

Research

The effect of recombination on the evolution of a population of *Neisseria meningitidis*

Neil MacAlasdair,^{1,8} Maiju Pesonen,^{2,8} Ola Brynildsrud,^{3,4} Vegard Eldholm,³ Paul A. Kristiansen,^{3,9} Jukka Corander,^{1,5,6} Dominique A. Caugant,^{3,7} and Stephen D. Bentley¹

¹Parasites and Microbes, Wellcome Sanger Institute, Cambridge CB10 1SA, United Kingdom; ²Oslo Centre for Biostatistics and Epidemiology (OCBE), Oslo University Hospital Research Support Services, Blindern, 0317 Oslo, Norway; ³Division for Infection Control and Environmental Health, Norwegian Institute of Public Health, 0213 Oslo, Norway; ⁴Department of Food Safety and Infection Biology, Faculty of Veterinary Science, Norwegian University of Life Science, 0454 Oslo, Norway; ⁵University of Oslo, Department of Biostatistics, Blindern, 0317 Oslo, Norway; ⁶Helsinki Institute for Information Technology HIIT, Department of Mathematics and Statistics, University of Helsinki, 00014 Helsinki, Finland; ⁷Department of Community Medicine, Faculty of Medicine, University of Oslo, Blindern, 0316 Oslo, Norway

Neisseria meningitidis (the meningococcus) is a major human pathogen with a history of high invasive disease burden, particularly in sub-Saharan Africa. Our current understanding of the evolution of meningococcal genomes is limited by the rarity of large-scale genomic population studies and lack of in-depth investigation of the genomic events associated with routine pathogen transmission. Here, we fill this knowledge gap by a detailed analysis of 2839 meningococcal genomes obtained through a carriage study of over 50,000 samples collected systematically in Burkina Faso, West Africa, before, during, and after the serogroup A vaccine rollout, 2009–2012. Our findings indicate that the meningococcal genome is highly dynamic, with highly recombinant loci and frequent gene sharing across deeply separated lineages in a structured population. Furthermore, our findings illustrate how population structure can correlate with genome flexibility, as some lineages in Burkina Faso are orders of magnitude more recombinant than others. We also examine the effect of selection on the population, in particular how it is correlated with recombination. We find that recombination principally acts to prevent the accumulation of deleterious mutations, although we do also find an example of recombination acting to speed the adaptation of a gene. In general, we show the importance of recombination in the evolution of a geographically expansive population with deep population structure in a short timescale. This has important consequences for our ability to both foresee the outcomes of vaccination programs and, using surveillance data, predict when lineages of the meningococcus are likely to become a public health concern.

[Supplemental material is available for this article.]

Neisseria meningitidis, the meningococcus, is a species of bacteria found exclusively in humans. It can cause meningitis, an infection of the membranes covering the brain and spinal cord, as well as septicemia (Stephens et al. 2007). These infections are difficult to treat, even with antimicrobials, and have a high case fatality rate. Of the 12 serogroups defined on the basis of the structure of the capsular polysaccharide, six (A, B, C, W, X, and Y) are responsible for nearly all cases of invasive meningococcal disease (IMD) worldwide. In contrast to strains that are capable of causing disease, non-disease-causing carriage isolates are typically unencapsulated. However, most infections of encapsulated and unencapsulated *N. meningitidis* are asymptomatic, with the bacteria being carried in the oropharynx of human populations without causing disease with a prevalence of ~5%–10% (Christensen et al. 2010). It is likely that essentially all individuals will be colonized by potentially IMD-causing bacteria once or even several

times during their lifetimes, so there are an uncertain number of carriage infections and transmission events in a human population. This presents a challenge for controlling the disease, and in order to reduce the incidence of IMD, effective polysaccharide-conjugate vaccines against serogroups A, C, W, and Y have been developed and introduced in national vaccination programs. These vaccines are, however, expensive and not affordable for low-income countries. Therefore, a monovalent conjugate serogroup A vaccine was produced and successfully introduced in large-scale vaccination campaigns in countries of the so-called “meningitis belt” of sub-Saharan Africa (Diomandé et al. 2015; Trotter et al. 2017) a region stretching from the Gambia and Senegal to Ethiopia (Molesworth et al. 2002).

Prior to the vaccination campaigns that started at the end of 2010, the overall incidence of meningococcal meningitis in the region was substantially higher than anywhere else in the world and included epidemics that occurred in the winter months every five to 12 years (Trotter and Greenwood 2007). Though the vaccine has been very effective at controlling meningitis epidemics caused by serogroup A, the main cause of IMD in the meningitis belt

⁸These authors contributed equally to this work.

⁹Present address: Coalition for Epidemic Preparedness Innovations, 0473 Oslo, Norway

Corresponding author: nm12@sanger.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.264465.120>. Freely available online through the *Genome Research* Open Access option.

© 2021 MacAlasdair et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

(Diomandé et al. 2015; Trotter et al. 2017), other serogroups (C, W, and X) have emerged or expanded in the region, reducing the initial impact of the vaccine (Topaz et al. 2019). There is also concern that virulent strains circulating in the population might switch capsule or that less virulent strains not covered by the current vaccine might acquire virulence genes (Bärnes et al. 2017; Brynildsrud et al. 2019).

Both of these potential scenarios are driven by the ability of bacteria from the genus *Neisseria* to be naturally transformable and to readily recombine their DNA with one another (Oberfell and Seifert 2015), in concert with selection (Arnold et al. 2020). Various mechanisms of recombination have been described in *N. meningitidis* (Schoen et al. 2009; Marri et al. 2010; Joseph et al. 2011), involving abundant and diverse repetitive DNA sequences in its chromosome. The evolutionary and epidemiological effects of recombination in *N. meningitidis* have been studied in some detail (Marri et al. 2010; Joseph et al. 2011; Retchless et al. 2018), but less work has been undertaken to describe how the extent of recombination varies both between different lineages of *N. meningitidis* and between different regions of its complete genome in a single circulating carriage population. In particular, there is little understanding of how this recombination affects the process of natural selection. Studies across diverse populations and species have suggested that there is likely some variation (Castillo-Ramírez et al. 2012; Ezewudo et al. 2015), and this variation in recombination rate is particularly relevant amid the disruption of population structure caused by large-scale vaccine introduction (Potts et al. 2018).

Burkina Faso, located in the center of the meningitis belt, historically has had a high burden of disease caused by serogroup A meningococci (Nicolas et al. 2005) and was one of the first countries to introduce the serogroup A conjugate vaccine in a mass vaccination campaign in 2010 (Kristiansen et al. 2013). Since then, the incidence of IMD has decreased overall, but there have been meningitis outbreaks caused by serogroups W and X, belonging, respectively, to the sequence types (STs) 11 and 181 (Kristiansen et al. 2013).

Here, we present a detailed population genetic analysis, focusing on recombination, in a collection of 2838 *N. meningitidis* carriage isolates collected from three areas of Burkina Faso over the course of the implementation of the serogroup A vaccine, from 2009 to 2012. This collection has been previously studied using molecular typing techniques (Kristiansen et al. 2013, 2014) which identified ST-181 as the dominant lineage in this population during the time period when sampling was performed. It further showed that the vaccine was effective at reducing the incidence of its target, serogroup A isolates, with none of the previously prevalent serogroup A ST-2859 clone detected after vaccination. This was, however, associated with an increase in the incidence of the disease-causing serogroup X ST-181 complex and also with the introduction and expansion of the disease-causing serogroup W ST-11 complex. In this study, we use whole-genome sequencing with Illumina short-read technology to generate de novo assemblies for each iso-

late, which are then used to construct phylogenies, infer recombination events, and perform tests for selection. Our study recapitulates the finding of previous molecular studies on this population, but leveraging whole-genome data, this study additionally sought to determine whether lineages that make up this population have significantly different recombination rate phenotypes; if different lineages have specific recombination hotspot regions—loci where much more recombination takes place compared to elsewhere in their genomes; to what extent recombination is occurring between the different lineages which make up this population; and finally, whether it is possible to ascertain the evolutionary causes and effects for recombination in this population. Though recombination in *N. meningitidis* has been a known phenomenon and studied for some time (Zhu et al. 1999), we believe this study to be the first to detail the extent of variation in recombination rate within a sampled population and to characterize how and why recombination affects the evolution of the population.

