the main exception is the presence in *H. parainfluenzae* of *dinB*, which encodes an error-prone DNA polymerase. In *N. meningitidis*, the mutation or over-expression of *dinB* does not alter the rate of tetranucleotide repeat-mediated phase variation [46] but its long term effect on SSR in *H. parainfluenzae* remains to be tested.

The presence of multiple phase variable LPS-related genes in *H. influenzae* contributes to the expression of heterogeneous glycoforms within a population of cells that would be available for short term natural selection. Through having a much restricted repertoire of outer core LPS genes and no phase variation, *H. parainfluenzae* may lack some of the properties endowed on *H. influenzae* that enable it to thrive under circumstances within the host that could favour disease.

Experimental procedures

Bacterial strains and culture

H. parainfluenzae and hybrid strains were a gift from Derrick Crook (University of Oxford) and were isolated in 2001 from the throats of healthy children in Oxfordshire, UK (strains 2, Hy6, 8, 10, Hy11, 14, 16, 20, 22 and 35) and Kulari, Gambia (strains 13, 15, 17, 18, 19, 24, 26, 30, 31 and 34). The strains have been numbered for convenience and their alternative names are given in Online Resource Table S2. The genome sequence of *H. parainfluenzae* strain 26, also known as T3T1, is available as EMBL accession number FQ312002. Of the three *H. influenzae* strains used for comparison, Rd is a type d-derived laboratory strain lacking a capsule, whilst R2846 and R2866 are non-typeable isolates from OM and invasive disease respectively [47, 48]. *Haemophilus* strains were grown at 37 °C on brain heart infusion (BHI) agar (1 %w/v) supplemented with 10 % (v/v) Levinthal's reagent, which provides NAD. Kanamycin (15 $\mu g ml^{-1}$), erythromycin (20 $\mu g ml^{-1}$) or tetracycline (2 $\mu g ml^{-1}$) were added when required for the selection of transformants. *E. coli* DH5α was cultured at 37 °C in LB broth or agar (1 %w/v), with kanamycin (50 $\mu g ml^{-1}$) or erythromycin (300 $\mu g ml^{-1}$) added where necessary.

DNA extraction

Genomic DNA (gDNA) was prepared from *Haemophilus* bacteria as described by Preston *et al.* [49]. Plasmid DNA was isolated from *E. coli* by alkaline lysis [50].

Polymerase chain reaction (PCR)

PCR amplification was carried out using Taq DNA Polymerase and buffer (Invitrogen), 40 ng template gDNA, 0.4 μM each primer (Sigma), 0.4 mM each dNTP and 2.5 mM $MgCl_2$. DNA was amplified for 30 cycles comprising 1 min each of denaturation (94 °C), annealing (50 °C) and extension (72 °C); for expected products of >1.5 kb the extension time was increased to 3 min. Primers used for PCR analysis are listed in Online Resource Table S3.

DNA sequencing and analysis

PCR products and plasmids were sequenced using an ABI-3730 DNA analyser with BigDye Terminator v3.1 (Applied Biosystems). DNA sequence reads were assembled with ContigExpress (VectorNTI, Invitrogen). Sequence similarity analysis was performed using the NCBI basic local alignment search tool (BLAST) [51] with the default algorithm parameters. Protein sequences were aligned with ClustalW [52] and presented using GENEDOC 2.7 [53].

Southern analysis

MfeI-restricted gDNA was fractionated on a 0.8 % agarose gel, visualised under UV light to ensure equal loading per lane and transferred to a nylon membrane (Hybond N+, GE Healthcare) by Southern blotting [50]. For the analysis of LPS synthesis genes, internal gene fragments (0.5–1 kb) were amplified from *H. influenzae* or *H. parainfluenzae* gDNA using PCR and purified using a QIAEX II gel extraction kit (Qiagen). Blots were hybridised with horseradish peroxidase (HRP)-labelled PCR products for 16 h at 42 °C, washed stringently at 42 °C according to the manufacturer's instructions, and visualised on Hyperfilm by HRP-catalysed luminol production (ECL Direct kit, GE Healthcare). For the analysis of tetranucleotide repeats, 20-base oligonucleotides (Sigma) were radiolabelled and used as probes as described by Fox *et al.* [54].

