

# SCIENTIFIC REPORTS



OPEN

## Acquisition of the capsule locus by horizontal gene transfer in *Neisseria meningitidis* is often accompanied by the loss of UDP-GalNAc synthesis

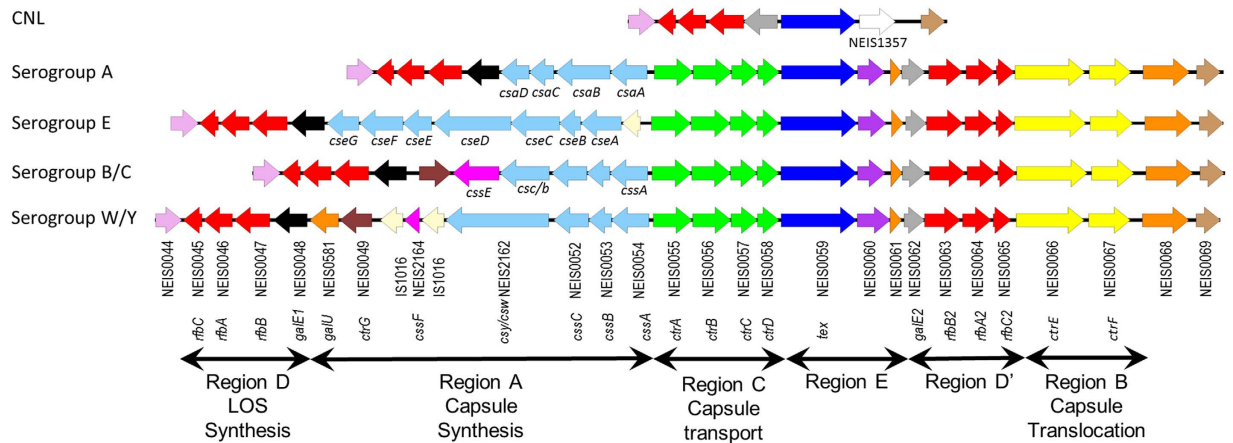
Stephanie N. Bartley<sup>1,2,3,\*</sup>, Shakeel Mowlaboccus<sup>1,2,\*</sup>, Christopher A. Mullally<sup>1,2</sup>, Keith A. Stubbs<sup>3</sup>, Alice Vrieling<sup>3</sup>, Martin C. J. Maiden<sup>4</sup>, Odile B. Harrison<sup>4</sup>, Timothy T. Perkins<sup>1,2</sup> & Charlene M. Kahler<sup>1,2</sup>

Pathogenic meningococci have acquired a 24 kb capsule synthesis island (*cps*) by horizontal gene transfer which consists of a synthetic locus and associated capsule transport genes flanked by repetitive Regions D and D'. Regions D and D' contain an intact gene encoding a UDP-galactose epimerase (*galE1*) and a truncated remnant (*galE2*), respectively. In this study, GalE protein alleles were shown to be either mono-functional, synthesising UDP-galactose (UDP-Gal), or bi-functional, synthesising UDP-Gal and UDP-galactosamine (UDP-GalNAc). Meningococci possessing a capsule null locus (*cnl*) typically possessed a single bi-functional *galE*. Separation of functionality between *galE1* and *galE2* alleles in meningococcal isolates was retained for all serogroups except serogroup E which has a synthetic requirement for UDP-GalNAc. The truncated *galE2* remnant in Region D' was also phylogenetically related to the bi-functional *galE* of the *cnl* locus suggesting common ancestry. A model is proposed in which the illegitimate recombination of the *cps* island into the *galE* allele of the *cnl* locus results in the formation of Region D' containing the truncated *galE2* locus and the capture of the *cps* island en bloc. The retention of the duplicated Regions D and D' enables inversion of the synthetic locus within the *cps* island during bacterial growth.

The *Neisseria* genus contains eleven species that colonize the oropharyngeal and urogenital mucosa of humans and of these, *Neisseria gonorrhoeae* and *N. meningitidis*, are of medical importance. *N. gonorrhoeae* colonises the mucosal surfaces of the urogenital tract in both males and females<sup>1</sup>. *N. meningitidis* is predominantly an opportunistic pathogen, asymptotically colonising the mucosa of the nasopharynx of approximately 10% of the adult population<sup>2,3</sup> but occasionally invades the host, resulting in septicaemia and meningitis<sup>4</sup>.

In both *N. meningitidis* and *N. gonorrhoeae*, interactions with host cells are modulated by the glycome on the bacterial surface which includes the lipooligosaccharide (LOS) and the glycosylation status of the type IV pilin<sup>5</sup>. Meningococci are distinguished from gonococci by the presence of a capsule polysaccharide synthesis (*cps*) island which encodes a capsule that facilitates systemic infection<sup>5</sup>. The *cps* island has a lower G/C-content than the core meningococcal genome consistent with acquisition via horizontal gene transfer (HGT) by recombination<sup>6</sup>. The general organisation of the *cps* island is as follows: Region A is responsible for the synthesis of the capsule polymer; Region B (*ctrEF*) and Region C (*ctrABCD*) contain genes responsible for transporting the capsule polymer to the bacterial surface while Region D and Region D' contain an intact and truncated remnant copy of a gene encoding UDP-galactose epimerase, known as *galE1* and *galE2*, respectively. Region E contains a gene of unknown function termed *tex* which is a homologue of a transcription factor. GalE is necessary for the synthesis

<sup>1</sup>School of Biomedical Sciences, University of Western Australia, Perth, Australia. <sup>2</sup>The Marshall Centre for Infectious Disease Research and Training, University of Western Australia, Perth, Australia. <sup>3</sup>School of Molecular Sciences, University of Western Australia, Perth, Australia. <sup>4</sup>University of Oxford, Department of Zoology, South Parks Road, Oxford OX1 3PS, United Kingdom. <sup>\*</sup>These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.M.K. (email: Charlene.Kahler@uwa.edu.au)



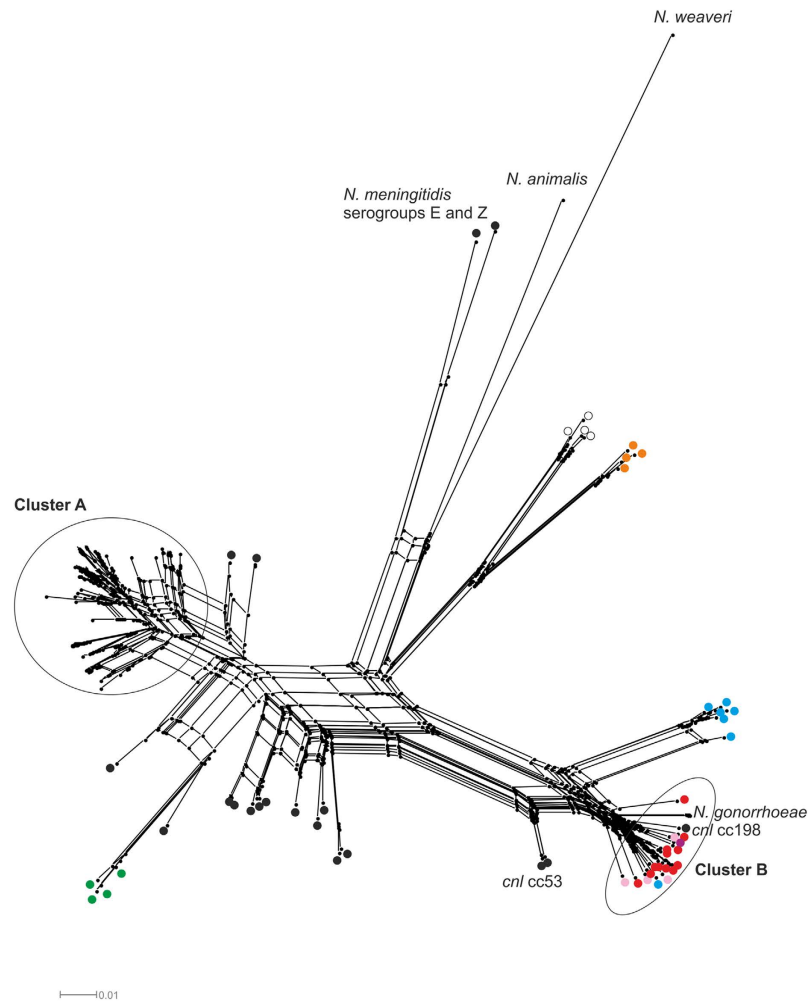
**Figure 1. Genetic organisation of *cnl* and *cps* loci in *N. meningitidis*.** Schematic organisation of the capsule synthesis (*cps*) island in meningococci. Serogroups A, B/C, W/Y and E are represented. Region A (light blue) is accompanied by Region C (bright green) encoding four capsule transport encoding genes *ctrA-D*. Region E (navy) encodes *tex*, a gene of unknown function homologue of a transcription factor. Region A-C-E is flanked by an inverted repeat segment, Region D and Region D'. Region D contains *galE1* (black), encoding the UDP-galactose 4-epimerase and three genes termed *rfbBAC* (red) of unknown function. Region D' contains a truncated *galE2* (grey) and *rfbB2-A2-C2* (red). Region B (yellow) contains *ctrE* and *ctrF* which encode KDO transferases to initiate polysaccharide biosynthesis. Region A (light blue) contains polysaccharide biosynthesis genes specific for each serogroup and *ctrG* (brown) which is found in *cps* islands encoding sialic acid biosynthesis pathways. The four genes *csaA*, *csaB*, *csaC* and *csaD* are required for the synthesis of a homopolymer of *O*-acetylated,  $\alpha 1 \rightarrow 6$ -linked ManNAc 1-phosphate of serogroup A meningococci. Serogroup B, C, W and Y meningococci contain three genes required for sialic acid biosynthesis, *cssA*, *cssB* and *cssC*. Each cluster has a serogroup specific polysialyltransferase, capsule synthesis serogroup B (*csb*), capsule synthesis serogroup C (*csc*), capsule synthesis serogroup W (*csw*) and capsule synthesis serogroup Y (*csy*). Serogroup C polysaccharides are *O*-acetylated by *CssE* while serogroup Y/W are acetylated by *CssF*. The biosynthesis cluster for serogroup W and Y also contains a *galU* encoding a UTP: $\alpha$ -D-glucose-1-phosphate uridylyltransferase. Serogroup B consists of a homopolymer of  $\alpha 2 \rightarrow 8$  linked sialic acid, serogroup C consists of  $\alpha 2 \rightarrow 9$  linked sialic acid, serogroup W consists of repeating units of 4-*O*- $\alpha$ -D-galactopyranosyl- $\beta$ -D-*N*-acetylneuraminic acid and serogroup Y consists of repeating units of 4-*O*- $\alpha$ -D-glucopyranosyl- $\beta$ -D-*N*-acetylneuraminic acid. Serogroup E polysaccharide consists of alternating D-galactosamine and 2-keto-3-deoxyoctulosonate (KDO) residues synthesized by the seven genes *cseA-G*. The capsule null locus (*cnl*) of non-disease causing meningococci consists of *galE-rfbBAC* and *tex*. The figure was drawn using Easyfig<sup>50</sup> and is similar to the figure presented by Harrison *et al.*<sup>46</sup>.

of UDP-galactose (UDP-Gal) which is utilised in for the synthesis of capsule polymers from serogroup W/Y, LOS and protein glycosylation<sup>5</sup>.