Results

Population structure

The PopPUNK clustering of the successfully sequenced Burkina Faso collection of 2838 carriage isolates returned 17 clusters, five of which were single isolate and three that had fewer than 10 isolates. These eight clusters were not considered in the downstream per-cluster analyses—the remaining nine clusters accounting for 99% of the isolate collection. These clusters broadly correspond to the dominant serogroups and sequence types (STs), as indicated in Figure 1: with cluster 1 being composed entirely of serogroup X and 96% of ST-181; cluster 2 composed 99.8% of serogroup Y and 98% of ST-4375; cluster 3 composed entirely of serogroup W and 98% of ST-11; cluster 4 composed 99% of serogroup W and 93% of ST-2881; cluster 5 composed 99% of serogroup Y and 59% of ST-767; cluster 6 composed entirely of serogroup A and 99% of ST-2859; cluster 7 composed wholly of nongroupable (NG) or capsule-null (cni) isolates and 82% of ST-192; cluster 8 composed entirely of NG/cni and 100% ST-198; and cluster 9 composed 100% of NG/cni and 52% of ST-4899. Among these clusters, we found

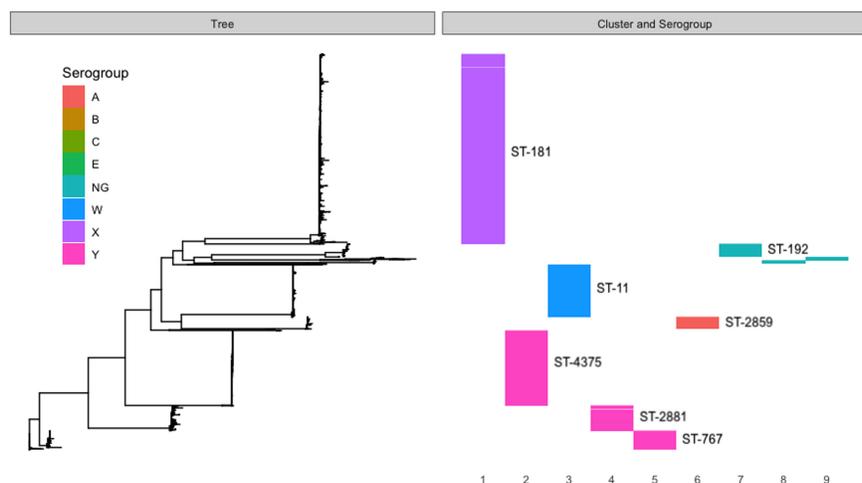


Figure 1. Core genome phylogeny, annotated with PopPUNK-inferred clusters, serogroups, and sequence types for all Burkina Faso isolates in the collection. Clusters are numbered 1–9 from left to right, serogroups are indicated with colors as per the legend, and sequence types for the seven largest clusters are indicated with labels. NG indicates that the isolates were nonserogroupable.

the major disease-causing lineages occurring in Burkina Faso pre- (clusters 3 and 6) and post-vaccine introduction (clusters 1 and 3) (Kristiansen et al. 2013). Disease surveillance since the implementation of the serogroup A vaccine in Burkina Faso has shown these lineages, clusters 1 and 3, to be the predominant causative strains of meningococcal IMD (Diallo et al. 2017). Based on the dominant serogroups and serogroup predictions for each cluster, we can see further that the clusters contain between 1% and 24% isolates which have likely varied their capsular phenotype. Figure 2 includes the entire Burkina Faso collection, as well as the global whole-genome sequences collected as described in the Methods. From the figure, we can see that the Burkina Faso population is composed of nine independently evolving lineages, where isolates from each cluster are more closely related to globally sampled isolates from their respective clusters as opposed to isolates sampled from other clusters within Burkina Faso.

Cluster recombination rates

Average recombination event per mutation event (ρ/θ) rates for each cluster, as well as the overall average rate between all clusters, are shown in Table 1. These were calculated by averaging the per-branch ρ/θ rates inferred by Gubbins. In general, the rates were on the order of 10^{-2} with the average rate for the entire collection being 0.0745. The per-cluster average recombination rate ranged from 0.0082 to 0.1930, a difference of two orders of magnitude. The Kruskal–Wallis test on the per-branch per-mutation rates for each cluster returned an H -statistic of 440.977, with the associated P -value of 3.17×10^{-90} , so there were at least two groups that were significantly different from one another. As the estimated average recombination rates are scaled to the different mutation rates of each cluster, the difference in recombination rates suggested that the clusters must be inherently different in terms of their recombination rate.

To determine which clusters were significantly different from one another, Dunn's test was used for post hoc statistical testing. The full results from this set of pairwise tests are shown in Figure 3 with markers for a significant difference from another cluster on the top of the individual violin plots. In general, the rates of re-

Table 1. Average ρ/θ , recombination events per mutation, and θ , number of nonrecombinant mutations per branch, for each cluster as determined by cluster-by-cluster Gubbins analysis

Cluster (dominant serogroup and ST)	ρ/θ	Number of isolates	θ
Cluster 1 (X/ST-181)	0.1053	1361	3.72
Cluster 2 (Y/ST-4375)	0.0082	539	6.63
Cluster 3 (W/ST-11)	0.0249	373	1.21
Cluster 4 (W/ST-2881)	0.1460	180	8.99
Cluster 5 (Y/ST-767)	0.0710	135	8.19
Cluster 6 (A/ST-2859)	0.0542	86	4.16
Cluster 7 (NG/ST-192)	0.0524	95	25
Cluster 8 (NG/ST-198)	0.0154	22	27.7
Cluster 9 (NG/ST4899)	0.1930	23	49.1
Average	0.0745		14.97

combination span three orders of magnitude, from 10^{-1} to 10^{-3} , with clusters 1, 4, and 9 being on the order of 10^{-1} , cluster 2 being on the order of 10^{-3} , and the remainder being on the order of 10^{-2} . This post hoc difference testing suggested that there are essentially three recombination phenotypes—the highly recombinant lineages (and serogroups): 1 (X/ST-181), 4 (W/ST-2881), and 9 (NG/ST4899), the moderately recombinant clusters/serogroups: 3(W/ST-11), 5(Y/ST-767), 6(A/ST-2859), 7(NG/ST-192), and 8(NG/ST-198), and the relatively nonrecombinant cluster 2 (Y/ST-4375).

Within-cluster loci with elevated recombination rates

Recombination across the genome for each cluster is summarized in Figure 4, and the genes contained within each cluster's candidate hotspots are fully described in Table 2. From these results, specifically those shown in Figure 4, we can see that in most clusters there are regions of the genome with substantially more recombination than the background level, though we cannot distinguish their relative levels of significance. Seven of the clusters had obvious peaks in the Manhattan plot of recombinations, whereas clusters 2 and 6 did not. Inspecting the annotation of the reference genomes for each cluster revealed a number of genes frequently present within these regions with elevated levels of recombination. In particular, *pilES*, the pilin genes, were present in clusters 1, 4, 5, 7, 8, and 9. Clusters 1, 4, 5, and 7 had regions containing genes associated with iron uptake. The transferrin-binding protein gene *tbpB* was present in clusters 1, 4, 5, and 7, and various portions of the bipartite outer membrane hemoglobin receptor gene *hpuA/B* were present in clusters 4 and 5.

Recombination between clusters

The network of recombinant sequence region counts between donor and recipient clusters inferred from each gene in the pan-genome of the nine major clusters in the Burkina Faso collection is illustrated in Figure 5. A loose correlation between the size of each cluster and the number of contiguous recombinant sequences identified within was seen, though there were exceptions to this

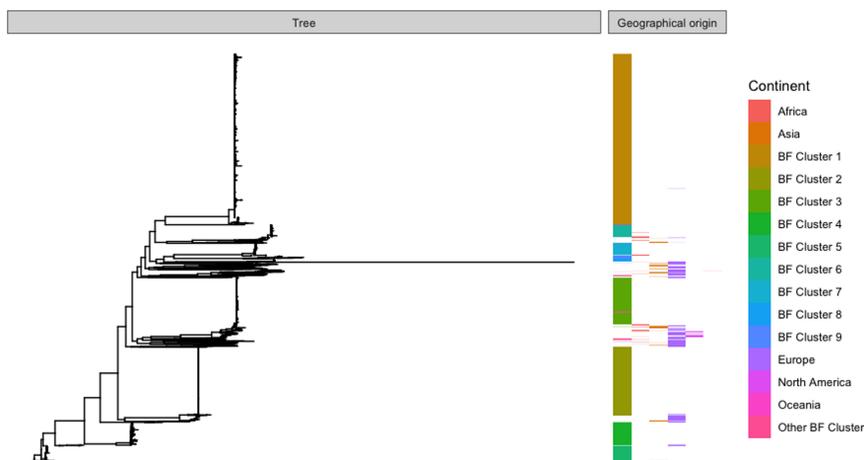


Figure 2. Core genome phylogeny of the Burkina Faso collection with the global background collection, arranged in order by first Burkina Faso and then alphabetical by continent, and annotated with colors for cluster, for Burkina Faso isolates, and continent of origin, for the global background collection (cf. Supplemental Table S4).

