

RESEARCH ARTICLE

# Tracking a serial killer: Integrating phylogenetic relationships, epidemiology, and geography for two invasive meningococcal disease outbreaks

Ifeoma Ezeoke<sup>1</sup>, Madeline R. Galac<sup>2</sup>, Ying Lin<sup>3</sup>, Alvin T. Liem<sup>4,5</sup>, Pierce A. Roth<sup>4,5</sup>, Andrew Kilianski<sup>2</sup>, Henry S. Gibbons<sup>4</sup>, Danielle Bloch<sup>1</sup>, John Kornblum<sup>3</sup>, Paula Del Rosso<sup>1</sup>, Daniel A. Janies<sup>2</sup>, Don Weiss<sup>1\*</sup>

**1** Bureau of Communicable Disease, Department of Health and Mental Hygiene, New York, NY, United States of America, **2** Department of Bioinformatics and Genomics, University of North Carolina, Charlotte, NC, United States of America, **3** Bureau of Public Health Laboratory, Department of Health and Mental Hygiene, New York, NY, United States of America, **4** Department of Microbiology, US Army Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD, United States of America, **5** DCS Corporation, Alexandria, VA, United States of America

☉ These authors contributed equally to this work.

\* [DWeiss@health.nyc.gov](mailto:DWeiss@health.nyc.gov)



**OPEN ACCESS**

**Citation:** Ezeoke I, Galac MR, Lin Y, Liem AT, Roth PA, Kilianski A, et al. (2018) Tracking a serial killer: Integrating phylogenetic relationships, epidemiology, and geography for two invasive meningococcal disease outbreaks. *PLoS ONE* 13 (11): e0202615. <https://doi.org/10.1371/journal.pone.0202615>

**Editor:** Baochuan Lin, Defense Threat Reduction Agency, UNITED STATES

**Received:** August 1, 2018

**Accepted:** November 5, 2018

**Published:** November 28, 2018

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** Genbank accession numbers are listed in the manuscript. The New York City Department of Health and Mental Hygiene surveillance data are subject to legal restrictions protecting personally identifiable health information. Researchers may request this data from the following contact: [BCD\\_syndromic\\_data@health.nyc.gov](mailto:BCD_syndromic_data@health.nyc.gov).

**Funding:** D.W., Y.L. and J.S.K. were supported by New York City tax levy funding, I.E. [grant number

## Abstract

### Background

While overall rates of meningococcal disease have been declining in the United States for the past several decades, New York City (NYC) has experienced two serogroup C meningococcal disease outbreaks in 2005–2006 and in 2010–2013. The outbreaks were centered within drug use and sexual networks, were difficult to control, and required vaccine campaigns.

### Methods

Whole Genome Sequencing (WGS) was used to analyze preserved meningococcal isolates collected before and during the two outbreaks. We integrated and analyzed epidemiologic, geographic, and genomic data to better understand transmission networks among patients. Betweenness centrality was used as a metric to understand the most important geographic nodes in the transmission networks. Comparative genomics was used to identify genes associated with the outbreaks.

### Results

*Neisseria meningitidis* serogroup C (ST11/ET-37) was responsible for both outbreaks with each outbreak having distinct phylogenetic clusters. WGS did identify some misclassifications of isolates that were more distant from the outbreak strains, as well as those that should have been included based on high genomic similarity. Genomes for the second outbreak were more similar than the first and no polymorphism was found to either be unique or specific to either outbreak lineage. Betweenness centrality as applied to transmission

3U50CK000407-01] and D.B. [grant number NU50CK000407-03] were supported by the Centers for Disease Control and Prevention Epidemiology and Laboratory Capacity Grant. P.D. R. was supported by the Centers for Disease Control and Prevention Public Health Emergency Preparedness Cooperative Agreement [grant number NU90TP921922-01-00]. A.K. was supported by the US National Academy of Science and the US Defense Threat Reduction Agency (DTRA project CB2047) as a National Research Council (NRC) fellow. A.T.L. and P.A.R. were supported by the United States Army via a contract with DCS Corporation. The salaries of A.T.L. and P.A.R. were paid by the US Army. DCS Corporation, nor any other funder, did not have any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the PLOS online author contributions section.

**Competing interests:** All authors have no competing interests to declare. After the project data collection and analysis was completed, during the final phase of manuscript editing, D.A.J. was appointed to the board of the North Carolina Biotechnology Center. D.A.J. reaffirmed that he does not have any competing interests. The affiliation with DCS [A.T.L. and P.A.R.] does not alter our adherence to PLOS ONE policies on sharing data and materials.

networks based on phylogenetic analysis demonstrated that the outbreaks were transmitted within focal communities in NYC with few transmission events to other locations.