Non-pathogenic meningococci frequently possess a capsule null locus (*cnl*) with a single copy of Region D with an intact *galE* and Region E (Fig. 1)<sup>7</sup>. Evolutionary studies have suggested that meningococci acquired the *cps* island from potential donors such as *Pasteurella multocida* and *Haemophilus influenzae* through mosaic HGT events<sup>8–10</sup>. Although the ancestral events leading to the formation of the *cps* island are unclear, serogroup switching can result in the replacement of the entire synthetic Region A in instances where serogroups B or C switch to serogroups A, W or Y (Fig. 1) indicating recombination events in this region continue to occur<sup>11,12</sup>.

Gonococci also possess a *cnl* locus (Fig. 1), and have the additional ability to synthesise UDP-*N*-acetyl galactosamine (UDP-GalNAc) which is used to terminate the non-reducing terminus of LOS<sup>13</sup>. In meningococci, GalNAc is not a component of LOS<sup>5,14</sup>, but is required for the synthesis of meningococcal serogroup E and Z capsules<sup>15,16</sup>. Although many of the biosynthetic pathways of *Neisseria* spp. have been elucidated, the mechanism by which UDP-GalNAc is synthesised has not been proposed. Bacteria synthesise UDP-GalNAc via either a bifunctional UDP-galactose 4-epimerase (*GalE*) or a UDP-GalNAc 4-epimerase (*GNE*). Since *Neisseria* spp. do not harbour a *GNE* homologue, the neisserial *GalE1* could be a bi-functional epimerase that synthesizes both UDP-Gal and UDP-GalNAc from the substrates UDP-glucose (UDP-Glc) and UDP-*N*-acetylglucosamine (UDP-GlcNAc), respectively.

This study examined the functional diversity of the *galE* locus and found that *Neisseria* spp. possess both mono-functional and bi-functional UDP-galactose epimerases. While *N. gonorrhoeae* and meningococci with a *cnl* locus possess bi-functional *galE* alleles which are phylogenetically related, meningococci carrying the *cps* locus possessed both bi- and mono-functional alleles for *galE1* and *galE2*. The phylogenetic relations of the *galE* alleles and their associative properties with serogroup and clonal complex has provided further evidence for the hypothetical model of the recombination events during HGT of the *cps* island.



**Figure 2. Phylogenetic reconstruction using *galeE* alleles from commensal and pathogenic *Neisseria* spp. using an un-rooted neighbour-net algorithm.** *N. gonorrhoeae* and commensal species are represented by circles with different colours denoting species. These were: *N. meningitidis* is black; *N. lactamica* is red; *N. polysacchareae* is pink; *N. bergeri* is mauve; *N. cinerea* is blue; *N. elongata* is orange; *N. animalis* and *N. weaveri* are grey; *N. mucosa* is white and *N. subflava* is green. *N. meningitidis* isolates are found in cluster A but also on branches in-between clusters A and B as well as the *cnl* in cluster B. *N. gonorrhoeae* are only found in one branch of cluster B and are labelled as *N. gonorrhoeae* without a symbol.

## Results

***galeE1* alleles form two distinct phylogenetic clusters in *Neisseria* species.** The distribution of *galeE1* alleles (NEIS0048) across the genus *Neisseria* was assessed using a defined set of 194 isolates for which genomic data was available in the PubMLST database ([www.pubmlst.org/neisseria](http://www.pubmlst.org/neisseria), Supplementary Table S1). A total of 107 unique *galeE* alleles, comprising 65 *galeE1* alleles from meningococci and 42 *galeE* alleles from the other species, represented the top 24 most commonly occurring NEIS0048 alleles within the PubMLST database (> 0.38% occurrences). An un-rooted neighbour-net tree (Fig. 2) revealed that the alleles formed two main clusters. Cluster A contained encapsulated meningococci ( $p$ -distance = 0.092), while cluster B contained *N. gonorrhoeae* and the non-pathogenic *Neisseria* species, *N. lactamica* and *N. polysacchareae* as well as unencapsulated meningococci possessing the *cnl* locus (overall mean  $p$ -distance = 0.063) (Fig. 2). Several other less distinct clusters with deeper roots were observed which corresponded with *N. subflava*, *N. oralis*, *N. elongata*, and *N. cinerea*. Two meningococcal *galeE1* variants belonging to serogroup E and Z meningococci were found on the same branch as those from *N. animalis* and *N. weaveri* which were more distantly related to Cluster A and B (overall  $p$ -distance = 0.266). Overall, the distribution of the *galeE1* alleles into at least two distinct clusters suggest that these alleles are under diversifying selection which may be related to function between species.

***Neisseria* spp. possess mono- and bi-functional GalE epimerases.** To investigate whether NEIS0048 alleles from each phylogenetic cluster had different functions, the *galeE1* of *N. gonorrhoeae* strain FA1090 (GalE1 allele 17 = GalE\_17) from cluster B and *N. meningitidis* strain MC58 (GalE1 allele 2 = GalE\_2) from cluster A were cloned into pET15b to create GalE\_17::Hisx6 and GalE\_2::Hisx6. GalE\_17::Hisx6 could epimerise both UDP-Glc and UDP-GlcNAc indicating that this variant is bi-functional, with the equilibrium favouring the

Enzyme	Substrate	Equilibrium ratio UDP-Glc:UDP-Gal	Equilibrium ratio UDP-GlcNAc:UDP-GalNAc
Gonococcal GalE_17	UDP-Glucose	39:61	—
	UDP-GlcNAc	—	32:68
GalE_17 (S299F)	UDP-Glucose	59:41	—
	UDP-GlcNAc	—	99.2:0.8
Meningococcal GalE_2	UDP-Glucose	63:37	—
	UDP-GlcNAc	—	99.5:0.5
GalE_2 (F300S)	UDP-Glucose	67:33	—
	UDP-GlcNAc	—	73:27
GalE_236	UDP-Glucose	23:77	—
	UDP-GlcNAc	—	25:75

**Table 1. Analysis of UDP-Glucose and UDP-GlcNAc 4-epimerase activity of GalE alleles and their mutants.**

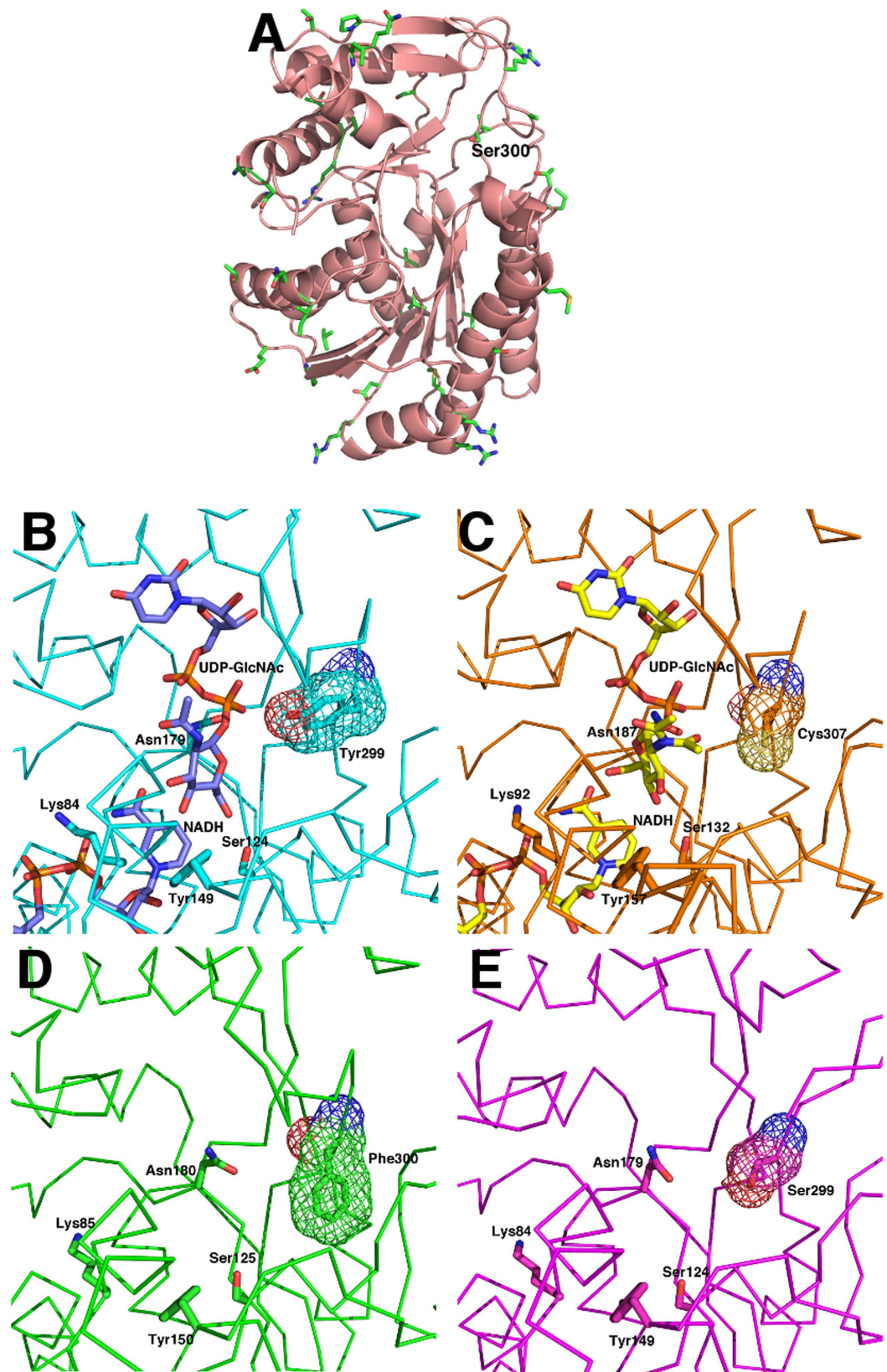
production of UDP-Gal and UDP-GalNAc (Table 1). In comparison, GalE\_2::Hisx6 would only accept UDP-Glc as a substrate and could not epimerise UDP-GlcNAc indicating that this enzyme is mono-functional. Unlike the gonococcal GalE\_17 allele, the meningococcal GalE\_2 had an equilibrium favouring the formation of UDP-Glc at 63% of all products.