General cloning and transformation procedures

PCR products were cloned in *E. coli* using the pSCA or pSCA-amp-kan vector system (Stratagene) following manufacturer's guidelines. Further DNA manipulation was carried out using restriction enzymes and T4 DNA ligase from Roche and New England Biolabs according to the manufacturer's instructions, before amplification of constructs in chemically competent *E. coli* DH5α. *H. parainfluenzae* strains were transformed with 500 ng linearised plasmid DNA using the static aerobic method described by Gromkova and Goodgal [55], using 20 mM $MgSO_4$. Correct transformants were confirmed by PCR.

Preparation of plasmids and *H. parainfluenzae* mutant strains

Plasmid pRY29 was constructed in order to introduce the *lic1* locus into *H. parainfluenzae* by the homologous recombination of flanking DNA. The target sequence (*tetB*, part of the Tn10 transposon) was amplified from *H. parainfluenzae* T3T1 gDNA using primers P1 and P2 (Online Resource Table S4) and cloned in pSCA-amp-kan. Genes *lic1A-lic1D* (amplified from *H. influenzae* Rd gDNA using primers P3 and P4) and *ermC* (erythromycin resistance gene released from pER2) [56] were inserted in tandem into the centre of the target sequence. *H. parainfluenzae* strain 19 was transformed with pRY29 to produce the PCho-expressing mutant strains 19.4 and 19.5. *H. parainfluenzae* strain 20, which does not contain *tetB*, was first transformed with gDNA from strain T3T1 and plated on media containing tetracycline to select for Tn10 insertion. Transformant 20.1 was then transformed with pRY29 to produce the PCho-expressing strains 20.12 and 20.13.

A double mutant strain, 20.21, with the genotype *H. parainfluenzae* 20 Δ*wcfS lic1*$^+$ was constructed for phase variation

experiments because *H. parainfluenzae wcfS* mutant colonies dissociate more readily than wild type strains from nitrocellulose membranes, leading to a lower background signal with mAb TEPC15 (RY, unpublished data). *WcfS* (also known as *wbaP*) is involved in O-antigen biosynthesis [18]. For the *wcfS* deletion plasmid (pRY41), primers P5 and P6 were used to amplify a fragment of the *H. parainfluenzae* T3T1 O-antigen locus, and the product was cloned in pSCA. Part of the plasmid was amplified with primers P7 and P8 and the PCR product was ligated to a kanamycin resistance gene derived from pUC4kan. This replaced base pairs 114–344 of *wcfS* with the kanamycin resistance gene. In strain 20.21 the *lic1* locus was inserted into a stable, single-copy chromosomal location (the pseudogene *PARA_18810/18820*) using the plasmid construct pRY27. This is identical to pRY29 with the exception of the flanking sequence for recombination, which was amplified from *H. parainfluenzae* T3T1 gDNA using primers P9 and P10.

Colony and dot immunoblotting

For dot immunoblots, bacteria grown on BHI agar were suspended in PBS, adjusted to equal optical densities and boiled for 5 min to lyse. 3 μl of each suspension was pipetted onto a gridded Protran membrane (Whatman). PCho was detected using mAb TEPC15 (Sigma; 1:2000 dilution). Galα(1–4)βGal was detected using mAb 4C4 (1:200 dilution; provided by E.J. Hansen, University of Texas), whilst Gal extending from HepIII was detected using mAb L6A9 (1:2000 dilution; NRC, Ottawa). The secondary antibody for all experiments was goat anti-mouse IgA alkaline phosphatase antibody (Sigma; 1:20,000 dilution), visualised with BCIP/NBT Substrate Solution (Perkin Elmer). Colony immunoblots were performed as described previously [57] using mAb TEPC15 as above.

Electrophoretic analysis of LPS

Cell lysates were treated with proteinase K (50 $\mu g ml^{-1}$) at 60 °C for 3 h, then fractionated by tricine SDS-PAGE [58]. LPS was visualised by staining the gel with silver (GE Healthcare) following the manufacturer's instructions.

Sequence data

The *H. parainfluenzae* DNA sequence data have been submitted to the GenBank database under the following accession numbers: strain 8 *lpsB*, JQ622291; strain 10 *lpsB*, JQ622292; strain 15 *lpsB*, JQ622293; strain 20 *lpsB*, JQ622294; strain 35 *lpsB*, JQ622295; strain 13 *lic2C*, JQ622296; strain 17 *losA1*, JQ622297; strain 17 *losB1*, JQ622298; strain 17 *lic2A*, JQ622299; strain 17 *lgtC*, JQ622300.