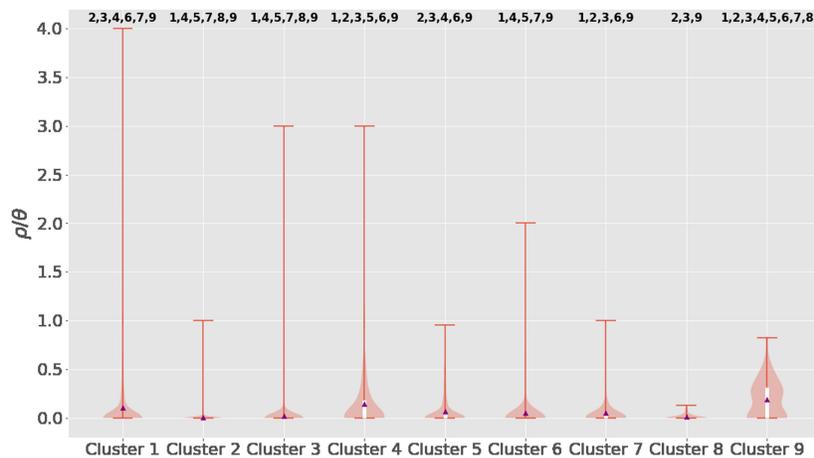


Figure 3. Violin plot of the per-isolate ρ/θ , recombination events per mutation event, as calculated by Gubbins for each cluster. The average ρ/θ per cluster is indicated by the purple triangles and these are also enumerated in Table 1. The top and bottom of the white boxes indicate the third and first quartiles, respectively, and the whiskers of the plot represent the maximum and minimum values. The orange background shading represents the distribution of inferred recombination rates within each cluster. Significant differences between clusters, as determined by a Kruskal–Wallis nonparametric analysis of variance on all the per-branch rates for each cluster, followed by post hoc statistical testing for differences between groups using Dunn’s test and the conservative Holm–Bonferroni correction for multiple testing, is indicated by cluster numbers above each cluster’s violin plot.

pattern that reflected the per-cluster estimates of recombination rate for each cluster. For instance, cluster 5 was involved in more exchanges of contiguous recombinant sequence than cluster 4, despite being smaller at 135 isolates compared to 280. This was also true of the smallest clusters 8 and 9, where 9 was involved in many more exchanges of contiguous recombinant sequence than 8.

Among the main lineages (clusters 1–9), clusters 1, 3, and 4 received more DNA by recombination than they donated, whereas 2, 5, 6, 7, 8, and 9 donated more DNA than they received. The results indicated that clusters with lower recombination rates, which make up a smaller proportion of the population (e.g., clusters 2, 5–9), donate more of the identified contiguous recombinant sequences than they receive, whereas clusters with a higher recombination rate or that are larger in population size, namely 1 and 4, and 1 and 3, respectively, receive more DNA than they donate.

Across the pan-genome of 2861 genes, the average number of contiguous recombinant sequences identified per gene was 63.03, with 1084 genes not having any recombinant sequences detected, and the number of recombinant sequences being detected per gene ranging from 1 to 9015, in genes where a nonzero number of recombinant sequences were detected. Six genes had more than 5000 separate contiguous recombinant sequences inferred: an IS200-like transposase, the flavinyl transferase *apbE*, an unnamed P1 outer membrane protein, the microcin resistance gene *tldD*, the adhesin gene *mafA*, and the sRNA regulatory protein *yhbJ*.

Selection and recombination in the pan-genome

d_N/d_S was successfully estimated with SNPGenie for 1804 genes of the 2861 genes in the pan-genome, with the remaining genes being singletons or not having enough diversity to allow the calculation to be performed. The average d_N/d_S for the pan-genome was 1.563, with values for individual genes ranging from 0 to 396.66. Relatively few genes, 119 of 1804, had a d_N/d_S greater than one (Supplemental Table S2). Eleven genes had unusually high estimates for d_N/d_S —over 70—and the next highest d_N/d_S estimate

for a gene was 27. Six of these genes were hypothetical proteins without a known annotation, two genes were phage-associated and had some pseudo-genized copies, and the last three genes were *hpallM*, a restriction enzyme, an uncharacterized peptidase, and the *exbD* membrane protein. As these genes are generally not well-described, they likely include artifacts from the annotation or pan-genome inference stage or are simply unusual regions of the genome that do not evolve like other genes, resulting in unusually high d_N/d_S estimates for these genes. Consequently, these 11 genes were excluded from figures and further analysis.

Even with those genes excluded, correlating the d_N/d_S and number of separate recombinant sequence regions identified per gene in the pan-genome with a nonparametric Spearman’s rank correlation returned a highly significant negative correlation coefficient of -0.122 , with $P=2.21 \times 10^{-6}$. As can be seen in Figure 6, however, one gene has a d_N/d_S

value of over 10 and 42 separate recombinant sequence regions identified within. This is *dpmA*, a DNA methylation modification enzyme. Additional analysis of this gene with BUSTED and FUBAR confirmed that it is under gene-wide selection (BUSTED, $P=0.00015$) and that a single amino acid site is under selection (FUBAR, posterior probability=0.9039). The 42 branches of the phylogeny which featured the recombination were also tested with aBSREL to see whether they were under selection, and one of these branches was found to be under selection ($P=0.00059$).

Discussion

The first, immediate conclusion from the composition of our sample of *N. meningitidis* carriage isolates from Burkina Faso is that, over the course of the sampling period, the population was composed primarily of nine independently evolving lineages. Bacteria of these lineages were more closely related to isolates of the same lineage from other parts of the world than to isolates from other lineages within Burkina Faso, as indicated by Figure 2, reflecting the deeply stratified population structure of *N. meningitidis*, where clonal lineages with a distant common ancestor coexist in space and time, evolving largely independently. Seven of the lineages making up the Burkina Faso population were present in all four years of sampling. Two lineages, however, were not consistently detected. Cluster 3, the serogroup W/ST-11 lineage, was first detected in our collection in 2011 and then also in 2012 but was not present in 2009 or 2010, although it caused a large epidemic in Burkina Faso in 2002 (Koumaré et al. 2007). This may therefore represent either a re-introduction of the lineage after it had gone locally extinct or an increase in population size to be once again detectable after the population perturbation caused by vaccination. Cluster 6, the serogroup A lineage, is the other cluster which was not detected throughout our sampling period, being found in 2009, 2010, and a single isolate in 2012. The isolate in 2012 was substantially divergent from the rest of the lineage

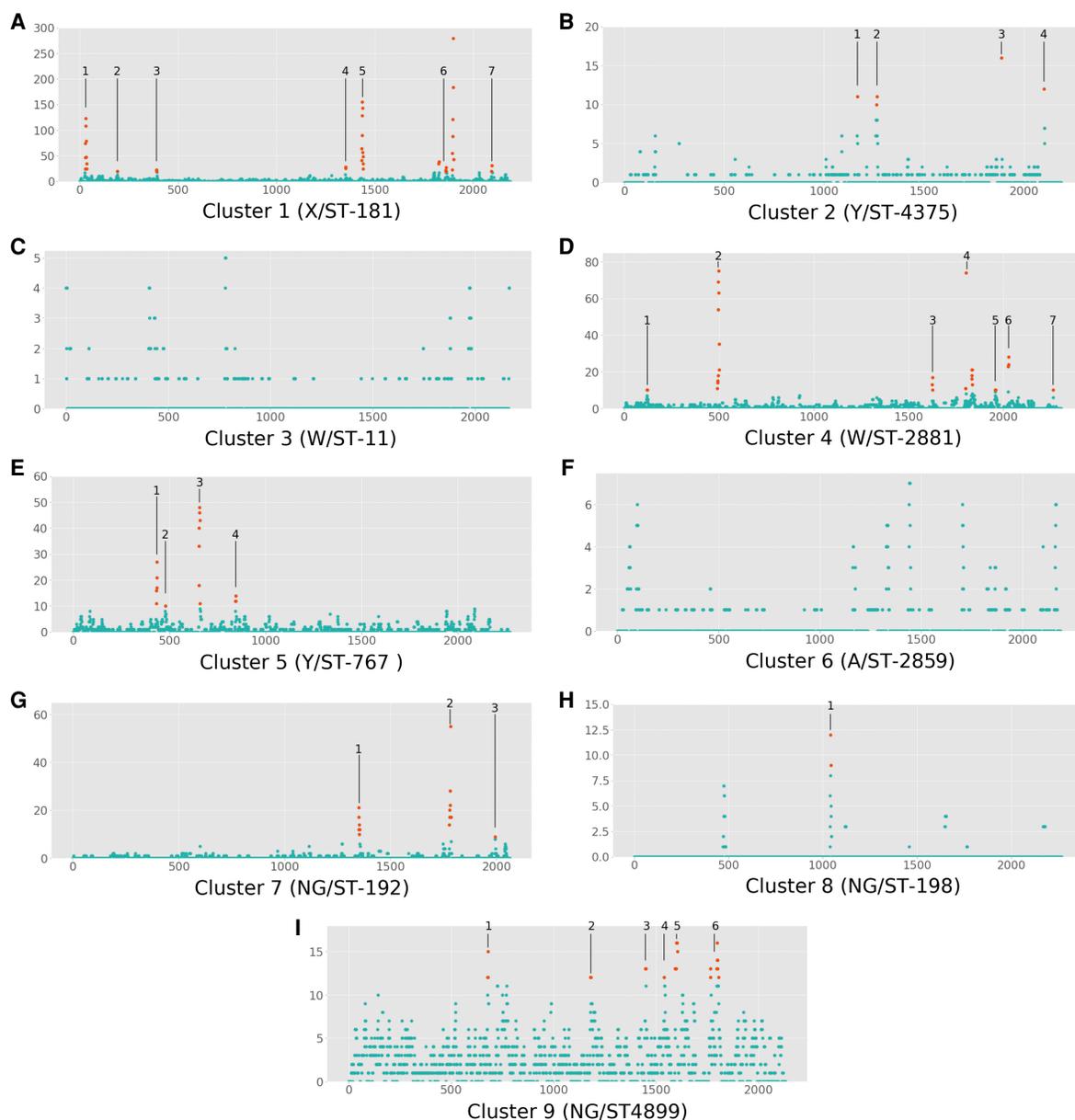


Figure 4. Per-cluster Manhattan plots of the number of recombinations per discrete 1000-base pair window in each cluster's unique reference genome. Plots for Clusters 1–9 are in panels A–I, respectively. The number of recombinations in a window is indicated on the y-axis, and the position on the genome, in kilobases, is on the x-axis. Regions with elevated rates of recombination, determined manually based on obvious peaks in the plot, are highlighted in orange, and the genes within each numbered region are fully described in Table 2.