**Mono- and bi-functionality of GalE epimerases is determined by a single amino acid residue in the active site.** Previous investigations of known mono-functional and bi-functional GalE proteins have shown that a single amino acid residue is the main factor in the ability of the enzyme to possess singular or dual modes of activity. Thoden *et al.*<sup>17</sup> compared the GalE protein from *H. sapiens* (Group 2) to the *E. coli* (Group 1) homologue and demonstrated that four conserved residues hold the sugar in a productive conformation (Supplementary Fig. 1) while a fifth residue affected whether UDP-Glc or UDP-GlcNAc could enter the active site cleft. In *E. coli* GalE and the human GalE, this position is occupied by a tyrosine (Y299) residue or a cysteine (C307), respectively. When the tyrosine residue of *E. coli* GalE was mutated to cysteine (Y299C) it switched the native enzyme from a mono-functional to a bi-functional mode while the converse occurred in *Yersinia enterocolitica* GalE<sup>18</sup>.

To understand the contribution of amino acid residues to the active site of GalE, 107 *galE* alleles from ten *Neisseria* spp. (Supplementary Table S1) were translated and aligned using CLUSTALW2<sup>19</sup>. As expected, residues K85, S125, Y150 and N180 which correspond to the four residues co-ordinating the pyranose ring in the active site were absolutely conserved (Supplementary Fig. 1). The fifth position was occupied by a serine (S299) residue in gonococcal GalE amino acid sequences and phenylalanine (F300) in most meningococcal alleles. Since the serine and phenylalanine residues are similar in size to cysteine in bi-functional enzymes and tyrosine in mono-functional enzymes, respectively, we proposed that these residues govern the functionality of GalE. Using PHYRE2<sup>20</sup>, homology models using GalE amino acid sequences of *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strain MC58 were built. These models and the crystal structures of the *E. coli* and human enzymes in the presence of NADH and UDP-GlcNAc are shown in Fig. 3. To address whether variability within the amino acid sequences of neisseria GalE proteins contributed to the binding site pocket, 38 non-conserved residues derived from the alignment of 103 GalE1 alleles (excluding the four most distantly related alleles from *N. weaveri* GalE\_115, *N. animalis* GalE\_110, and *N. meningitidis* GalE\_14 and GalE\_15) were mapped to the predicted structure of GalE\_2 from *N. meningitidis* MC58 (Fig. 3A). The non-conserved residues were localized predominantly to the external surface of the protein and were distant from the active site. Conservation, both in structure and in ligand binding residues, was evident for all four enzymes with the exception of positions F300 in meningococci and S299 in gonococci. Comparison of the structures suggests that when a small residue is present (C307 in humans or S299 in gonococci), both the UDP-GlcNAc and UDP-Glc are able to bind productively and hence the enzyme is bi-functional (Fig. 3C,E). In contrast, in the presence of an aromatic side chain (Y299 in *E. coli* or F300 in meningococci), the UDP-GlcNAc does not bind correctly in the active site to allow efficient hydride transfer from the sugar moiety to the nicotinamide and, thus the enzyme is mono-functional for only UDP-Glc (Fig. 3B,D).

To test the role of S299 and F300 in determining substrate specificity of neisserial GalE enzymes, GalE\_17(S299F)::Hisx6 and GalE\_2(F300S)::Hisx6 mutants were constructed using site-directed mutagenesis and these recombinant enzymes were purified. The GalE\_17(S299F)::Hisx6 lost the ability to epimerise UDP-GlcNAc while retaining the ability to convert UDP-Glc to UDP-Gal, although the equilibrium was altered to favour the formation of UDP-Glc compared to the wild-type enzyme (from 39% to 59%). Meningococcal GalE\_2(F300S)::Hisx6 gained the ability to epimerise both UDP-Glc and UDP-GlcNAc although the observed equilibrium resulted in less UDP-GalNAc for GalE\_2(F300S)::Hisx6 than gonococcal GalE\_17::Hisx6 (Table 1).

**The distribution of mono- and bi-functional GalE1 (NEIS0048) is related to species.** A Neighbor-Joining phylogenetic tree of 89 GalE1 amino acid sequences (Supplementary Table S1, Supplementary Fig. 2) revealed at least three major lineages Clades 1, 2, and 3, in addition to two sets of deeply branched outliers. One group of outliers contained alleles from *N. weaveri*, *N. animalis* with serogroup E and serogroup Z expressing meningococci. The GalE1 alleles from the serogroup E and serogroup Z expressing meningococci both have a cysteine in the active site cleft consistent with the proposed bi-functionality of GalE1



**Figure 3.** Structural views of mono-functional and bi-functional GalE from different organisms. (A) The predicted structure of GalE from *N. meningitidis* MC58 obtained using PHYRE2. The structure is depicted as a ribbon representation showing the secondary structure elements. The side chains for residues which are not conserved amongst all neisseria species are shown in stick representation. Position 300 in *N. gonorrhoeae* strain FA1090 is labelled. (B) The active site region of GalE from *E. coli* in complex with NADH and UDP-GlcNAc (PDB code 1LRJ)<sup>17</sup>. The protein alpha carbon trace is shown. The bound ligands and side chains for residues

involved in interaction with the bound ligands are indicated by stick representation. A mesh surface is shown for Tyr299. (C) The active site region of GalE from *H. sapiens* in complex with NADH and UDP-GlcNAc (PDB code 1HZJ)<sup>17</sup>. The bound ligands and side chains for residues involved in interaction with the bound ligands are indicated by stick representation. A mesh surface is shown for Cys307. (D) The active site region of the PHYRE2 predicted structure of GalE from *N. meningitidis* MC58. The conserved side chains for residues implicated in interaction with the substrates are indicated by stick representation. A mesh surface is shown for Phe300. (E) The active site region of the PHYRE2 predicted structure of GalE from *N. gonorrhoeae* FA1090. The conserved side chains for residues implicated in interaction with the substrates are indicated by stick representation. A mesh surface is shown for Ser299.

and the requirement for UDP-GalNAc for capsule synthesis. The GalE from *N. weaveri* also contained a cysteine in the active site cleft and was predicted to be bi-functional. The GalE proteins from *N. animalis* and *N. subflava* were both predicted to be mono-functional due to the presence of a tyrosine or phenylalanine residue, respectively, in the active site cleft.

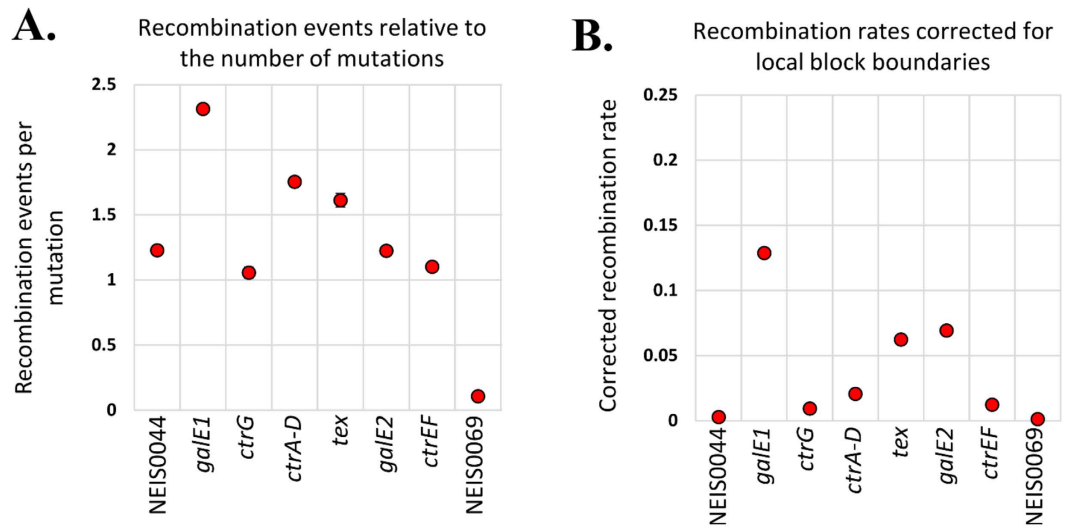
Clade 1 containing GalE amino acid alleles from *N. elongata* and *N. mucosa*, was predicted to be bi-functional due to the presence of either a valine or a cysteine residue in the active site cleft. Clade 2 contained alleles from *N. cinerea*, *N. polysacchara*, *N. lactamica*, *N. gonorrhoeae*, *N. bergeri*, and five meningococci. All of these were assigned bi-functionality due to an active site serine residue characteristic of the bi-functional alleles of *N. gonorrhoeae*. GalE alleles 16, 24, 37, 129 and 166 were present in meningococcal *cnl* isolates. Clade 3 contained only GalE1 amino acid sequences from *N. meningitidis* which were either mono- (contained a phenylalanine in the active site) or bi-functional (contained a serine or cysteine in the active site). Two distantly related protein alleles from *N. meningitidis*, GalE\_41 and GalE\_34, from a serogroup A and a serogroup C isolate, respectively, were most similar to protein alleles from *N. subflava*. The bi-functional alleles in clade 3 were associated with cc174 isolates expressing serogroup Y capsules (allele 131), a variety of serogroup B isolates from various genetic lineages (alleles 46, 49 and 236), a cc1 isolate expressing serogroup A capsule (allele 50), and serogroup I and K expressing isolates (allele 4 and 5 respectively). GalE\_236 was selected as a representative of the bi-functional alleles in clade 3 for assessment of epimerase functionality. It was cloned, expressed as a His-tagged protein, assessed by HPLC and was shown to be bi-functional (Table 1).