Acknowledgments This work was supported by grants from the Medical Research Council, UK. We thank Derrick Crook and Abdel Elamin for providing strains, Peter Power for bioinformatic assistance, Elke Schweda and Richard Moxon for helpful discussions, and the Weatherall Institute of Molecular Medicine Sequencing Service, John Radcliffe Hospital, Oxford for DNA sequencing.

References

1. Rivers, T.M.: Influenza-like bacilli. Growth of influenza-like bacilli on media containing only an autoclave-labile substance as an accessory food factor. Bull. Johns Hopkins Hosp. **33**, 429–431 (1922)
2. Norskov-Lauritsen, N., Bruun, B., Kilian, M.: Multilocus sequence phylogenetic study of the genus *Haemophilus* with description of *Haemophilus pittmaniae* sp. nov. Int. J. Syst. Evol. Microbiol. **55** (Pt 1), 449–456 (2005)
3. Black, C.T., Kupferschmid, J.P., West, K.W., Grosfeld, J.L.: *Haemophilus parainfluenzae* infections in children, with the report of a unique case. Rev. Infect. Dis. **10**(2), 342–346 (1988)
4. Darras-Joly, C., Lortholary, O., Mainardi, J.L., Etienne, J., Guillevin, L., Acar, J.: *Haemophilus endocarditis*: report of 42 cases in adults and review. *Haemophilus* Endocarditis Study Group. Clin. Infect. Dis. **24**(6), 1087–1094 (1997)
5. Cooney, T.G., Harwood, B.R., Meisner, D.J.: *Haemophilus parainfluenzae* thoracic empyema. Arch. Intern. Med. **141**(7), 940–941 (1981)
6. Frankard, J., Rodriguez-Villalobos, H., Struelens, M.J., Jacobs, F.: *Haemophilus parainfluenzae*: an underdiagnosed pathogen of biliary tract infections? Eur. J. Clin. Microbiol. Infect. Dis. **23**(1), 46–48 (2004)
7. Scheifele, D.W., Fussell, S.J.: Frequency of ampicillin-resistant *Haemophilus parainfluenzae* in children. J. Infect. Dis. **143**(3), 495–498 (1981)
8. Taylor, D.C., Cripps, A.W., Clancy, R.L., Murree-Allen, K., Hensley, M.J., Saunders, N.A., Sutherland, D.C.: Biotypes of *Haemophilus parainfluenzae* from the respiratory secretions in chronic bronchitis. J. Med. Microbiol. **36**(4), 279–282 (1992)
9. Deadman, M.E., Hermant, P., Engskog, M., Makepeace, K., Moxon, E.R., Schweda, E.K., Hood, D.W.: Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. Infect. Immun. **77** (6), 2376–2384 (2009)
10. Deadman, M.E., Lundstrom, S.L., Schweda, E.K., Moxon, E.R., Hood, D.W.: 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**(40), 29455–29467 (2006)
11. Helander, I.M., Lindner, B., Brade, H., Altmann, K., Lindberg, A.A., Rietschel, E.T., Zahringer, U.: Chemical structure of the lipopolysaccharide of *Haemophilus influenzae* strain I-69 Rd-/b+. Description of a novel deep-rough chemotype. Eur. J. Biochem. **177**(3), 483–492 (1988)
12. Schweda, E.K., Richards, J.C., Hood, D.W., Moxon, E.R.: Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*: implication in virulence. Int. J. Med. Microbiol. **297**(5), 297–306 (2007)
13. Weiser, J.N., Lindberg, A.A., Manning, E.J., Hansen, E.J., Moxon, E.R.: Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae*. Infect. Immun. **57**(10), 3045–3052 (1989)
14. Power, P.M., Sweetman, W.A., Gallacher, N.J., Woodhall, M.R., Kumar, G.A., Moxon, E.R., Hood, D.W.: Simple sequence repeats