(Supplemental Fig. S1) and belonged to a different sequence type, ST-7, whereas the earlier isolates of the lineage belonged to ST-2859. Given that this lineage was the target of the vaccine introduced in Burkina Faso at the end of 2010, it seems likely that the original serogroup A lineage was extinct by 2011 in Burkina Faso and the 2012 isolate represents a re-introduction from abroad. The dynamic changes in the composition of our population underscore the need for carriage surveillance in monitoring the evolution of *N. meningitidis* in order to guide future intervention. Though there were shifts in the population composition of our sample, including lineages which were restricted to a subset of the sampling period, a majority of the lineages that made up the study population were present throughout and therefore do reflect

a fluctuating yet consistent population in Burkina Faso across the time period of our sampling.

The significant differences in the levels of recombination between different clusters is an important and interesting result. As far as we are aware, this is the first study of a large, contemporaneous population of *N. meningitidis* that has used whole-genome data to estimate and compare the recombination rates of different lineages within the population. Though it has been shown in other species and with more disparate sampling, we show here that, even within narrow geographical and temporal constraints, the lineages composing this population of *N. meningitidis* differ in their propensity to recombine by orders of magnitude. This has important implications for how likely it is for any bacterium

Table 2. Regions with elevated rates of recombination found by inspecting discrete window Manhattan plots of recombination counts for peaks

Cluster (dominant serogroup and ST)	No. of hotspots	Genes in hotspots
Cluster 1 (X/ST-181)	7	1: (Prolyl endopeptidase, <i>argA</i> , MJ0042 family finger-like domain gene, <i>pyrE</i> , putative uracil-DNA glycosylase, ribosomal-protein-alanine acetyltransferase, <i>yeaZ</i> , hypothetical lipoprotein, <i>fda</i> , <i>xerC</i> , <i>dxs</i> , <i>miaB</i>) 2: (<i>adhA</i> , <i>fimA</i> , <i>macA</i>) 3: (<i>tbpB</i> , <i>lbpA</i>) 4: (<i>glmM</i> , <i>dhpS</i> , <i>aroF</i> , putative AsmA-like protein) 5: (<i>tbpA</i> , <i>tbp2</i> , <i>glr</i> , putative <i>yjeE</i> -like ATPase, <i>amiC</i> , Fic/DOC cell division protein, <i>rlmL</i>) 6: (<i>ntrX</i> , <i>kinB</i> , <i>lipB</i> , <i>lipA</i> , <i>rbcC</i> , <i>rfaA</i> , <i>pilT</i> , <i>yccS</i> , <i>pilS5</i> , <i>pilE</i> , <i>pilS2</i> , <i>pilE</i> , <i>lpxC</i> , <i>yqjI</i>) 7: (<i>porB</i> , <i>fetA</i> , <i>trmE</i>)
Cluster 2 (Y/ST-4375)	4	1: (IS1106 transposase) 2: (<i>tspB</i> , phage-associated protein) 3: (<i>lpxC</i>) 4: (ISNme1 transposase, Fic/DOC hypothetical protein)
Cluster 3 (W/ST-11)	0	NA
Cluster 4 (W/ST-2881)	7	1: (16S rRNA, tRNA-Ile, tRNA-Ala) 2: (Fic/DOC cell division protein, <i>amiC</i> , hypothetical ATPase, <i>glr</i> , hypothetical protein, <i>tbpB</i> , <i>tbpA</i>) 3: (<i>fda</i> , lipoprotein, <i>yeaZ</i> , putative acetyltransferase) 4: (<i>lpxC</i> , <i>pilE</i> , hypothetical protein) 5: (<i>yccS</i> , <i>pilT</i> , <i>yggS</i>) 6: (<i>dtpT</i> , hypothetical periplasmic protein, <i>hpuB</i> , <i>hpuA</i>) 7: (<i>pilC</i>)
Cluster 5 (Y/ST-767)	4	1: (<i>pilE</i> , <i>pilS6</i> , <i>pilS5</i> , <i>pilE_2</i> , fimbrial protein P9-2 precursor, hypothetical protein, <i>pilE_3</i>) 2: (<i>PilC</i>) 3: (<i>tbpA</i> , <i>tbpB</i> , fimbrial protein P9-2 precursor), (<i>hpuB</i> , hemoglobin-haptoglobin-utilization protein) 4: (heme oxygenase, putative lipoprotein, putative <i>tbp</i> -like)
Cluster 6 (A/ST-2859)	0	NA
Cluster 7 (NG/ST-192)	3	1: (<i>tbpA</i> , <i>tbpB</i> , <i>glr</i>) 2: (<i>pilE_1</i> , <i>pilE_2</i> , <i>pilE_3</i> , <i>pilE_4</i> , <i>pilE_5</i> , <i>pilE_6</i> , hypothetical protein, <i>pilE_7</i> , <i>lpxC</i> , IS1106 transposase, IS1106A3 transposase) 3: (hemoglobin-haptoglobin-utilization protein)
Cluster 8 (NG/ST-198)	1	1: (<i>pilE_2</i> , <i>pilE_3</i> , <i>pilE_4</i> , <i>pilS5</i>)
Cluster 9 (NG/ST-4899)	6	1: (uracil-DNA glycosylase, <i>pyrE</i> , MJ0042 family finger-like protein, <i>argA</i>) 2: (<i>dapA</i> , hypothetical lipoprotein, <i>pip</i>) 3: (<i>glmM</i> , <i>dhpS</i> , AsmA-like protein) 4: (<i>aniA</i> , <i>nor</i>) 5: (<i>lbpA</i> , <i>tbpB</i> , <i>tbpA</i> , <i>tbpB_2</i>) 6: (<i>pilC</i> , <i>pilE_4</i> , <i>pilE_5</i> , <i>lpxC</i> , <i>yqjI</i> , <i>alx</i>)

The genes contained within a single region are indicated with a set of parentheses.

from a given lineage to pick up genetic material from another lineage, potentially a disease-causing one. This is of particular importance with regard to the efficacy of current vaccines and the design of future ones.

The rate of recombination within clusters was not entirely determined by the clusters' relative size in the population—and hence stochastic opportunity, assuming that the proportions of lineages in our sample reflect those in the general population—and though recombination rate is doubtlessly affected by many factors, this does suggest that there are lineage-specific genetic components affecting the rate of recombination in *N. meningitidis*. Though there exists an understanding of the mechanisms that underlie recombination in the species (Oberfell and Seifert 2015), knowledge of how these and other genetic factors that affect the rate of recombination differ between lineages has not yet been fully explored. This awareness would not only be important to our understanding of the evolution of *N. meningitidis* and other *Neisseria*, but it would also allow for a more detailed monitoring of the populations of these serious human pathogens. Finally, it is worth noting that the highest rate of recombination, more than twice the average, was seen in one of the NG clusters, cluster 9. Nonserogroupable *N. meningitidis*, which may lack the whole capsule operon or simply have loss-of-function mutations in key capsular genes, are rarely implicated in disease. However, given the extremely high recombination rate of the nongroupable cluster 9, some NG/cnl *N. meningitidis* bacteria may play a larger role in the evolution of a geographical population than their frequency would suggest. Therefore, it may prove worthwhile to monitor NG/cnl *N. meningitidis* more closely in future study.

Recombination in the pan-genome of all the lineages follows a pattern that is loosely consistent with the estimated within-cluster rates. Clusters with higher estimates of average recombination rate, ρ/θ , also generally have more inferred within-gene separate contiguous recombinant sequence in the pan-genome than clusters with lower estimates of ρ/θ , when the population size of the cluster is taken into account. For instance, clusters 8 and 9 have

a similar population size but differ greatly in the number of between-cluster recombinant sequences detected in each cluster. It is clear, though, that the primary factor that determines how much recombination lineages undergo in the pan-genome is the proportion of each cluster within the population at large, and hence the stochastic opportunity of recombining, particularly in terms of the likelihood of an encounter between two random individuals of different lineages. This is in contrast to the within-lineage recombination rates which were not correlated with population size. The recombination pattern in some clusters also suggests that the inherent recombination rate of a lineage affects how recombination proceeds between itself and other lineages. For example, cluster 2, despite being the second largest cluster, donates DNA more often than it receives DNA, like the much smaller clusters 5–9. This is as one would predict, as the donation of DNA in a recombination event is governed principally by the rate at which bacteria from different lineages physically encounter one another in the same host, whereas receiving DNA is further affected by how likely the receiving bacteria is able to take up the DNA. Therefore, in lineages with low inherent recombination rates, such as cluster 2, we expect to see a bias toward donating rather than receiving of DNA, compared to what is expected based on the size of the cluster. Though this effect is clearly dwarfed by the effect of cluster size in general, this suggests that more recombinant lineages of *N. meningitidis*, may be able to incorporate and maintain more variation from other lineages, including from lineages that themselves are not particularly recombinant (Supplemental Fig. S2). Although this may not be particularly surprising, it suggests that the inherent recombination rate has important consequences for how flexible the genome of a lineage is likely to be in response to different selection pressures and how likely it is that it may become more virulent, antibiotic-resistant, or escape a vaccine.