### Region D and Region D' of encapsulated meningococci generally encode mono- and bi-functional GalE, respectively.

In addition to Region D containing the *gale1* allele, encapsulated meningococci also possess the inverted repeat Region D' which contains a truncated remnant *gale2* in an operon with *rfbBAC'* (Fig. 1). The *gale2* pseudogene has lost the first 402 nucleotides but retains the last 615 nucleotides, containing the active site motif which can be used to assign ancestral functionality. To unambiguously assign a function to *gale2* alleles, *gale1* and *gale2* genes were manually curated in 1196 meningococci representing 10 common clonal complexes. This yielded 65 *gale1* alleles and 24 *gale2* alleles (data not shown). Of the 65 *gale1* alleles in 1196 isolates, 75% were associated with mono-functionality (F300) with the remainder being bi-functional (S300). The GalE1 alleles from serogroups B, W and Y were predominantly mono-functional (Fisher's exact test, two-tailed,  $p < 0.05$ ). However, no significant association of GalE1 functionality was noted for serogroups A and C although the trend in this small dataset was towards mono-functionality (Supplementary Table S2). Bi-functional GalE1 alleles were associated with all serogroup E expressing isolates (Fisher's exact test, two-tailed,  $p < 0.0001$ ) due to the requirement for UDP-GalNAc for the synthesis of this capsule polymer. Of the 24 *gale2* alleles, 71% were associated with bi-functionality with the remainder being mono-functional. Overall, 92.4% of meningococcal strains carried a mono-functional *gale1* and a bi-functional remnant *gale2* allele. There is a strong retention (Fisher's exact test, two-tailed,  $p < 0.05$  for each serogroup) of separate ancestral functionality between *gale1* and *gale2* alleles in meningococcal isolates that expressed capsules other than serogroup E.

**Recombination rates are highest in *gale1*, *gale2* and *tex* in the *cps* island.** Since meningococcal *gale1* and *gale2* are highly conserved at the nucleotide level, homologous recombination between the two genes may randomly switch functionality between *gale1* and *gale2* loci. To examine this further, recombination events per mutation ( $r/m$ ) was used to calculate the number of recombination events relative to mutations ( $\rho/\theta$ ) and the recombination events relative to the number of mutations within *gale1* and *gale2* with respect to other *cps* island regions (Supplementary Table S3, Fig. 4). Recombination events occurred twice as often as point mutations in *gale1*, 1.5 times as often in *ctrA-D* and *tex* while recombination and mutation equally contributed to diversity within *ctrG*, *gale2* and *ctrEF*. Diversity within NEIS0044, located on the boundary of Region D containing *gale1*, was equally driven by recombination and mutation, whereas diversity of NEIS0069 on the boundary of Region B was mostly caused by mutation (Fig. 4A). When recombination rates were corrected for local block boundaries, *gale1*, *tex* and *gale2* had the highest recombination rates relative to the other genes in the *cps* locus (Fig. 4B). Therefore, although recombination does contribute to the diversity of *gale1* and *gale2* alleles, the exchange of functionality between the two loci is infrequent since the functional separation of the two loci is maintained in the clonal complexes that do not possess serogroup E.

**Meningococcal *gale2* represents a potential site for the capture of the *cps* island.** The origin of Region D' containing *gale2* is currently unknown. While the majority of encapsulated meningococci contained *gale2* with a motif consistent with an ancestral bi-functional activity, most *cnl* meningococci from various genetic lineages including cc53 and cc198 possessed bi-functional alleles (GalE\_37, GalE\_16 and GalE\_116, GalE\_161, GalE\_281, GalE\_362 alleles, GalE\_378, and GalE\_397). An alignment of the conserved portions of the *gale1* and



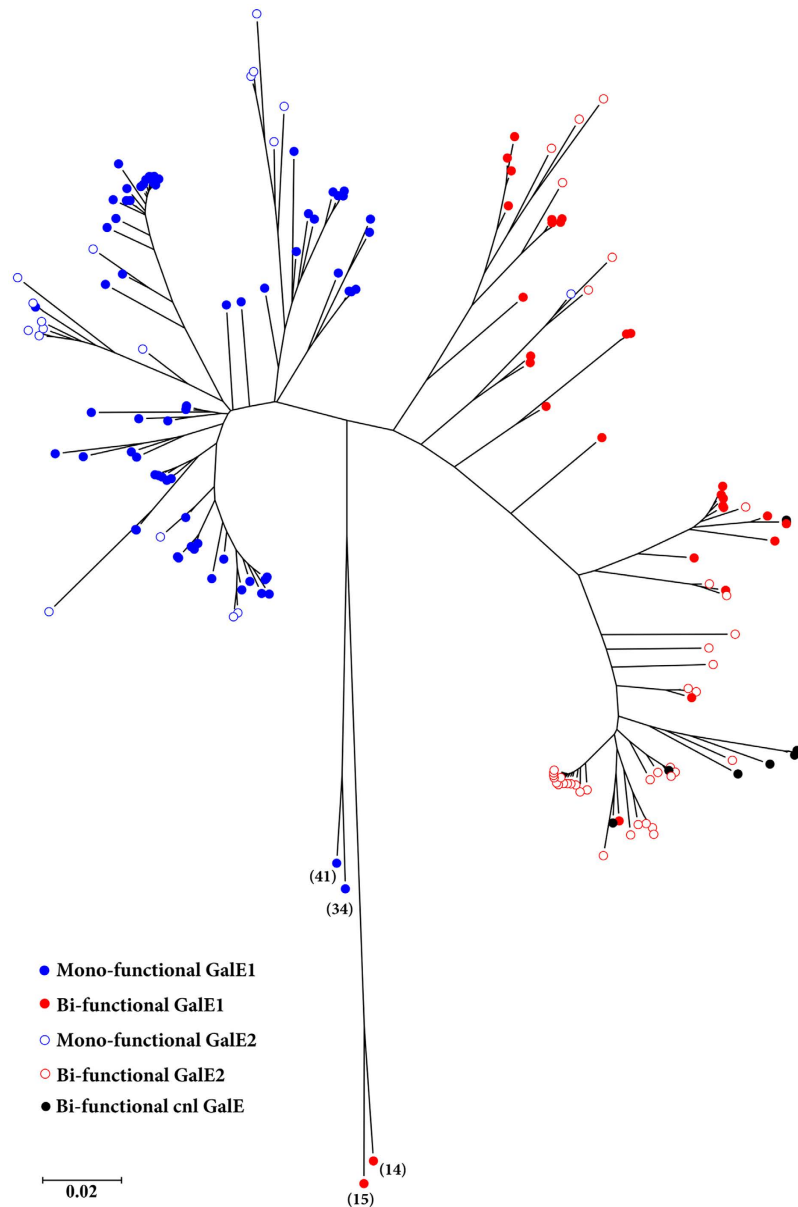
**Figure 4. Recombination and mutation rates across the *cps* locus genes.** Parsed ClonalFrame output detailing the origin of the rates of change for each locus. Panel A shows the relative rate of occurrence of recombination and mutation for each locus such that a value greater than 1.0 on the vertical axis ( $\rho/\theta$ ) indicates a higher frequency of recombination than mutation. Panel B shows the intensity of recombination for each locus as inferred by ClonalOrigin such that high values on the vertical axis ( $\rho_s/\theta_s$ ) indicate hotspots for recombination.

*galE2* alleles revealed a bi-furcated phylogeny with the bi-functional *galE1/galE2* alleles and mono-functional alleles forming separate clusters (Fig. 5). Interestingly, the *galE* alleles from *cnl* meningococci clustered with the bi-functional *galE2* alleles. This observation generated the hypothesis that the *galE2* of Region D' in pathogenic meningococci is a remnant of *galE* from the original *cnl* locus of the ancestral recipient genome during the capture of the *cps* island. The *cnl* site is always located between two boundaries: NEIS0044 upstream of the *galE1-rfbBAC* cluster and NEIS0069 downstream of *tex* (Fig. 1). The capture of the *cps* island was postulated to lead to the deletion of the 5' end of *galE-tex*- NEIS1357 in the *cnl* locus to create NEIS0044-*rfbCAB'*-*galE2* linked to the *cps* island ending at the downstream boundary of NEIS0069. In this model, the *cps* island HGT event would require the introduction of Region A-C-E-D-B by recombination into the *galE* of the *cnl* locus.

One outcome of this model would be the expectation of strong linkage disequilibrium between the two *galE* loci with a strong association with serogroup as these would be linked en bloc during the HGT event. A plot of the association of *galE1* and *galE2* allelic pairs with serogroup in 1196 isolates representing each of the 10 clonal complexes showed that each clonal complex was characterized by a predominant *galE1/serogroup/galE2* combination (Supplementary Fig. S3). In summary, 84.2% of cc1 is represented by GalE1<sub>39</sub>/GalE2<sub>30</sub>, 74.2% of cc11 is represented by GalE1<sub>10</sub>/GalE2<sub>8</sub>, 46.6% of cc22 is represented by GalE1<sub>8</sub>/GalE2<sub>13</sub>, 88.8% of cc23 is represented by GalE1<sub>27</sub>/GalE2<sub>23</sub>, 73.2% of cc41/44 is represented by GalE1<sub>20</sub>/GalE2<sub>13</sub>, 65.3% of cc60 is represented by GalE1<sub>10</sub>/GalE2<sub>13</sub>, and 78.4% of cc269 is represented by GalE1<sub>120</sub>/GalE2<sub>7</sub>. A strong associative property with clonal complex was observed for *galE1/galE2* allelic pairs (Cramer's V coefficient,  $V = 0.985$ ), *tex* ( $V = 0.963$ ), *ctrE* ( $V = 0.948$ ) and *ctrF* ( $V = 0.770$ ) suggesting that the correlation of alleles in the *cps* island with a specific clonal complex was formed independently during the acquisition of the *cps* island into the founder of each clonal complex.