in *Haemophilus influenzae*. Infect. Genet. Evol. **9**(2), 216–228 (2009)
15. Weiser, J.N., Love, J.M., Moxon, E.R.: The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell **59**(4), 657–665 (1989)
16. Roberts, M.C., Mintz, C.S., Morse, S.A.: Characterization of *Haemophilus parainfluenzae* strains with low-Mr or ladder-like lipopolysaccharides. J. Gen. Microbiol. **132**(3), 611–616 (1986)
17. Pollard, A., St Michael, F., Connor, L., Nichols, W., Cox, A.: Structural characterization of *Haemophilus parainfluenzae* lipooligosaccharide and elucidation of its role in adherence using an outer core mutant. Can. J. Microbiol. **54**(11), 906–917 (2008)
18. Vitiazeva, V., Twelkmeyer, B., Young, R., Hood, D.W., Schweda, E.K.: Structural studies of the lipopolysaccharide from *Haemophilus parainfluenzae* strain 20. Carbohydr. Res. **346**(14), 2228–2236 (2011)
19. Cipolla, L., Polissi, A., Airoldi, C., Galliani, P., Sperandeo, P., Nicotra, F.: The Kdo biosynthetic pathway toward OM biogenesis as target in antibacterial drug design and development. Curr. Drug Discov. Technol. **6**(1), 19–33 (2009)
20. White, K.A., Kaltashov, I.A., Cotter, R.J., Raetz, C.R.: A monofunctional 3-deoxy-D-*manno*-octulosonic acid (Kdo) transferase and a Kdo kinase in extracts of *Haemophilus influenzae*. J. Biol. Chem. **272**(26), 16555–16563 (1997)
21. Wang, X., Quinn, P.J.: Lipopolysaccharide: Biosynthetic pathway and structure modification. Prog. Lipid Res. **49**(2), 97–107 (2010)
22. Gronow, S., Brabetz, W., Lindner, B., Brade, H.: OpsX from *Haemophilus influenzae* represents a novel type of heptosyltransferase I in lipopolysaccharide biosynthesis. J. Bacteriol. **187**(17), 6242–6247 (2005)
23. Wright, J.C., Hood, D.W., Randle, G.A., Makepeace, K., Cox, A.D., Li, J., Chalmers, R., Richards, J.C., Moxon, E.R.: *lpt6*, a gene required for addition of phosphoethanolamine to inner-core lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae*. J. Bacteriol. **186**(20), 6970–6982 (2004)
24. Hood, D.W., Deadman, M.E., Cox, A.D., Makepeace, K., Martin, A., Richards, J.C., Moxon, E.R.: 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**(Pt 7), 2089–2097 (2004)
25. Hood, D.W., Deadman, M.E., Engskog, M.K., Vitiazeva, V., Makepeace, K., Schweda, E.K., Moxon, R.: Genes required for the synthesis of heptose-containing oligosaccharide outer core extensions in *Haemophilus influenzae* lipopolysaccharide. Microbiology **156**, 3421–3431 (2010)
26. Griffin, R., Bayliss, C.D., Herbert, M.A., Cox, A.D., Makepeace, K., Richards, J.C., Hood, D.W., Moxon, E.R.: Digalactoside expression in the lipopolysaccharide of *Haemophilus influenzae* and its role in intravascular survival. Infect. Immun. **73**(10), 7022–7026 (2005)
27. Chen, C., Kittichotirat, W., Si, Y., Bumgarner, R.: Genome sequence of *Aggregatibacter actinomycetemcomitans* serotype c strain D11S-1. J. Bacteriol. **191**(23), 7378–7379 (2009)
28. McCrea, K.W., Xie, J., Marrs, C.F., Gilsdorf, J.R.: Prevalence of genetic differences in phosphorylcholine expression between nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. BMC Microbiol. **10**, 286 (2010)
29. Vitovski, S., Dunkin, K.T., Howard, A.J., Sayers, J.R.: Nontypeable *Haemophilus influenzae* in carriage and disease: a difference in IgA1 protease activity levels. JAMA **287**(13), 1699–1705 (2002)
30. Liljemark, W.F., Bloomquist, C.G., Lai, C.H.: Clustering of an outer membrane adhesin of *Haemophilus parainfluenzae*. Infect. Immun. **60**(2), 687–689 (1992)
31. Bryant, C.E., Spring, D.R., Gangloff, M., Gay, N.J.: The molecular basis of the host response to lipopolysaccharide. Nat. Rev. Microbiol. **8**(1), 8–14 (2010)