The radically different amounts of recombination in different regions across the genome of most lineages within this population, combined with the discovery of consistent genes within regions of elevated recombination, suggest two things: first,

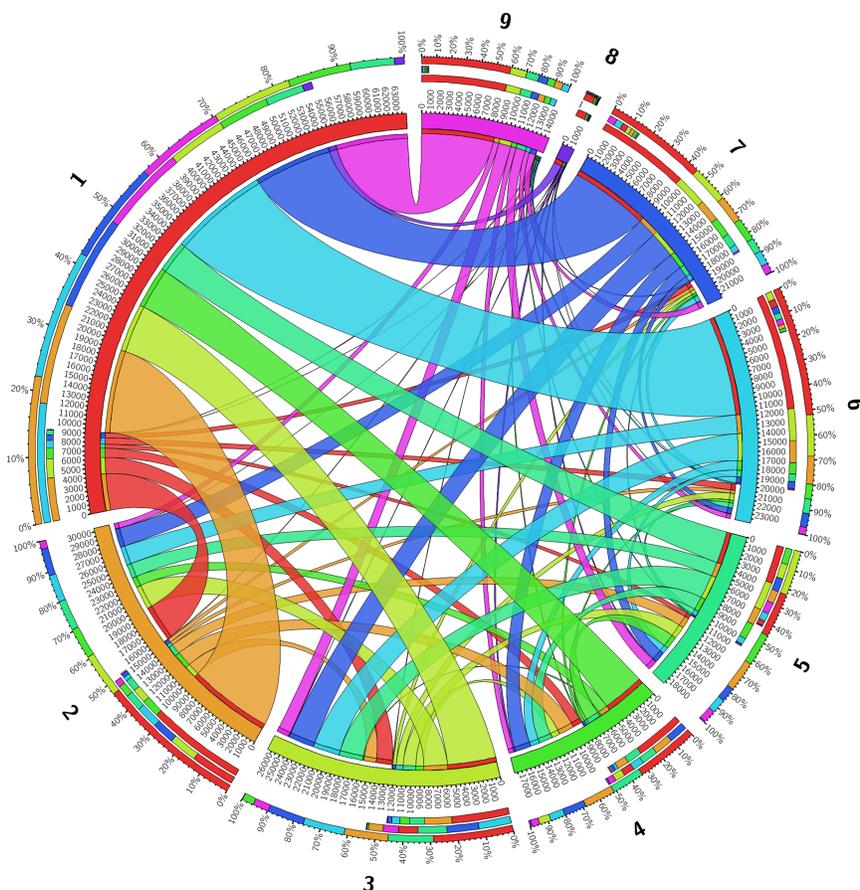


Figure 5. Chord diagram of the count of recombinant regions identified by fastGEAR between the clusters in the Burkina Faso *N. meningitidis* carriage collection. Clusters are positioned on the main circle of the diagram, with the arc length of the cluster indicating the number of recombination events. Linkages between clusters represent the number of recombinant regions occurring between those clusters with their width representing the number of regions and the color indicating the donating lineage. The three stacked bars *outside* the main diagram indicate, from *outermost* to *innermost*, the proportion of the total number of recombinant regions in each cluster colored by the other cluster involved, those same proportions only for the count of recombinant regions received, and those proportions for regions donated in the focal cluster.

that recombination in the *N. meningitidis* genome is elevated in certain regions; and second, that multiple lineages within the population are experiencing routine ongoing adaptation to human hosts. In particular, the elevated rates of recombination in pilin and iron uptake genes—found across clusters 1, 2, 4, 5, 7, 8, and 9, as well as in previous work (Linz et al. 2000)—corroborates that these genes are crucial determinants of the ability of *N. meningitidis* to survive within a human host. Pilin genes, expressed on the surface of the bacteria, are targeted by the immune system (Wachter and Hill 2016), and the *pilE* gene is further known to be highly recombinant (Aho et al. 2005), due in part to the various mechanisms of phase variation. Human hosts are typically very iron-deficient environments for bacteria to grow in, resulting in a strong selective pressure on iron uptake pathways (Perkins-Balding et al. 2004; Yu et al. 2014), and iron-binding proteins have been shown to be very diverse and have repeats which may promote both intra- and inter-genomic recombination close by (Acevedo et al. 2014). Further work should explore methods of determining which of these regions of elevated recombination are statistically significant and consistently so across data sets which

are more temporally and geographically diverse, to fully explore the genetic and evolutionary factors driving elevated recombination rates in these regions.

The pattern of recombination events in the per-gene analysis of the pan-genome is largely as expected given the within-cluster results, with an IS200 transposase, surface-exposed genes such as the adhesin *mafA*, and the microcin resistance gene *tldD* among the top five most recombinant genes. Other genes expected to be highly recombinant, such as the iron uptake genes *tbp2/tpbB* and the surface-exposed T cell stimulating protein *tspB*, also have many inferred recombinations and are among the top 15 most recombinant genes. Although these genes are known to be generally very recombinant in *Neisseria*, it is also likely that the disruption of the population by vaccination has caused substantial shifts in the population size of each lineage which, in most cases, apart from cluster 6, was likely a population expansion. During this demographic shift, lineages will have had more opportunity to selectively adapt to different hosts and also persist to exchange DNA, and these events are presumably contributing to the consistently high levels of recombination detected in different genes, particularly immune targets, across the different lineages.

The d_N/d_S values for each gene in the pan-genome follow a pattern that is broadly expected, with most values being very low, and relatively few genes—119—having a $d_N/d_S > 1$. It is difficult to draw specific conclusions from a simple gene-wide estimate of d_N/d_S , but the significant negative correlation with the number of

separate contiguous recombinant sequences for each gene in the pan-genome and the gene's estimated d_N/d_S can inform us as to the general evolutionary role played by recombination in this

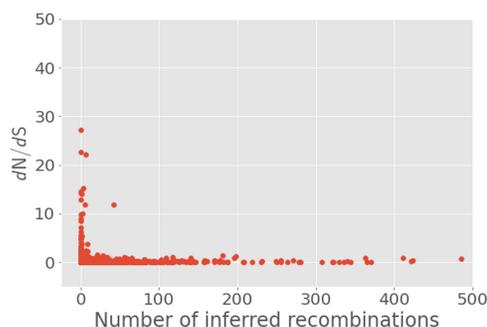


Figure 6. Scatter plot of number of recombination events versus d_N/d_S for genes in the pan-genome of the entire Burkina Faso collection of carriage isolates. Genes with abnormally high estimates of d_N/d_S were excluded from the plot as described in the text.

population. One of the evolutionary explanations of the maintenance of recombination in populations has been that it prevents the accumulation of deleterious mutations (Felsenstein 1974). Given the significant negative correlation count of recombinant sequences and d_N/d_S in the pan-genome, and the fact that nonsynonymous mutations are generally more likely to be deleterious, avoiding the accumulation of deleterious mutations, or enhancing the effect of negative selection, seems to be the primary role played by recombination in this population. In spite of this, a single gene in the pan genome, the *dpnA* DNA methylation modification enzyme, has both a substantial number of recombinant sequences, 42, and an elevated d_N/d_S estimate of 11.8. All of these sequence regions identified as recombinant originate from an unknown bacterium outside the collection, and 41 ended up in the cluster 1 lineage, with the final one ending up in cluster 2. Further analysis of this gene with the BUSTED algorithm for detecting gene-wide episodes of selection confirmed that it has been under selection, and using the FUBAR method for detecting specific sites under selection confirmed that the *dpnA* gene has definitely been under selection in this population. Though the specific site under selection detected by FUBAR did not overlap with any of the sequence regions identified as recombinant, testing the branches of the gene phylogeny with the recombination present with aBSREL found one of the lineages possessing the recombinant sequence to be under selection as well. This example of a high recombination gene under selection is an exception to the general pattern of recombination in the Burkina Faso population of *N. meningitidis* and suggests that recombination may be acting here to speed adaptation instead of simply preventing the accumulation of deleterious mutations. Though recombination acting in this manner has long been theorized to be true for all recombining organisms under the Fisher-Muller model of recombination, and shown experimentally with plasmids in *Escherichia coli* (Cooper 2007), we demonstrate this phenomenon here in a sample from a natural population.