This hypothetical model for the acquisition of the *cps* island in modern clonal complexes also predicted that the original organization of the *cps* island upon acquisition could have been NEIS0044-Region D' (containing *galE2* bi-functional)-E-C-A-D (*galE1* mono-functional)-B. However, the most common arrangement found in closed genomes of strains MC58 and FAM18 has NEIS0044 followed by Region D containing *galE*- A- C- E- D'- B- NEIS0069 (Fig. 6) which suggests that the synthesis cassette, A-C-E, may undergo inversion in this site. Colony PCR confirmed that both *cps* island arrangements were detectable in plate grown cultures of strains MC58, FAM18 and NMB (Supplementary Fig. S4). Complete sequencing of these PCR products did not detect recombination events within the *galE1* and *galE2* genes which is consistent with potential breakpoints in the flanking *rfbBAC* gene clusters of Region D and Region D'. However, since the paired *rfbBAC* and *rfbBAC'* regions in each strain was almost identical, the recombination site was not detected.

**Meningococcal *galE1* and *tex* represent potential sites for homologous recombination during serogroup switching events.** Mustapha *et al.*<sup>12</sup> have recently described the recombination events detected upon the integration of the serogroup W cassette into the cc11 lineage which originally expressed serogroup C (Fig. 1). They showed that the left hand boundary for the integration of Region A for serogroup W synthesis occurred in *galE1* while *galE2* remained unchanged. To examine whether this is a common occurrence in serogroup switching, the association of the *galE1/galE2* allelic pairs with serogroup was plotted (Supplementary Fig. S3). The dominating serogroup in each lineage is strongly associated with only one *galE1/galE2* allelic pair. For example, 93.5% of cc11:serogroup C isolates possessed GalE1<sub>27</sub>/GalE2<sub>8</sub> allelic pair whereas 100% of cc11:serogroup W isolates possessed GalE1<sub>10</sub>/GalE2<sub>8</sub>. Capsule switching events which



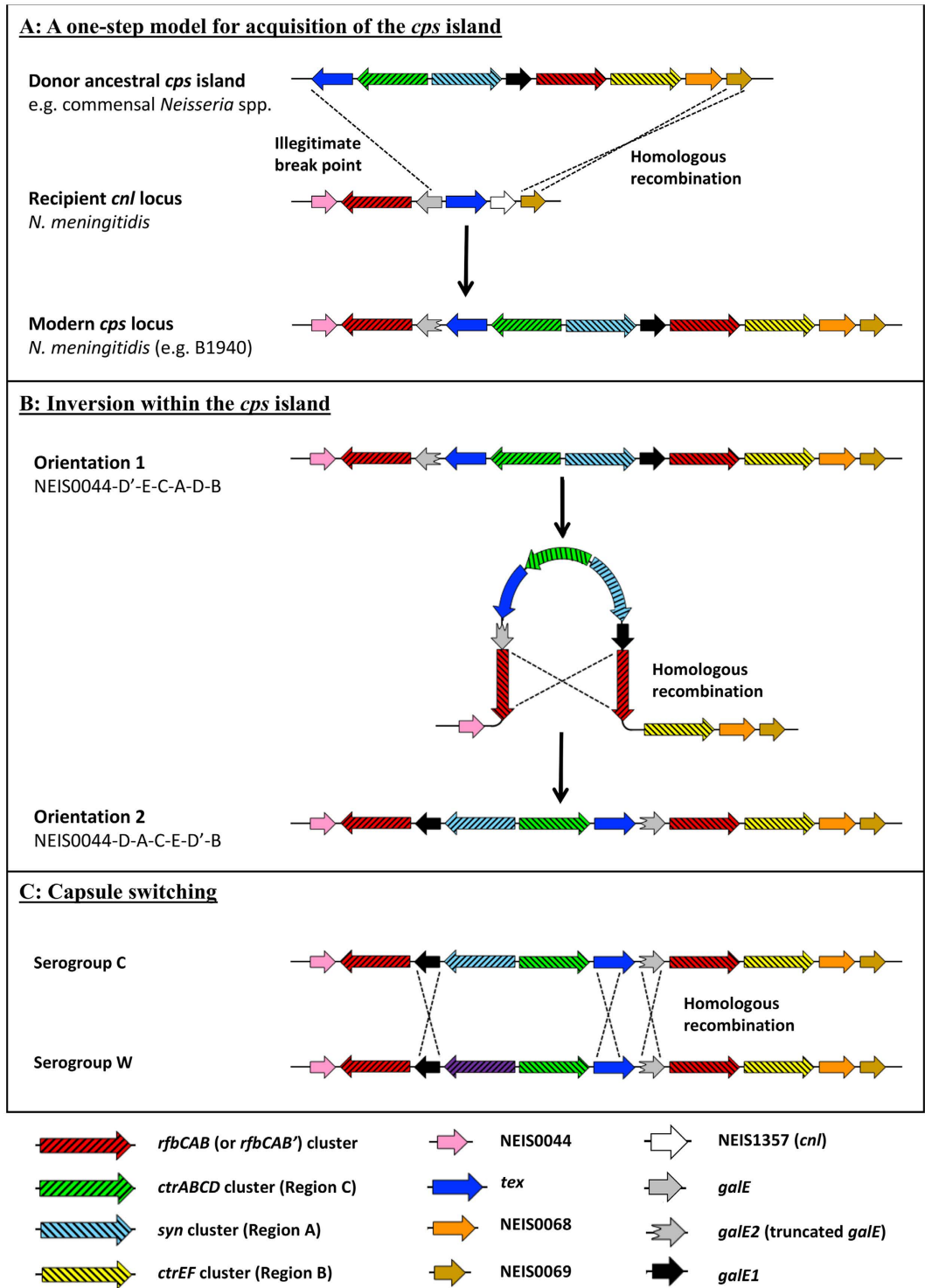
**Figure 5. Phylogenetic tree of *gale1* and *gale2* alleles from *N. meningitidis*.** The phylogenetic tree was constructed using the Neighbor-Joining method with 1000 bootstraps and evolutionary distances were calculated by the maximum composite likelihood method. In order to generate the tree, the alignment of the nucleotides was performed using MUSCLE after trimming the *gale1* and *gale2* alleles to the last 615 nucleotides. The colours indicate the functionality of the protein associated with the alleles; the filled and empty circles indicates *gale1* and *gale2* alleles respectively. The asterisks indicate *gale1* alleles (given in brackets) from *cnl* isolates.

involved the acquisition of serogroup B into cc11 displayed a similar pattern with 75% retaining GalE2\_8 while the allele of GalE1 with Region A had been changed. Similarly, 16 events in cc60:serogroup E strains involving the acquisition of serogroup B synthetic regions showed a similar pattern of variation in GalE1 with conservation of the GalE2 alleles (Supplementary Table S4). Although the library of 1196 isolates contained 65 unique GalE1/GalE2 allelic pairs, only four GalE1/GalE2 pairs were found in multiple clonal complexes (Supplementary Fig. 3) suggesting that these exchanges had recombination boundaries outside of the *gale1/gale2* pairs.

## Discussion

The acquisition of the *cps* island by HGT into an ancestral non-pathogenic meningococcal lineage carrying a *cnl* locus similar to cc53 and cc198 isolates of *N. meningitidis* has been proposed by Schoen *et al.*<sup>6</sup> and others<sup>21,22–25</sup>. This hypothetical model proposes that the modern arrangement of the *cps* island is the outcome of at least two recombination events that resulted in the capture of Region A-C and Region B, located upstream and downstream of Region E (*tex*), respectively, in the capsule null locus of non-pathogenic meningococci<sup>8–10,26</sup>. In this study, we





**Figure 6. Proposed hypothetical model of events in the capture and re-organisation of the *cps* island in meningococci.** (A) A hypothetical E-C-A-D-B cassette, containing a mono-functional *galE1*, was formed in an ancestral isolate (donor). This was recombined into a recipient strain containing a *cnl* locus to create the arrangement of the *cps* island found in modern pathogenic *N. meningitidis* in which the A-C-E cassette is flanked by the repetitive regions D and D'. (B) The D'-E-C-A-D cassette may undergo inversion via homologous recombination since Region D and D' are able to form a hair-pin loop. (C) Serogroup switching involves the replacement of one serogroup synthetic cassette (eg. Serogroup C) with another (eg. Serogroup W). Multiple recombination events have been shown to occur within *galE1*, *tex* and *galE2* in multiple studies (see text).

provide evidence for a second hypothetical model which suggests, that once the ancestral *cps* island was formed, this structure became the donor sequence for en bloc transfer into the modern clonal complexes of *N. meningitidis*. Evidence for this hypothetical model is based upon the observation that the *gale1* and *gale2* alleles are phylogenetically distinct due to different enzymatic functions, and that these phylogenetic relationships can be traced through capsule null loci and in *cps* islands in the various clonal complexes of *N. meningitidis*.

Gonococci have a requirement for UDP-GalNAc which is a component of the gonococcal LOS<sup>27</sup>. The gonococcal GalE epimerase was shown to be bi-functional with the capacity to synthesise both UDP-Gal and UDP-GalNAc whereas the meningococcal GalE could only produce UDP-Gal. Exchanging a pivotal amino acid in the active site cleft (F300 in meningococci and S299 in gonococci) enabled the exchange of mono- and bi-functionality of the enzymes suggesting that the identity of the amino acid positioned at the mouth of the active site cleft determined the functional phenotype of the epimerase. Ishiyama *et al.*<sup>28</sup> proposed that changing the amino acid at this position of the active site cleft of GalE epimerase determined the ability of the UDP-linked *N*-acetyl-sugar to rotate and undergo epimerisation. Although there are other non-conserved amino acid positions within neisserial GalE variants, none of these were shown to influence the cleft leading to the active site (Fig. 3). Therefore, the use of the six amino acid residue motif surrounding the active site cleft residue of F300 (meningococci)/S299 (gonococci) or equivalent in commensal *Neisseria* spp. was considered sufficient to assign functionality to the *gale* alleles.