32. Khan, S.A., Everest, P., Servos, S., Foxwell, N., Zahringer, U., Brade, H., Rietschel, E.T., Dougan, G., Charles, I.G., Maskell, D.J.: A lethal role for lipid A in *Salmonella* infections. Mol. Microbiol. **29**(2), 571–579 (1998)
33. Tuyau, J.E., Sims, W.: Aspects of the pathogenicity of some oral and other haemophili. J. Med. Microbiol. **16**(4), 467–475 (1983)
34. Swords, W.E., Ketterer, M.R., Shao, J., Campbell, C.A., Weiser, J.N., Apicella, M.A.: Binding of the non-typeable *Haemophilus influenzae* lipooligosaccharide to the PAF receptor initiates host cell signalling. Cell. Microbiol. **3**(8), 525–536 (2001)
35. Howard, M.D., Willis, L., Wakarchuk, W., St Michael, F., Cox, A., Horne, W.T., Hontecillas, R., Bassaganya-Riera, J., Lorenz, E., Inzana, T.J.: Genetics and molecular specificity of sialylation of *Histophilus somni* lipooligosaccharide (LOS) and the effect of LOS sialylation on Toll-like receptor-4 signaling. Vet. Microbiol. **153**(1–2), 163–172 (2011)
36. Billips, B.K., Schaeffer, A.J., Klumpp, D.J.: Molecular basis of uropathogenic *Escherichia coli* evasion of the innate immune response in the bladder. Infect. Immun. **76**(9), 3891–3900 (2008)
37. Harper, M., Boyce, J.D., Cox, A.D., St Michael, F., Wilkie, I.W., Blackall, P.J., Adler, B.: *Pasteurella multocida* expresses two lipopolysaccharide glycoforms simultaneously, but only a single form is required for virulence: identification of two acceptor-specific heptosyl I transferases. Infect. Immun. **75**(8), 3885–3893 (2007)
38. Pettigrew, M.M., Foxman, B., Marrs, C.F., Gilsdorf, J.R.: 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**(7), 3551–3556 (2002)
39. Twelkmeyer, B., Deadman, M.E., Haque, E., Li, J., Hood, D.W., Schweda, E.K.: The role of *lic2B* in lipopolysaccharide biosynthesis in *Haemophilus influenzae* strain Eagan. Carbohydr. Res. **346**(10), 1262–1266 (2011)
40. Bouchet, V., Hood, D.W., Li, J., Brisson, J.R., Randle, G.A., Martin, A., Li, Z., Goldstein, R., Schweda, E.K., Pelton, S.I., Richards, J.C., Moxon, E.R.: Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharide and is a major virulence factor in experimental otitis media. Proc. Natl. Acad. Sci. U. S. A. **100**(15), 8898–8903 (2003)
41. Elswaifi, S.F., St Michael, F., Sreenivas, A., Cox, A., Carman, G.M., Inzana, T.J.: Molecular characterization of phosphorylcholine expression on the lipooligosaccharide of *Histophilus somni*. Microb. Pathog. **47**(4), 223–230 (2009)
42. Schenkein, H.A., Barbour, S.E., Berry, C.R., Kipps, B., Tew, J.G.: Invasion of human vascular endothelial cells by *Actinobacillus actinomycetemcomitans via* the receptor for platelet-activating factor. Infect. Immun. **68**(9), 5416–5419 (2000)
43. Weiser, J.N., Shchepetov, M., Chong, S.T.: Decoration of lipopolysaccharide with phosphorylcholine: a phase-variable characteristic of *Haemophilus influenzae*. Infect. Immun. **65**(3), 943–950 (1997)
44. Tong, H.H., Blue, L.E., James, M.A., Chen, Y.P., DeMaria, T.F.: Evaluation of phase variation of nontypeable *Haemophilus influenzae* lipooligosaccharide during nasopharyngeal colonization and development of otitis media in the chinchilla model. Infect. Immun. **68**(8), 4593–4597 (2000)
45. Fujita, K., Hirano, T., Kodama, S., Suzuki, M.: Prognostic impact of phosphorylcholine expression in nontypeable *Haemophilus influenzae* in otitis media with effusion. Acta Otolaryngol. **129**(8), 832–838 (2009)
46. Martin, P., Sun, L., Hood, D.W., Moxon, E.R.: Involvement of genes of genome maintenance in the regulation of phase variation frequencies in *Neisseria meningitidis*. Microbiology **150**(Pt 9), 3001–3012 (2004)
47. Erwin, A.L., Bonthuis, P.J., Geelhood, J.L., Nelson, K.L., McCrea, K.W., Gilsdorf, J.R., Smith, A.L.: Heterogeneity in tandem octanucleotides within *Haemophilus influenzae* lipopolysaccharide