Evidence of adaptation speeded by recombination was seen in the *dpnA* gene in the cluster 1 lineage. Though it is uncharacterized in *N. meningitidis*, the *dpnA* gene, as part of the DpnII restriction-modification system, has been studied in *Streptococcus pneumoniae*, another bacterial pathogen that primarily colonizes the nasopharynx of healthy humans. In *S. pneumoniae*, *dpnA* is specifically produced during natural competence for promoting recombination by methylating incoming DNA to protect it from the rest of the DpnII system (Johnston et al. 2013). Though we cannot assume that *dpnA* has an identical function in *N. meningitidis*, given the important consequences that DNA methylation is known to have for recombination in *N. meningitidis* (Seib et al. 2017; Kim et al. 2019) and the on-average higher rate of recombination in cluster 1 isolates with this *dpnA* recombination (0.159 vs. 0.105), it seems that *dpnA* recombination may have affected the overall recombination phenotype in the cluster 1 lineage.

The effects of recombination on the evolution of various genes in the *N. meningitidis* genome have been the subject of extensive research. This research has largely been focused on the effects of recombination in a specific set of genes and over a long period of time (Vigué and Eyre-Walker 2019). However, our understanding cannot be complete without also investigating how recombination affects the evolution of a population of *N. meningitidis* in the short to medium term. Without this, it would be difficult to predict how different populations will respond to vaccines, anticipate changes in virulence between different lineages, or know if a population has a high risk of developing antimicrobial

resistance. In this study, we have conducted such an analysis on a collection of *N. meningitidis* carriage samples from Burkina Faso. We have demonstrated how using high-throughput whole-genome sequencing and subsequent computational analyses on whole-genome data allow for the possibility of understanding the differential effects of recombination on the evolution of the various lineages within this population, how the lineages of the population are interacting with each other, and the evolutionary explanations which underlie these patterns.

Methods

Sample collection and laboratory analysis

To assess the impact of the monovalent serogroup A conjugate vaccine on meningococcal carriage, 50,811 samples were collected over the course of four years, from 2009 to 2012, in 10 rounds of sampling. This took place across three sites in Burkina Faso: the Bogodogo arrondissement of the capital city, Ouagadougou; 10 villages in the Kaya district, 100 km northeast of Ouagadougou; and 10 villages in the Dandé district, 350 km west of Ouagadougou. Oropharyngeal swabs were taken from healthy volunteers, and associated metadata were collected from the individuals. A total of 2848 meningococcal isolates were recovered from the 10 samplings (overall carriage rate 6.05%) and confirmed as *N. meningitidis* at the Norwegian Institute of Public Health (NIPH), Oslo (Kristiansen et al. 2011, 2013, 2014). The isolates were serogrouped using commercial antisera (Remel), most of them were characterized by multilocus sequence typing (Maiden et al. 1998), and all were stored frozen in Greaves medium (Craven et al. 1978) at -70°C .

Whole-genome sequencing, quality control, assembly, and annotation

DNA was extracted from 2839 of the 2848 collected isolates and these were sent to be sequenced at the Wellcome Sanger Institute, using Illumina HiSeq 2000 sequencing technology. Paired-end libraries were prepared with insert sizes of 500 base pairs, for 125-base pair reads, sequenced at a read depth of 100x. The quality of the raw sequencing reads was assessed using an internal quality control (QC) pipeline as well as a local implementation of Kraken (Wood and Salzberg 2014), the metagenomic sequence classifier, to check for contamination. Raw reads from 2838 isolates passed this QC and were then used to produce de novo genome assemblies (Supplemental Table S3) using the bacterial assembly pipeline at the Sanger Institute (Page et al. 2016). The resulting genome assemblies were annotated using Prokka (Seemann 2014), the prokaryotic annotation pipeline, to infer gene content. To identify the ST of all isolates from the sequence data, we used SRST2 (Inouye et al. 2014), the *Neisseria* PubMLST sequence typing database, and the raw reads for each isolate. Finally, in order to determine serogroups for each isolate in silico, we adapted seroBA (Epping et al. 2018) using the published capsule reference sequences (Harrison et al. 2013). In addition to the collected isolates from Burkina Faso, to give global genetic context, genomes of an additional 427 publicly available *N. meningitidis* isolates from five continents and maximizing serogroup diversity, as well as the FA1090 *Neisseria gonorrhoeae* reference as an outgroup, were downloaded from the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena/browser/home>) and processed using the same quality control, assembly, annotation pipeline, and further analyses as above where applicable (Supplemental Table S4).

Selected isolates from five of the clusters in the collection were resequenced with Oxford Nanopore MinION at the NIPH,

with R9.4 (FLO-MIN106) flowcells and the SQK-RBK001 rapid barcoding kit for 1D reads, in order to produce reference genomes for those clusters. De-multiplexing and base-calling was performed using Albacore version 2.1.2 (Sahoo 2017), and then porechop version 0.2.3 (Wick et al. 2017a) was used to remove chimeric reads and adapters from the nanopore reads. The reference genomes were then assembled by hybrid assembly using both the original Illumina sequence and the nanopore reads, with Unicycler (Wick et al. 2017b) before being annotated with Prokka (Seemann 2014) like the rest of the sequenced isolates.

Clustering and phylogeny inference

To assess the population structure of the collection of carrier isolates, we clustered the sequence assemblies using PopPUNK (Lees et al. 2019). For four clusters among the nine largest carried forward for further analysis, an Oxford Nanopore-sequenced reference was not available. Therefore, the assembly with the lowest number of contigs was selected for further scaffolding, using the MeDuSa multidraft assembly scaffolder (Bosi et al. 2015) and the rest of the isolate assemblies from that cluster. The resulting scaffolds for each cluster were then joined with gaps of 1000 blank nucleotides. The resulting sequences, or the Oxford Nanopore-sequenced references, were then used as reference genomes in producing whole-genome pseudoalignments for each cluster. We did this using a custom mapping, variant calling, and local realignment around indels pipeline using BWA-MEM (Li 2013), SAMtools mpileup (Li 2011), and MUSCLE (Edgar 2004), and then used the resulting whole-genome pseudoalignments to infer phylogenies for each cluster, using Gubbins (Croucher et al. 2015), and RAXML, its underlying dependency (Stamatakis 2014).

In order to construct a phylogeny for the entire collection and also the collection combined with global isolates, we used the recently published Panaroo pan-genome pipeline (Tonkin-Hill et al. 2020) to infer a set of core genes for the entire collection. We concatenated alignments of all the core genes and then used IQ-TREE (Nguyen et al. 2015), with the substitution model, GTR + F+I + G4, inferred by ModelFinder (Kalyaanamoorthy et al. 2017), to infer phylogenies for both the entire Burkina Faso collection and the Burkina Faso collection plus global isolates.

Recombination, selection, and pan-genome analyses

Recombination was first analyzed cluster-by-cluster, using whole-genome pseudoalignments of each cluster, generated as described above, and Gubbins (Croucher et al. 2015). Gubbins outputs an estimate of ρ/θ , the number of recombination events per mutation event, for each branch of the phylogeny (Supplemental Table S1). To test if any of the differences in recombination rates between clusters were statistically significant, we used a Kruskal–Wallis nonparametric analysis of variance (Kruskal and Wallis 1952) on all of the estimated per-branch rates for each cluster, followed by Dunn’s test for post hoc statistical testing (Dunn 1964) for differences between groups and the conservative Holm–Bonferroni correction for multiple testing (Holm 1979). Candidate recombination hotspot regions in the reference genome of each cluster were found by producing Manhattan plots of the number of recombination events per 1000-base pair discrete window across the genome, manually looking for peaks in these plots, and comparing them to a Prokka annotation of the genome to establish which genes, if any, are within these windows.

To then analyze the recombinations between clusters in the entire Burkina Faso collection, we ran fastGEAR (Mostowy et al. 2017) on alignments of each gene of the Burkina Faso collection’s pan-genome, as inferred by Panaroo (Tonkin-Hill et al. 2020) and

aligned using MAFFT (Kato and Standley 2013), using the whole-genome clustering inferred by PopPUNK (Lees et al. 2019). fastGEAR proceeds base pair-by-base pair in a multiple sequence alignment to identify recombinant regions by identifying the most likely lineage of origin for each base pair and, as such, identifies contiguous sequence regions whose origin is more likely to be from a different lineage than a given isolate’s own and hence is recombinant. Although this approach does lead to some uncertainty with regard to how the number of these regions correspond to the number of physical events, this is not a confounding bias. The further advantage of fastGEAR as a method for detecting recombination is in how it infers recombination explicitly with directionality, allowing the flow of recombination events between clusters to be visualized as a network. This was done using Circos Table Viewer and Circos (Krzywinski et al. 2009). To assess the effect of selection on the pan-genome, d_N/d_S ratios were also inferred for each gene using the same alignments as input into fastGEAR and the implementation of the Nei–Gojobori method (Nei and Gojobori 1986) of calculating d_N/d_S as implemented in SNPGenie (Nelson et al. 2015). These d_N/d_S results were then correlated with the number of recombination events per gene using a nonparametric Spearman’s rank correlation (Spearman 1904). Genes with an elevated SNPGenie d_N/d_S estimate and a high number of recombinations were further analyzed with the HyPhy package (Pond and Muse 2005), in particular, the BUSTED (Murrell et al. 2015) method for detecting gene-wide episodic selection, the FUBAR (Murrell et al. 2013) method for finding codons under selection, and the aBSREL (Smith et al. 2015) method to find selected branches in a phylogeny.