Functional and phylogenetic analysis of *gale1* revealed that all *Neisseria* spp. possessed the gene, but that the functionality of the epimerase was predicted to vary across species (Fig. 2). Although the role of UDP-Gal and UDP-GalNAc is well described for the pathogenic *Neisseria* spp., very little is known regarding other species. Immunotyping studies have proposed that *N. cinerea*, *N. polysacchara* and *N. lactamica*, which possess putative bi-functional *gale1* alleles, synthesise a LOS containing Gal and GalNAc residues but no structural studies have been done to confirm this<sup>29–31</sup>. The bi-functional meningococcal *gale1* alleles (Supplementary Fig. S2), were found in the presence of serogroup A, B, E, Y, I and K synthetic cassettes. A representative of this group, GalE\_236 allele, had an equilibrium similar to that of the bi-functional allele of *N. gonorrhoeae*, thereby suggesting that the synthesis of UDP-GalNAc is the favoured outcome for this reaction (Table 1). This allele was found in *N. meningitidis* strain NMB which was shown to synthesise UDP-GalNAc by Lee *et al.*<sup>32</sup>. Although UDP-GalNAc is not present in LOS of this isolate<sup>33</sup>, it may be involved in another synthetic pathway such as protein glycosylation which has been recently detected in commensal *Neisseria* spp.<sup>34,35</sup>. However, since the GalNAc lipopolysaccharide transferase, *lgtD*, has been found in some isolates of meningococci, the capacity to add GalNAc to LOS could occur if these strains also possessed a bi-functional *gale1*<sup>36</sup>. The phylogeny of the meningococcal *gale* alleles (Fig. 5) suggests that bi-functional *gale1* alleles found in serogroups A/B/C/Y/W may have arisen sporadically through recombination with the *gale2* remnant or through point mutation of the active site of the *gale1* allele (Fig. 4). In contrast, the bi-functional *gale1*\_14 allele found in serogroup E- expressing isolates is phylogenetically distant from these bi-functional *gale1* alleles (Fig. 5) suggesting a separate heritage potentially with the closest relative, *N. weaveri* (Fig. 1).

The phylogenetic relationship of the bi-functional *gale2* alleles from pathogenic meningococci and the *gale* alleles in the *cnl* of non-pathogenic isolates led to the hypothesis that Region D' was the original locus into which the *cps* island recombined (Fig. 6A). Since statistical association analysis demonstrated linkage disequilibrium between *gale1* and *gale2* as well as *gale1/gale2* pairs within each clonal complex, this suggested that the pairing of *gale1* with *gale2* occurred upon the acquisition of the *cps* island into the *cnl* locus of the ancestral strain of each clonal complex. One event that could result in this arrangement is the illegitimate recombination of an ancestral *cps* island consisting of Region A-C-E-D-B into *gale* on the left boundary and NEIS0069 on the right boundary of the recipient *cnl* locus (Fig. 6A). The strong associative properties of the *gale1/gale2* (Region A), *tex* (Region E) and *ctrEF* (Region B) alleles with clonal complex supports the concept of en bloc transfer and provides an explanation for the creation of Region D' due to an illegitimate recombination event that truncates the *gale* allele of the *cnl* locus to become the remnant *gale2* locus (Fig. 6A).

The initial organisation of the *cps* locus predicted by this model is NEIS0044-region D'(*gale2*)- E- C- A- (*gale1*) D- B- NEIS0069 which is the same as that of the closed genome of strain B1940<sup>23</sup>. However, the closed genomes of strains MC58 and FAM18 are reported as NEIS0044-region D (*gale1*)- A- C- E- (*gale2*) D'- B- NEIS0069, suggesting that an inversion of the A-C-E block between region D and region D' could occur. This was confirmed by directional PCR of the *cps* locus using plate grown strains MC58 and FAM18 (Supplementary Fig. S4). Since single colony PCR revealed that each colony contained a mixture of *cps* orientations (data not shown), it appears that this process is dynamic, occurring during chromosomal replication when single stranded DNA is present for RecA mediated homologous recombination (Fig. 6B). Unless the duplicated Region D' is deleted, the locus is not able to be fixed in any particular orientation. Colony immunoblots of strain NMB which is not phase variable for capsule expression did not detect any mixed phenotypes consistent with previous studies on the expression of capsule by this strain (data not shown<sup>37–39</sup>). It is interesting to note that the *cps* island is inserted 54 kb downstream of the origin of replication which is surrounded by a number of genomic inversions<sup>40</sup>. In *E. coli*, Ivanova *et al.*<sup>41</sup> have recently shown that inversions surrounding the origin of replication occur to resolve collisions between highly transcribed genes and the direction of chromosomal replication, thus improving fitness. It is currently unclear whether this mechanism applies to this situation in *N. meningitidis* but is worth considering in future investigations of this region.

The proposed en bloc model for the transfer of the *cps* island into meningococcal clonal complexes relies upon the concept previously advanced by Schoen *et al.*<sup>6,7,42</sup> that the mosaic structure is created in an ancestral meningococcal isolate. Recently, Harrison *et al.* (personal communication) have detected similar *cps* islands consisting of regions A-C-E-D-B, and lacking Region D', in commensal *Neisseria* spp. which could also be a source of the donor element for the en bloc transfer hypothesis. Currently the PubMLST database does not contain meningococcal isolates lacking Region D' as the majority of isolates represented in this collection have been isolated from cases

of disease. It will be interesting to examine a larger collection of non-disease causing meningococcal isolates for the presence of strains bearing incomplete *cps* islands as these may provide further evidence for the model. This model was also used to examine serogroup switching, which from the observations of Mustapha *et al.*<sup>12</sup> disrupt the linkage disequilibrium of *galE1* and *galE2*. An examination of other serogroup switching events in the database indicated that this was a general phenomenon associated with this HGT event.

In conclusion, we propose evidence for a hypothetical model in which the acquisition of the *cps* island via HGT results in the loss of UDP-GalNAc synthesis from UDP-GlcNAc in most pathogenic meningococcal isolates, except for those expressing serogroup E capsules. UDP-GlcNAc is an entry metabolite for the synthesis of cell wall components, peptidoglycan and lipid A, in addition to sialic acid found in serogroup B, C, W and Y capsular types. Presumptively, the inactivation of the bi-functional *galE* allele of the *cnl* locus and replacement with a mono-functional *galE1* allele releases a pool of UDP-GlcNAc that can be redirected into the sialic acid capsule biosynthesis pathway while minimising the metabolic fitness cost associated with the acquisition of the *cps* island in pathogenic meningococci. However, other accessory changes to central metabolism are likely to be required. Mustapha *et al.*<sup>12</sup> noted that the capsule switching event in cc11 from serogroup C to W included central metabolic genes such as the pyruvate kinase. In addition, Schoen *et al.*<sup>43</sup> found that the central metabolism of pathogenic meningococci is diverse and that multiple adaptive changes to metabolism have occurred thus providing a framework in which to test our theory regarding the fitness costs associated with the acquisition of the *cps* synthesis pathway.

## Materials and Methods

**Bacterial strains and growth conditions.** Meningococcal strains were cultured under aerobic conditions with 5% CO<sub>2</sub> at 37 °C on GC agar (GCA) or GC broth (GCB) (Oxoid) supplemented with 0.4% glucose, 0.01% glutamine, 0.2 mg of cocarboxylase per litre, and 5 mg of Fe(NO<sub>3</sub>)<sub>3</sub> per litre. The wild-type strains and constructed mutants used in this study are shown in Supplementary Table S5. Antibiotic selection for meningococcal mutants was performed on GCA containing 100 µg/ml of kanamycin (sulfate salt), 60 µg/ml of spectinomycin, 5 µg/ml of tetracycline or 2 µg/ml of erythromycin (Sigma). *Escherichia coli* DH5α was used as a host for all DNA manipulations and Rosetta™ (DE3) as a host for all protein expression. *E. coli* strains were routinely grown on Luria-Bertani broth (LBB) and agar (LBA, Oxoid) which, where appropriate, was supplemented with antibiotics at the following concentrations: ampicillin at 100 µg/ml, spectinomycin at 50 µg/ml, kanamycin at 50 µg/ml, erythromycin at 300 µg/ml, tetracycline at 12.5 µg/ml and chloramphenicol at 30 µg/ml (Sigma).

**Construction of expression strains.** The plasmids constructed and used in this study are listed in Supplementary Table S5. Genomic DNA was isolated from *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strains NMB and MC58 using the Purelink™ genomic DNA purification kit (Invitrogen) according to manufacturer's instructions. The gene *galE* was amplified with the primers KAP580 (5'-GGAATCCATATGAAAAAATTCCTCGTTACCG-3') and KAP581 (5'-CGCGGATCCTTAATCGTCGTAGCCATTCGG-3') for MC58, KAP582 (5'-GGAATCCATATGACCGTCTGATTACCG-3') and KAP583 (5'-CGCGGATCCTTAATCCCCATATCTGCCG-3') for FA1090 and KAP563 (5'-GGAATCCATATGCCCTATACGGAAGATATG-3') and KAP564 (5'-CGCGGATCCTTAATCCCCATATCCGTTGGG-3') for NMB, such that there was a 5' *NdeI* site overlapping the start codon and a 3' *BamHI* site downstream of the stop codon incorporated into the PCR fragment. Each PCR product was directionally cloned into the *NdeI* and *BamHI* sites of pET15b, resulting in the 5' fusion of the open reading frame with a hexahistidine motif, resulting in plasmids pCMK730, pCMK729 and pCMK771. The *galE* of *N. gonorrhoeae* and *N. meningitidis* that have been cloned into pET15b were sequenced and their sequence found to be identical to the native gene.

The cloned *galE* of *N. gonorrhoeae* strain FA1090 and *N. meningitidis* strain MC58 were mutagenised using site directed mutagenesis to change a single nucleotide. The primers KAP427 (5'-CGACTTGGCGTGTCTTCTATGCCGACCC-3') and KAP428 (5'-GGGTCGGCATAGAAA CACGCCAAGTTCG-3') to mutate the gonococcal *galE*, and KAP429 (5'-GGTGATTTGGCGTGTCTTCTATGCCGACCC-3') and KAP430 (5'-GGGTCGGCATAGGAGCACGCCAAATCACC-3') to mutate the meningococcal *galE* were used to amplify the plasmids pCMK729 and pCMK730 respectively using the non-strand displacing polymerase Phusion (New England Biolabs). The template DNA was digested with *DpnI*, prior to transformation into DH5α. The resulting plasmids were sequenced and confirmed to contain only the introduced mutation. These plasmids were called pCMK733 and pCMK734 respectively.