## Conclusions

*Neisseria meningitidis* is an ever changing pathogen and comparative genomic analyses can help elucidate how it spreads geographically to facilitate targeted interventions to interrupt transmission.

## Background

The incidence of meningococcal disease in the United States has declined steadily since 1996 [1]. Against this backdrop, New York City (NYC) experienced 2 outbreaks of invasive meningococcal disease (IMD) caused by serogroup C *Neisseria meningitidis* (MenC) that prompted vaccination interventions [2, 3]. The first NYC MenC outbreak (Ob1) occurred in the years 2005–2006 and affected current and former drug users and their close contacts [2]. The second community outbreak of IMD (Ob2) began four years later and affected 22 men who have sex with men (MSM) [3]. Previous outbreaks of IMD among MSM were seen in the early 2000s in Toronto [4] and Chicago [5] and after the NYC outbreak, clusters of IMD in MSM were recognized in Southern California [6], Chicago [7], Germany [8], France [9], and Spain [10].

*N. meningitidis* (Nm) is found in the nasopharynx in up to 35% of healthy adults [11] and it is this colonization state that is believed to be responsible for IMD transmission and outbreaks. Secondary IMD cases are rare and outbreaks occur in either organizational settings, such as universities, or in members of discreet community groups, such as MSM. Pulsed-field gel electrophoresis (PFGE) has been used in NYC to detect links among IMD cases not identified by case investigations [12]. However, PFGE may misclassify due to a lack of discriminatory power [13]. Additionally, PFGE does not provide details on polymorphisms in the Nm genome that can be used to track its spread in space and time.

Advances in whole genome sequencing (WGS) technology have allowed epidemiologists to uncover disease transmission patterns through phylogenetic analyses. WGS has successfully been used to inform outbreaks caused by tuberculosis [14], cholera [15], *Klebsiella* [16], *Legionella* [17], and *Listeria* [18]. In an IMD outbreak associated with the 23rd World Scout Jamboree held in Japan, WGS was used to elucidate relationships among cases in Scotland and Sweden [19]. Also, WGS has led to recent reports suggesting that *N. meningitidis* has adapted to become a urogenital pathogen. Taha et al. sequenced *N. meningitidis* isolates from the 2012–13 MSM outbreak in Germany and France and several antecedent rectal and genitourinary cultures [20]. They found evidence of an in-frame *aniA* gene in MenC similar to that present in *Neisseria gonorrhoeae* that confers the ability to survive under anaerobic conditions [20]. Following a large cluster of *N. meningitidis* urethritis in heterosexual men in Ohio, U.S., WGS identified an insertion sequence that had replaced much of the *cps* capsular locus of a MenC strain, rendering it unencapsulated and nongroupable [21,22].

We sequenced a collection of archived *N. meningitidis* isolates' genomes from NYC from 2003–2013 and analyzed phylogenetic relationships of the isolates based on core genome alignment, patient risk factors, and demographics to explore spatial connections among cases not evident from the epidemiologic data alone. The study was undertaken to better understand how IMD transmission occurs within a large urban center with the goal of improving IMD outbreak response, prevention, and control.

## Materials and methods

### Study population and case investigation

IMD is a reportable condition in NYC. Medical providers and laboratories are required to report suspected and confirmed cases within 24 hours as defined by the Council of State and Territorial Epidemiologists (CSTE)/Centers for Disease Control and Prevention's (CDC) case definition [23]. All reports are investigated to confirm the diagnosis, identify close contacts who qualify for post-exposure prophylaxis, and to uncover links that could indicate an outbreak. Sterile site isolates are sent to the NYC Public Health Laboratory for serogrouping and molecular typing. A significant proportion of IMD cases have negative cultures [24] and beginning in 2006, clinical samples were routinely sought on culture-negative cases for RT-PCR testing at the NY State Wadsworth Center [25]. Non-sterile *N. meningitidis* reports (pharyngitis, urethritis, proctitis) were occasionally received and processed as above on a case-by-case basis. Isolates from non-NYC residents who had an epidemiologic link to NYC were sought for molecular characterization and comparison. Ob1-associated cases were defined as those meeting the CSTE/CDC case definition with *N. meningitidis* serogroup C and PFGE findings of  $\geq 85\%$  similarity to the designated outbreak strain, or for probable cases (culture negative), patients who were household contacts to or shared an epidemiologic link with an outbreak-associated case [2]. Ob2-associated cases were defined as confirmed or probable *N. meningitidis* serogroup C MSM NYC residents with onset during August 2010–February 2013 [3].