The SciPy (Virtanen et al. 2020), pandas (McKinney 2010), and Matplotlib v. 2.2.3 (Hunter 2007) Python libraries were used throughout this study for statistics, data manipulation, and data visualization, respectively. The R (R Core Team 2019) package ggtree (Yu et al. 2017) was used to draw phylogenies.

Data access

All raw sequencing data generated in this study have been submitted to the European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/ena>) under the study accession number PRJEB12668.

Metadata, assemblies, and sequence types for all the isolates in the Burkina Faso collection are available on *Neisseria* PubMLST (<https://pubmlst.org/neisseria/>), as per Kristiansen et al. (2013). Whole-genome sequenced isolates are tagged with their ENA run accessions. Interactive notebooks for recreating analyses are included as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

The authors thank Ingerid Kirkeleite at NIPH for DNA extraction from all the samples for Illumina and Oxford Nanopore sequencing, DNA pipelines at Wellcome Sanger Institute (WSI) for library preparation and sequencing, Pathogen informatics at WSI for systems administration support, and members of teams 81 and 284 at the WSI for helpful comments and advice. This research was funded by the Wellcome Trust (206194 to S.D.B.), the European Research Council (742158 to J.C.), and the Research Council of Norway (220829 to D.A.C.).

Author contributions: N.M. and M.P. carried out the computational analyses. N.M., M.P., O.B., V.E., J.C., D.A.C., and S.D.B.

critically interpreted the results. N.M. wrote the first draft of the manuscript with contributions from O.B., D.A.C., and S.D.B. N.M., M.P., O.B., V.E., J.C., D.A.C., and S.D.B. reviewed and edited the manuscript. The study was designed by O.B., P.A.K., J.C., D.A.C., and S.D.B. Samples were collected by P.A.K. and D.A.C., as well as others cited in Kristiansen et al. (2013).

References

- Acevedo RNV, Ronpirin C, Kandler JL, Shafer WM, Cornelissen CN. 2014. Identification of regulatory elements that control expression of the *tbpa* operon in *Neisseria gonorrhoeae*. *J Bacteriol* **196**: 2762–2774. doi:10.1128/JB.01693-14
- Aho EL, Urwin R, Batcheller AE, Holmgren AM, Havig K, Kulakoski AM, Vomhof EE, Longfors NS, Erickson CB, Anderson ZK, et al. 2005. Neisserial pilin genes display extensive interspecies diversity. *FEMS Microbiol Lett* **249**: 327–334. doi:10.1016/j.femsle.2005.06.035
- Arnold B, Sohail M, Wadsworth C, Corander J, Hanage WP, Sunyaev S, Grad YH. 2020. Fine-scale haplotype structure reveals strong signatures of positive selection in a recombining bacterial pathogen. *Mol Biol Evol* **37**: 417–428. doi:10.1093/molbev/msz225
- Bärnes GK, Brynildsrud OB, Børud B, Workalemahu B, Kristiansen PA, Beyene D, Aseffa A, Caugant DA. 2017. Whole genome sequencing reveals within-host genetic changes in paired meningococcal carriage isolates from Ethiopia. *BMC Genomics* **18**: 407. doi:10.1186/s12864-017-3806-3
- Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffold. *Bioinformatics* **31**: 2443–2451. doi:10.1093/bioinformatics/btv171
- Brynildsrud OB, Eldholm V, Rakhimova A, Kristiansen PA, Caugant DA. 2019. Gauging the epidemic potential of a widely circulating non-invasive meningococcal strain in Africa. *Microb Genom* **5**: e000290. doi:10.1099/mgen.0.000290
- Castillo-Ramírez S, Corander J, Marttinen P, Aldeljawi M, Hanage WP, Westh H, Boye K, Gulay Z, Bentley SD, Parkhill J, et al. 2012. Phylogeographic variation in recombination rates within a global clone of methicillin-resistant *Staphylococcus aureus*. *Genome Biol* **13**: R126. doi:10.1186/gb-2012-13-12-r126
- Christensen H, May M, Bowen L, Hickman M, Trotter CL. 2010. Meningococcal carriage by age: a systematic review and meta-analysis. *Lancet Infect Dis* **10**: 853–861. doi:10.1016/S1473-3099(10)70251-6
- Cooper TF. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol* **5**: e225. doi:10.1371/journal.pbio.0050225
- Craven DE, Frasch CE, Robbins JB, Feldman HA. 1978. Serogroup identification of *Neisseria meningitidis*: comparison of an antiserum agar method with bacterial slide agglutination. *J Clin Microbiol* **7**: 410–414. doi:10.1128/jcm.7.5.410-414.1978
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**: e15. doi:10.1093/nar/gku1196
- Diallo AO, Soeters HM, Yameogo I, Sawadogo G, Aké F, Lingani C, Wang X, Bitá A, Fall A, Sangaré L, et al. 2017. Bacterial meningitis epidemiology and return of *Neisseria meningitidis* serogroup A cases in Burkina Faso in the five years following MenAfriVac mass vaccination campaign. *PLoS One* **12**: e0187466. doi:10.1371/journal.pone.0187466
- Diomandé FV, Djingarey MH, Daugla DM, Novak RT, Kristiansen PA, Collard JM, Gamougam K, Kandolo D, Mbakuliyemo N, Mayer L, et al. 2015. Public health impact after the introduction of PsA-TT: the first 4 years. *Clin Infect Dis* **61**: S467–S472. doi:10.1093/cid/civ499
- Dunn O. 1964. Multiple comparisons using rank sums. *Technometrics* **6**: 241–252. doi:10.1080/00401706.1964.10490181
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797. doi:10.1093/nar/gkh340
- Epping L, Van Tonder AJ, Gladstone RA, The Global Pneumococcal Sequencing Consortium, Bentley SD, Page AJ, Keane JA. 2018. SerBA: rapid high-throughput serotyping of *Streptococcus pneumoniae* from whole genome sequence data. *Microb Genom* **4**: e000186. doi:10.1099/mgen.0.000186
- Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, Del Rio C, Didelot X, Dillon JA, Selden RF, Shafer WM, Turingan RS, et al. 2015. Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. *PeerJ* **3**: e806. doi:10.7717/peerj.806
- Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics* **78**: 737–756. doi:10.1093/genetics/78.2.737
- Harrison OB, Claus H, Jiang Y, Bennett JS, Bratcher HB, Jolley KA, Corton C, Care R, Poolman JT, Zollinger WD, et al. 2013. Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerg Infect Dis* **19**: 566–573. doi:10.3201/eid1904.111799
- Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat* **6**: 65–70.
- Hunter JD. 2007. Matplotlib: a 2D graphics environment. *Comput Sci Eng* **9**: 90–95. doi:10.1109/MCSE.2007.55
- Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**: 90. doi:10.1186/s13073-014-0090-6
- Johnston C, Polard P, Claverys JP. 2013. The DpnI/DpnII pneumococcal system, defense against foreign attack without compromising genetic exchange. *Mobile Genet Elem* **3**: e25582. doi:10.4161/mge.25582
- Joseph B, Schwarz RF, Linke B, Blom J, Becker A, Claus H, Goesmann A, Froesch M, Müller T, Vogel U, et al. 2011. Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome. *PLoS One* **6**: e18441. doi:10.1371/journal.pone.0018441
- Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermini LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587–589. doi:10.1038/nmeth.4285
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780. doi:10.1093/molbev/mst010
- Kim WJ, Higashi D, Goytia M, Rendón MA, Pilligua-Lucas M, Bronnimann M, McLean JA, Duncan J, Trees D, Jerse AE, et al. 2019. Commensal *Neisseria gonorrhoeae* through a DNA-dependent mechanism. *Cell Host Microbe* **26**: 228–239.e8. doi:10.1016/j.chom.2019.07.003
- Koumaré B, Ouedraogo-Traoré R, Sanou I, Yada A, Sow I, Lusamba P, Traoré E, Dabal M, Santamaria M, Hacen MM, et al. 2007. The first large epidemic of meningococcal disease caused by serogroup W135, Burkina Faso, 2002. *Vaccine* **25**: A37–A41. doi:10.1016/j.vaccine.2007.04.038
- Kristiansen PA, Diomandé F, Wei SC, Ouedraogo R, Sangaré L, Sanou I, Kandolo D, Kaboré P, Clark TA, Ouédraogo AS, et al. 2011. Baseline meningococcal carriage in Burkina Faso before the introduction of a meningococcal serogroup A conjugate vaccine. *Clin Vaccine Immunol* **18**: 435–443. doi:10.1128/CVI.00479-10
- Kristiansen PA, Ba AK, Sanou I, Ouédraogo AS, Ouédraogo R, Sangaré L, Diomandé F, Kandolo D, Thomas JD, Clark TA, et al. 2013. Phenotypic and genotypic characterization of meningococcal carriage and disease isolates in Burkina Faso after mass vaccination with a serogroup A conjugate vaccine. *BMC Infect Dis* **13**: 363. doi:10.1186/1471-2334-13-363
- Kristiansen PA, Ba AK, Ouédraogo AS, Sanou I, Ouédraogo R, Sangaré L, Diomandé F, Kandolo D, Saga IM, Misegades L, et al. 2014. Persistent low carriage of serogroup A *Neisseria meningitidis* two years after mass vaccination with the meningococcal conjugate vaccine, MenAfriVac. *BMC Infect Dis* **14**: 663. doi:10.1186/s12879-014-0663-4
- Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc* **47**: 583–621. doi:10.1080/01621459.1952.10483441
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645. doi:10.1101/gr.092759.109
- Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, Corander J, Bentley SD, Croucher NJ. 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* **29**: 304–316. doi:10.1101/gr.241455.118
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993. doi:10.1093/bioinformatics/btr509
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio.GN]. doi:10.6084/M9.FIGSHARE.963153.V1
- Lin Z, Schenker M, Zhu P, Achtman M. 2000. Frequent interspecific genetic exchange between commensal *Neisseriae* and *Neisseria meningitidis*. *Mol Microbiol* **36**: 1049–1058. doi:10.1046/j.1365-2958.2000.01932.x
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci* **95**: 3140–3145. doi:10.1073/pnas.95.6.3140
- Marri PR, Paniscus M, Weyand NJ, Rendón MA, Calton CM, Hernández DR, Higashi DL, Sodergren E, Weinstock GM, Rounsley SD, et al. 2010. Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS One* **5**: e11835. doi:10.1371/journal.pone.0011835