Expression strains were constructed by transforming the plasmids pCMK729, pCMK730, pCMK771, pCMK733 and pCMK734 into BL21-DE3 Rosetta. Inducible protein expression from these strains were confirmed by growing the strains to mid-log phase in 10 ml LBB, splitting the culture in two, and inducing one culture with 0.3 mM IPTG for 1 hr. The cells from 1 ml of each culture was collected by centrifugation and resuspended in 100 µl of distilled water to which 25 µl of 5x Laemmli loading buffer was added. The samples were boiled for 10 mins prior to separation on 12% SDS-PAGE at 150 V, then stained with Coomassie Blue R250. Expression was determined by the presence of a band of increased intensity present in the induced sample compared to the un-induced sample at the correct molecular weight.

Expression of His-tagged GalE was confirmed by Western immunoblot. Whole cell lysates (750 ng) were separated by 15% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) by standard methods and transferred to nitrocellulose membranes. The membranes were blocked overnight with 2% BSA in TBS, then incubated with the monoclonal mouse anti-His IgG (Sigma) primary antibody at 1:1,000 dilution. Horse radish peroxidase-conjugated anti-Rabbit IgG and anti-Mouse IgG secondary antibodies (Santa Cruz Biotechnology) was used for detection and the membrane was developed with an ECL kit (GE Healthcare).

**Protein purification.** *E. coli* strains expressing meningococcal GalE proteins (both native and mutagenised) were grown overnight with shaking in 10 ml LBB containing 100 µg/ml ampicillin. An aliquot of 100 µl was used to inoculate 10 ml LBB containing 100 µg/ml ampicillin the following day and the strains grown to OD<sub>600</sub> 0.6–0.8 at which time protein expression was induced by the addition of IPTG to a final concentration of 0.3 mM. Induction occurred for 3 hr at 37 °C prior to the purification of protein using the Wizard HisLink™ Spin Protein Purification System (Promega) as per manufacturer's instructions. The gonococcal GalE could not be purified in this way due to the formation of inclusion bodies. Strains expressing the gonococcal GalE proteins were grown overnight with shaking in 10 ml LBB containing 100 µg/ml Ampicillin. An aliquot of 1 ml was used to inoculate 1 L LBB containing 100 µg/ml Ampicillin the following day and the strains grown to OD<sub>600</sub> 0.6–0.8 at which time protein expression was induced by the addition of IPTG to a final concentration of 0.3 mM. Induction occurred for 3 hr at 37 °C prior to the collection of all bacterial cells by centrifugation at 3000 rcf for 15 mins at 4 °C. The pellet was resuspended in 50 ml binding buffer (20 mM sodium phosphate, 0.5 M NaCl and 20 mM imidazole, pH 7.4) and sonicated 1 min on, 1 min off for 40 mins. Cell debris was removed by centrifugation and the supernatant applied to a HisTrap™ FF column. The column was washed with 10 volumes of binding buffer and the protein eluted in 10 volumes of elution buffer (20 mM sodium phosphate, 0.5 M NaCl and 500 mM imidazole, pH 7.4).

The purified proteins were dialysed against 100 volumes of dialysis buffer (10% glycerol, 20 mM Tris/HCl, pH 7.9), changed three times. The gonococcal GalE proteins were concentrated to 100 µl using a centricon-10 (Millipore, Germany). Protein concentrations were determined by Bradford assay and glycerol was added to a final concentration of 50% (v/v) to stabilize the enzyme.

**High Performance Liquid Chromatography.** Activity of the gonococcal and meningococcal GalE proteins, and the site directed mutagenised proteins was done using methods described in Dong *et al.*<sup>44</sup>. Briefly, the activated sugar substrates (UDP-Glc, UDP-Gal, UDP-GlcNAc and UDP-GalNAc, 1 mM) were incubated with 200 ng of purified protein in 20 mM Tris-HCl, 4 mM Mg<sup>2+</sup>, and 1 mM NAD<sup>+</sup> (pH 8.0) in a final volume of 50 µl. The reactions were performed at 37 °C for 2 h and were terminated by heating at 100 °C for 5 min. The products from the reaction were run on HPLC using a UV detector and compared to UDP-Glc, UDP-Gal, UDP-GlcNAc and UDP-GalNAc standards.

**Phylogenetic analysis of *galE1* and *galE2* alleles from *N. meningitidis*.** A dataset of 194 isolates of *Neisseria* spp. was selected from the Bacterial Isolate Genome Sequence Database (BIGSDB) located at PubMLST (<http://pubmlst.org/neisseria>)<sup>45</sup> and described in Supplementary Table S1. This dataset was selected to include the most common alleles of and provide the maximal diversity for NEIS0048 and NEISS0062 across 133 *N. meningitidis* isolates, 16 *N. gonorrhoeae* isolates, 17 *N. lactamica* isolates, seven *N. cinerea* isolates, six *N. polysacchara* isolates, four *N. oralis* isolates<sup>42,46</sup>, four *N. elongata* isolates, four *N. subflava* isolates and one isolate each from the *Neisseria* species *N. bergeri*, *N. animalis* and *N. weaveri*. All of these genomes were sequenced on an Illumina Genome Analyser II platform and assembled according to Jolley and Maiden<sup>45</sup>. However, because Region D and D' are duplicated regions, each contig was manually checked and exported as XMFA files containing aligned sequence blocks and then converted to a fastA format for import into MEGA version 5.0<sup>47</sup> and Splitstree<sup>48</sup>.

The *galE1* and *galE2* alleles were downloaded from the *Neisseria* PubMLST database (<http://pubmlst.org/neisseria/>)<sup>45</sup> as NEIS0048 and NEIS0062 respectively. The phylogeny of the GalE1 peptide sequences was reconstructed using the neighbor-joining method using the MEGA software package<sup>47</sup> with the following parameters: Bootstrap (1000 replicates), using a Poisson model and gamma distribution (parameter set to 1). For the phylogeny of the conserved regions of the *galE1* and *galE2* alleles, the sequences were trimmed using UGENE and the final 615 nucleotides were aligned using MUSCLE. Phylogenetic analysis was conducted by the neighbor-joining method using 1000 bootstrap tests. Evolutionary distances were calculated by the maximum composite likelihood method. The analysis was conducted with MEGA6.06.

**Clonal Frame analysis of recombination relative to mutation.** The distribution of NEIS0048 and NEIS0062 was assessed in 1196 pathogenic meningococcal isolates summarised in Supplementary Table S2. Aligned sequences were analysed using the default parameters of ClonalFrame<sup>49</sup> and the output parsed using the ParseCF script written by Barry Hall. Values calculated by ClonalFrame and ParseCF include the mutation rate ( $\theta$ ), the mutation rate per site ( $\theta$ /number of sites), the corrected recombination rate for local block boundaries ( $\rho$ /number of sites), the estimate of recombination relative to that of mutation ( $r/m$ ) and the estimate of the number of recombination events relative to mutations ( $\rho/\theta$ ).

**PCR amplification of *cps* island inversion.** KAP727 (5'-GAAGAATACCGCGAACTGACGC-3') is a forward primer specific to the 5' end of NEIS0044; KAP730 (5'-AGTTGGAAGAGCGCAAAGCCG-3') is a forward primer specific to the 5' end of *tex* (NEIS0059); KAP735 (5'-TTCAGACGGCAAGAGGGTACG-3') is a reverse primer specific to the 3' end of *ctrE* (NEIS0066). The PCR was performed using Phusion® High-Fidelity DNA Polymerase (New England Biolabs). PCR conditions for the assay were 2 min at 98 °C followed by 35 cycles of 98 °C for 10 seconds, 65 °C for 30 seconds and 72 °C for 10 minutes with a final extension time of 10 minutes at 72 °C.

## References

- Pavletic, A. J., Wolner-Hanssen, P., Paavonen, J., Hawes, S. E. & Eschenbach, D. A. Infertility following pelvic inflammatory disease. *Infect Dis Obstet Gynecol* **7**, 145–152 (1999).
- Cartwright, K. A., Stuart, J. M., Jones, D. M. & Noah, N. D. The Stonehouse survey: nasopharyngeal carriage of meningococci and *Neisseria lactamica*. *Epidemiol Infect* **99**, 591–601 (1987).
- Stephens, D. S. Uncovering the meningococcus: dynamics of carriage and disease. *The Lancet* **353**, 941–942 (1999).
- Stephens, D. S., Greenwood, B. & Brandtzaeg, P. Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *Lancet* **369**, 2196–2210, doi: 10.1016/S0140-6736(07)61016-2 (2007).