biosynthetic gene *losA* affects serum resistance. Infect. Immun. **74** (6), 3408–3414 (2006)
48. Erwin, A.L., Nelson, K.L., Mhlanga-Mutangadura, T., Bonthuis, P.J., Geelhood, J.L., Morlin, G., Unrath, W.C., Campos, J., Crook, D.W., Farley, M.M., Henderson, F.W., Jacobs, R.F., Muhlemann, K., Satola, S.W., van Alphen, L., Golomb, M., Smith, A.L.: Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. Infect. Immun. **73**(9), 5853–5863 (2005)
49. Preston, A., Maskell, D., Johnson, A., Moxon, E.R.: Altered lipopolysaccharide characteristic of the I69 phenotype in Haemophilus influenzae results from mutations in a novel gene, isn. J. Bacteriol. **178**(2), 396–402 (1996)
50. Sambrook, J., Fritsch, E.F., Maniatis, T.: Molecular Cloning. Cold Spring Harbor Laboratory Press, New York (1989)
51. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**(17), 3389–3402 (1997)
52. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G.: Clustal W and Clustal X version 2.0. Bioinformatics **23**(21), 2947–2948 (2007)
53. Nicholas, K.B., Nicholas, H.B., Deerfield, D.W.I.: GeneDoc: Analysis and visualization of genetic variation. EMBNEW.NEWS **4**, 14 (1997)
54. Fox, K.L., Yildirim, H.H., Deadman, M.E., Schweda, E.K., Moxon, E.R., Hood, D.W.: Novel lipopolysaccharide biosynthetic genes containing tetranucleotide repeats in *Haemophilus influenzae*, identification of a gene for adding O-acetyl groups. Mol. Microbiol. **58**(1), 207–216 (2005)
55. Gromkova, R., Goodgal, S.: Transformation by plasmid and chromosomal DNAs in *Haemophilus parainfluenzae*. Biochem. Biophys. Res. Commun. **88**(4), 1428–1434 (1979)
56. Jennings, M.P., van der Ley, P., Wilks, K.E., Maskell, D.J., Poolman, J.T., Moxon, E.R.: Cloning and molecular analysis of the *galE* gene of *Neisseria meningitidis* and its role in lipopolysaccharide biosynthesis. Mol. Microbiol. **10**(2), 361–369 (1993)
57. Fox, K.L., Li, J., Schweda, E.K., Vitiazeva, V., Makepeace, K., Jennings, M.P., Moxon, E.R., Hood, D.W.: Duplicate copies of lic1 direct the addition of multiple phosphocholine residues in the lipopolysaccharide of *Haemophilus influenzae*. Infect. Immun. **76** (2), 588–600 (2008)
58. Lesse, A.J., Campagnari, A.A., Bittner, W.E., Apicella, M.A.: Increased resolution of lipopolysaccharides and lipooligosaccharides utilizing tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis. J. Immunol. Methods **126**(1), 109–117 (1990)
59. Hood, D.W., Cox, A.D., Wakarchuk, W.W., Schur, M., Schweda, E.K., Walsh, S.L., Deadman, M.E., Martin, A., Moxon, E.R., Richards, J.C.: Genetic basis for expression of the major globotetraose-containing lipopolysaccharide from *H. influenzae* strain Rd (RM118). Glycobiology **11**(11), 957–967 (2001)
60. Risberg, A., Masoud, H., Martin, A., Richards, J.C., Moxon, E.R., Schweda, E.K.: Structural analysis of the lipopolysaccharide oligosaccharide epitopes expressed by a capsule-deficient strain of *Haemophilus influenzae* Rd. Eur. J. Biochem. **261**(1), 171–180 (1999)