- McKinney W. 2010. Data structures for statistical computing in Python. In *Proceedings of the Ninth Python in Science Conference*, pp. 51–56. Austin, TX. doi:10.25080/Majora-92bf1922-00a
- Molesworth AM, Thomson MC, Connor SJ, Cresswell MP, Morse AP, Shears P, Hart CA, Cuevas LE. 2002. Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa. *Trans R Soc Trop Med Hyg* **96**: 242–249. doi:10.1016/S0035-9203(02)90089-1
- Mostowy R, Croucher NJ, Andam CP, Corander J, Hanage WP, Marttinen P. 2017. Efficient inference of recent and ancestral recombination within bacterial populations. *Mol Biol Evol* **34**: 1167–1182. doi:10.1093/molbev/msx066
- Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a Fast, Unconstrained Bayesian AppRoximation for inferring selection. *Mol Biol Evol* **30**: 1196–1205. doi:10.1093/molbev/mst030
- Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol* **32**: 1365–1371. doi:10.1093/molbev/msv035
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426. doi:10.1093/oxfordjournals.molbev.a040410
- Nelson CW, Moncla LH, Hughes AL. 2015. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**: 3709–3711. doi:10.1093/bioinformatics/btv449
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274. doi:10.1093/molbev/msu300
- Nicolas P, Norheim G, Garnotel E, Djibo S, Caugant DA. 2005. Molecular epidemiology of *Neisseria meningitidis* isolated in the African meningitis belt between 1988 and 2003 shows dominance of sequence type 5 (ST-5) and ST-11 complexes. *J Clin Microbiol* **43**: 5129–5135. doi:10.1128/JCM.43.10.5129-5135.2005
- Obergfell KP, Seifert HS. 2015. Mobile DNA in the pathogenic *Neisseria*. *Microbiol Spectr* **3**: MDNA3-0015-2014. doi:10.1128/microbiolspec.MDNA3-0015-2014
- Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, Otto TD, Keane JA. 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microbial Genomics* **2**: e000083. doi:10.1099/mgen.0.000083
- Perkins-Balding D, Ratliff-Griffin M, Stojiljkovic I. 2004. Iron transport systems in *Neisseria meningitidis*. *Microbiol Mol Biol Rev* **68**: 154–171. doi:10.1128/MMBR.68.1.154-171.2004
- Pond SLK and Muse SV. 2005. *HyPhy*: hypothesis testing using phylogenies. In *Statistical methods in molecular evolution* (ed. Nielsen R), pp. 125–181. Springer, New York.
- Potts CC, Joseph SJ, Chang HY, Chen A, Vuong J, Hu F, Jenkins LT, Schmink S, Blain A, MacNeil JR, et al. 2018. Population structure of invasive *Neisseria meningitidis* in the United States, 2011–15. *J Infect* **77**: 427–434. doi:10.1016/j.jinf.2018.06.008
- R Core Team. 2019. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Retchless AC, Kretz CB, Chang HY, Bazan JA, Abrams AJ, Turner AN, Jenkins LT, Trees DL, Tzeng YL, Stephens DS, et al. 2018. Expansion of a urethritis-associated *Neisseria meningitidis* clade in the United States with concurrent acquisition of *N. gonorrhoeae* alleles. *BMC Genomics* **19**: 176. doi:10.1186/s12864-018-4560-x
- Sahoo N. 2017. “Sequence base-calling through Albacore software: a part of the Oxford Nanopore technology.” *PhD thesis*, Orissa University of Agriculture and Technology, Bhubaneswar, India.
- Schoen C, Tettelin H, Parkhill J, Frosch M. 2009. Genome flexibility in *Neisseria meningitidis*. *Vaccine* **27**: B103–B111. doi:10.1016/j.vaccine.2009.04.064
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069. doi:10.1093/bioinformatics/btu153
- Seib KL, Jen FEC, Scott AL, Tan A, Jennings MP. 2017. Phase variation of DNA methyltransferases and the regulation of virulence and immune evasion in the pathogenic *Neisseria*. *Pathog Dis* **75**: ftx080. doi:10.1093/femspd/ftx080
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* **32**: 1342–1353. doi:10.1093/molbev/msv022
- Spearman C. 1904. The proof and measurement of association between two things. *Am J Psychol* **15**: 72–101. doi:10.2307/1412159
- Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033
- Stephens DS, Greenwood B, Brandtzaeg P. 2007. Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *Lancet* **369**: 2196–2210. doi:10.1016/S0140-6736(07)61016-2
- Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo SW, Beaudoin C, Floto RA, et al. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* **21**. doi:10.1186/s13059-020-02090-4
- Topaz N, Caugant DA, Taha MK, Brynildsrud OB, Debech N, Hong E, Deghmane AE, Ouédraogo R, Ousmane S, Gamougame K, et al. 2019. Phylogenetic relationships and regional spread of meningococcal strains in the meningitis belt, 2011–2016. *EBioMedicine* **41**: 488–496. doi:10.1016/j.ebiom.2019.02.054
- Trotter CL, Greenwood BM. 2007. Meningococcal carriage in the African meningitis belt. *Lancet Infect Dis* **7**: 797–803. doi:10.1016/S1473-3099(07)70288-8
- Trotter CL, Lingani C, Fernandez K, Cooper LV, Bita A, Tevi-Benissan C, Ronveaux O, Préziosi MP, Stuart JM. 2017. Impact of MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of surveillance data. *Lancet Infect Dis* **17**: 867–872. doi:10.1016/S1473-3099(17)30301-8
- Vigué L, Eyre-Walker A. 2019. The comparative population genetics of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *PeerJ* **7**: e7216. doi:10.7717/peerj.7216
- Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, et al. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**: 261–272. doi:10.1038/s41592-019-0686-2
- Wachter J, Hill S. 2016. Positive selection pressure drives variation on the surface-exposed variable proteins of the pathogenic *Neisseria*. *PLoS One* **11**: e0161348. doi:10.1371/journal.pone.0161348
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017a. Completing bacterial genome assemblies with multiplex minion sequencing. *Microb Genomics* **3**: e000132. doi:10.1099/mgen.0.000132
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017b. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**: e1005595. doi:10.1371/journal.pcbi.1005595
- Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* **15**: R46. doi:10.1186/gb-2014-15-3-r46
- Yu D, Jin Y, Yin Z, Ren H, Zhou W, Liang L, Yue J. 2014. A genome-wide identification of genes undergoing recombination and positive selection in *Neisseria*. *BioMed Res Int* **2014**: 815672. doi:10.1155/2014/815672
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36. doi:10.1111/2041-210X.12628
- Zhu P, Morelli G, Achtman M. 1999. The *opa* and *ψopb* regions in *Neisseria*: genes, pseudogenes, deletions, insertion elements and DNA islands. *Mol Microbiol* **33**: 635–650. doi:10.1046/j.1365-2958.1999.01514.x

Received April 8, 2020; accepted in revised form April 22, 2021.



The effect of recombination on the evolution of a population of *Neisseria meningitidis*

Neil MacAlasdair, Maiju Pesonen, Ola Brynildsrud, et al.

Genome Res. 2021 31: 1258-1268 originally published online June 9, 2021

Access the most recent version at doi:[10.1101/gr.264465.120](https://doi.org/10.1101/gr.264465.120)

Supplemental Material

<http://genome.cshlp.org/content/suppl/2021/06/09/gr.264465.120.DC1>

References

This article cites 73 articles, 9 of which can be accessed free at:
<http://genome.cshlp.org/content/31/7/1258.full.html#ref-list-1>

Open Access

Freely available online through the *Genome Research* Open Access option.

Creative Commons License

This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>