5. Bartley, S. N. & Kahler, C. M. In *Pathogenic Neisseria: Genomics, molecular biology, and disease prevention* (eds Davies, J. K. & Kahler, C. M.) (Caister Academic Press, 2014).
6. Schoen, C. *et al.* Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc Natl Acad Sci USA* **105**, 3473–3478, doi: 10.1073/pnas.0800151105 (2008).
7. Claus, H., Maiden, M. C., Maag, R., Frosch, M. & Vogel, U. Many carried meningococci lack the genes required for capsule synthesis and transport. *Microbiology* **148**, 1813–1819 (2002).
8. Lam, T. T., Claus, H., Frosch, M. & Vogel, U. Sequence analysis of serotype-specific synthesis regions II of *Haemophilus influenzae* serotypes c and d: evidence for common ancestry of capsule synthesis in Pasteurellaceae and *Neisseria meningitidis*. *Res Microbiol* **162**, 483–487 (2011).
9. Putonti, C., Nowicki, B., Shaffer, M., Fofanov, Y. & Nowicki, S. Where does *Neisseria* acquire foreign DNA from: an examination of the source of genomic and pathogenic islands and the evolution of the *Neisseria* genus. *BMC evolutionary biology* **13**, 184, doi: 10.1186/1471-2148-13-184 (2013).
10. Schoen, C. *et al.* Whole-genome sequence of the transformable *Neisseria meningitidis* serogroup A strain WUE2594. *J Bacteriol* **193**, 2064–2065, doi: 10.1128/JB.00084-11 (2011).
11. Swartley, J. S. *et al.* Capsule switching of *Neisseria meningitidis*. *Proc Natl Acad Sci USA* **94**, 271–276 (1997).
12. Mustapha, M. M. *et al.* Genomic Investigation Reveals Highly Conserved, Mosaic, Recombination Events Associated with Capsular Switching among Invasive *Neisseria meningitidis* Serogroup W Sequence Type (ST)-11 Strains. *Genome biology and evolution* **8**, 2065–2075, doi: 10.1093/gbe/evw122 (2016).
13. Edwards, J. L. & Apicella, M. A. The molecular mechanisms used by *Neisseria gonorrhoeae* to initiate infection differ between men and women. *Clinical microbiology reviews* **17**, 965–981, table of contents, doi: 10.1128/CMR.17.4.965-981.2004 (2004).
14. Zhu, P. *et al.* Genetic diversity of three *lgt* loci for biosynthesis of lipooligosaccharide (LOS) in *Neisseria* species. *Microbiology* **148**, 1833–1844 (2002).
15. Bhattacharjee, A. K., Jennings, H. J. & Kenny, C. P. Structural elucidation of the 3-deoxy-D-manno-octulosonic acid containing meningococcal 29-e capsular polysaccharide antigen using carbon-13 nuclear magnetic resonance. *Biochemistry* **17**, 645–651 (1978).
16. Jennings, H. J., Rosell, K.-G. & Kenny, C. P. Structural elucidation of the capsular polysaccharide antigen of *Neisseria meningitidis* serogroup Z using <sup>13</sup>C nuclear magnetic resonance. *Canadian Journal of Chemistry* **57**, 2902–2907 (1979).
17. Thoden, J. B., Henderson, J. M., Fridovich-Keil, J. L. & Holden, H. M. Structural analysis of the Y299C mutant of *Escherichia coli* UDP-galactose 4-epimerase. Teaching an old dog new tricks. *J Biol Chem* **277**, 27528–27534 (2002).
18. Bengoechea, J. A. *et al.* Functional characterization of Gne (UDP-N-acetylglucosamine-4-epimerase), Wzz (chain length determinant), and Wzy (O-antigen polymerase) of *Yersinia enterocolitica* serotype O:8. *J Bacteriol* **184**, 4277–4287 (2002).
19. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597–600, doi: 10.1093/nar/gkt376 (2013).
20. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845–858, doi: 10.1038/nprot.2015.053 (2015).
21. Xu, Z. *et al.* Phylogenetic study of clonal complex (CC)198 capsule null locus (*cnl*) genomes: A distinctive group within the species *Neisseria meningitidis*. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases* **34**, 372–377, doi: 10.1016/j.meegid.2015.07.013 (2015).
22. Bennett, J. S. *et al.* A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology*, doi: 10.1099/mic.0.056077-0 (2012).
23. Harrison, O. B., Bray, J. E., Maiden, M. C. & Caugant, D. A. Genomic Analysis of the Evolution and Global Spread of Hyper-invasive Meningococcal Lineage 5. *EBioMedicine* **2**, 234–243, doi: 10.1016/j.ebiom.2015.01.004 (2015).
24. Dunning Hotopp, J. C. *et al.* Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes. *Microbiology* **152**, 3733–3749, doi: 10.1099/mic.0.29261-0 (2006).
25. Bennett, J. S. *et al.* Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020-06. *BMC genomics* **11**, 652, doi: 10.1186/1471-2164-11-652 (2010).
26. Frosch, M. & Vogel, U. In *Handbook of meningococcal disease* (eds Frosch, M. & Maiden, M. C.) Ch. 8, 145–162 (WILEY-VCH Verlag GmbH & Co. KGaA, 2006).
27. Kahler, C. M. & Stephens, D. S. Genetic basis for biosynthesis, structure, and function of meningococcal lipooligosaccharide (endotoxin). *Critical reviews in microbiology* **24**, 281–334 (1998).
28. Ishiyama, N., Creuzenet, C., Lam, J. S. & Berghuis, A. M. Crystal structure of WbpP, a genuine UDP-N-acetylglucosamine 4-epimerase from *Pseudomonas aeruginosa*: substrate specificity in UDP-hexose 4-epimerases. *J Biol Chem* **279**, 22635–22642 (2004).
29. Kim, J. J., Mandrell, R. E. & Griffiss, J. M. *Neisseria lactamica* and *Neisseria meningitidis* share lipooligosaccharide epitopes but lack common capsular and class 1, 2, and 3 protein epitopes. *Infect Immun* **57**, 602–608 (1989).
30. Stein, D. C., Miller, C. J., Bhoopalan, S. V. & Sommer, D. D. Sequence-based predictions of lipooligosaccharide diversity in the *Neisseriaceae* and their implication in pathogenicity. *PLoS One* **6**, e18923, doi: 10.1371/journal.pone.0018923 (2011).
31. Arking, D., Tong, Y. & Stein, D. C. Analysis of lipooligosaccharide biosynthesis in the *Neisseriaceae*. *J Bacteriol* **183**, 934–941, doi: 10.1128/JB.183.3.934-941.2001 (2001).
32. Lee, F. K. *et al.* Microheterogeneity of *Neisseria* lipooligosaccharide: analysis of a UDP-glucose 4-epimerase mutant of *Neisseria meningitidis* NMB. *Infect Immun* **63**, 2508–2515 (1995).
33. Rahman, M. M., Stephens, D. S., Kahler, C. M., Glushka, J. & Carlson, R. W. The lipooligosaccharide (LOS) of *Neisseria meningitidis* serogroup B strain NMB contains L2, L3, and novel oligosaccharides, and lacks the lipid-A 4'-phosphate substituent. *Carbohydrate research* **307**, 311–324 (1998).
34. Borud, B. *et al.* Genetic, structural, and antigenic analyses of glycan diversity in the O-linked protein glycosylation systems of human *Neisseria* species. *J Bacteriol* **192**, 2816–2829, doi: 10.1128/JB.00101-10 (2010).
35. Anonsen, J. H. *et al.* Characterization of a unique tetrasaccharide and distinct glycoproteome in the O-linked protein glycosylation system of *Neisseria elongata* subspecies *glycolytica*. *J Bacteriol*, doi: 10.1128/JB.00620-15 (2015).
36. Jennings, M. P. *et al.* The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. *Microbiology* **145** (Pt 11), 3013–3021, doi: 10.1099/00221287-145-11-3013 (1999).
37. Swartley, J. S., Ahn, J. H., Liu, L. J., Kahler, C. M. & Stephens, D. S. Expression of sialic acid and polysialic acid in serogroup B *Neisseria meningitidis*: divergent transcription of biosynthesis and transport operons through a common promoter region. *J Bacteriol* **178**, 4052–4059 (1996).
38. Tzeng, Y. L. *et al.* Transcriptional regulation of divergent capsule biosynthesis and transport operon promoters in serogroup B *Neisseria meningitidis*. *Infect Immun* **69**, 2502–2511 (2001).
39. Bartley, S. N. *et al.* Attachment and Invasion of *Neisseria meningitidis* to Host Cells Is Related to Surface Hydrophobicity, Bacterial Cell Size and Capsule. *PLoS One* **8**, e55798 (2013).
40. Bentley, S. D. *et al.* Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS genetics* **3**, e23, doi: 10.1371/journal.pgen.0030023 (2007).
41. Ivanova, D. *et al.* Shaping the landscape of the *Escherichia coli* chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res* **43**, 7865–7877, doi: 10.1093/nar/gkv704 (2015).

42. Wolfgang, W. J. *et al.* *Neisseria oralis* sp. nov., isolated from healthy gingival plaque and clinical samples. *International journal of systematic and evolutionary microbiology* **63**, 1323–1328, doi: 10.1099/ijs.0.041731-0 (2013).
43. Schoen, C., Kischkies, L., Elias, J. & Ampattu, B. J. Metabolism and virulence in *Neisseria meningitidis*. *Frontiers in cellular and infection microbiology* **4**, 114, doi: 10.3389/fcimb.2014.00114 (2014).
44. Dong, S., Chesnokova, O. N., Turnbough, C. L. Jr. & Pritchard, D. G. Identification of the UDP-N-acetylglucosamine 4-epimerase involved in exosporium protein glycosylation in *Bacillus anthracis*. *J Bacteriol* **191**, 7094–7101 (2009).
45. Jolley, K. A. & Maiden, M. C. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC bioinformatics* **11**, 595, doi: 10.1186/1471-2105-11-595 (2010).
46. Harrison, O. B. *et al.* Description and Nomenclature of *Neisseria meningitidis* Capsule Locus. *Emerg Infect Dis* **19**, 566–573 (2013).
47. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739 (2011).
48. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267 (2006).
49. Didelot, X. & Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251–1266, doi: 10.1534/genetics.106.063305 (2007).
50. Sullivan, M. J., Petty, N. K. & Beatson, S. A. Easyfig: a genome comparison visualizer. *Bioinformatics* **27**, 1009–1010, doi: 10.1093/bioinformatics/btr039 (2011).

## Acknowledgements

This publication made use of the *Neisseria* Multi Locus Sequence Typing website (<http://pubmlst.org/neisseria/>) which is funded by the Wellcome Trust Grant No. 104992 to MCJM and Keith Jolley. SNB and CAM were supported by PhD scholarships from the Amanda Young Foundation (<http://www.amandayoungfoundation.org.au/>). SM is supported by an International Postgraduate Research Student award from the University of Western Australia. KAS is funded by the Australian Research Council (FT100100291). CMK and SNB were supported by the National Health and Medical Research Council (APP546003). CMK and AV are supported by the National Health and Medical Research Council (APP1003697 and APP1078642). MCJM and OBH were supported by the Wellcome Trust (Grant 087622). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

S.N.B. completed the first draft of the manuscript. S.N.B. and K.A.S. performed the purification and enzymatic analyses contributing to Table 1. AV completed the structural biology and modelling of the enzyme contributing to Fig. 3. O.B.H. and M.J.C.M. completed the analysis of the phylogeny of the *galE* allele in all neisserial species see in Fig. 1. T.T.P., S.M. and C.A.M. completed the phylogeny, calculated the associative properties of the alleles with clonal complex and developed the final model for horizontal gene transfer as described in Figs 5 and 6. C.M.K. maintained oversight of the entire project, resources and final edit of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Bartley, S. N. *et al.* Acquisition of the capsule locus by horizontal gene transfer in *Neisseria meningitidis* is often accompanied by the loss of UDP-GalNAc synthesis. *Sci. Rep.* **7**, 44442; doi: 10.1038/srep44442 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017