IMD cases were investigated using a structured data collection instrument. Death from IMD was defined as occurring within 30 days of diagnosis and was obtained through a data match with the NYC Office of Vital Statistics death registry; HIV status was determined through matching with the NYC HIV Surveillance Registry. The earliest available MenC isolate was from 2003, therefore, isolates from 2003–2013 were eligible for inclusion in the study. A selection of other serogroups, non-NYC residents, and non-sterile sources were included for comparison.

### Isolate regrowth, DNA extraction, sequencing, assembly, and annotation

*N. meningitidis* isolates were preserved in Trypticase soy broth with 20% glycol at  $-70^{\circ}\text{C}$ . Isolates were removed from storage and plated on Chocolate II Agar (BBL). Following overnight growth at  $35^{\circ}$  in a humidified 5%  $\text{CO}_2$  incubator, single colonies were selected to inoculate a subculture. DNA was extracted from fresh *N. meningitidis* subculture colonies using Genomic-tip 20/G (Qiagen). Post-extraction quality control was performed to ensure concentration, purity, integrity, and sterility. WGS was performed using the HiSeq 2000 platform (Illumina) at the Edgewood Chemical and Biological Center (ECBC) sequencing facility. DNA was prepared for sequencing using the Nextera DNA Library Preparation kit (Illumina) and paired-end sequencing was performed on the dual-indexed isolate libraries (Illumina) according to the manufacturer's protocol. After sequence capture, reads were assembled *de novo* using either Velvet (1.2.10) [26] for higher coverage read files ( $>200\text{Mb}$ ) or CLC Genomics Workbench (6.5.1) (Qiagen) for lower coverage read files ( $<200\text{Mb}$ ). Consensus genomes were then annotated using RAST toolkit (1.3.0) [27, 28]. Sequencing personnel were blinded to the epidemiological and geographic metadata.

### Determining orthologs, core genome, and pan genome

Ortholog clusters were determined based on the comparison of all protein coding sequences with OrthoMCL (version 2.0.9) [29, 30] with an e-value cut off of  $10^{-5}$  on the BLAST+ (version 2.27) [31]. The output was then parsed with custom Python scripts to find the core genome for

different sets of isolates and genes specific to only isolates with particular metadata characteristics. The ortholog analysis also included 10 additional finished *N. meningitidis* genomes from GenBank and one draft genome from Paris reported to be a part of the MSM outbreak [32]. Assembled genomes were queried using BLAST against the PubMLST *Neisseria* database (<https://pubmlst.org/neisseria/>) to identify *N. meningitidis* clonal complex and the presence of a frameshift mutation in the *aniA* gene [33].

## Phylogeny

The phylogeny of the genomes included in the OrthoMCL analysis was determined based on the core genome. We used a strict definition of core genome as being those ortholog clusters that contain only a single gene from each genome, excluding clusters with any paralogs. No epidemiological data were taken into consideration for determining orthologs or the initial phylogeny. The DNA sequences for each cluster were individually aligned using MAFFT (version 7.130b) [34], then Trimal (version 1.2) [35] was used to obtain the mean percent identity from each individual alignment. Alignments with a mean percent identity less than 90% were inspected individually and those with poor alignments, suggesting that they may not be orthologs, were removed from further analysis. The remaining alignments were concatenated into one core genome alignment with FASconCAT (version 1.02) [36] to be used by RAxML (version 8) [37] to determine phylogenetic relationships using maximum likelihood optimality criterion. The same methods were used for determining the core genomes of a subset of genomes for analyses (i.e., Genome Wide Association Studies and Betweenness centrality (see below)).

## Comparison of isolates from the same patient

The OrthoMCL output was parsed to determine the core and unique genes for each set of isolates from the same patient. The strict core genome clusters were aligned individually using MAFFT [34] then a custom python script was used to find SNPs in each of the alignments. A SNP was declared only in instances where there was known nucleotide differences, not in cases where there was a gap or N. Alignments with high numbers of SNPs (greater than 10 for the gene) were examined by eye to ensure the number of SNPs was not due to poor sequence alignment.

## Genome wide association studies

Genome Wide Association Studies (GWAS) were done on two different input data types: (1) gene present or absent in a genome; (2) single nucleotide polymorphisms (SNPs) in the core genome. Custom Python scripts were written to parse the OrthoMCL output to determine if a gene was present or absent in each genome based on whether there was a gene listed within orthologous clusters and to produce PED and MAP files. Therefore, the number of genes considered is equivalent to the total number of gene clusters determined by OrthoMCL, 5,280 gene clusters. In the PED file, one base was assigned if the gene was present and a different base used if the gene was absent for that genome for each orthologous cluster. Custom python scripts were written to find SNPs in alignments made by MAFFT for each individual gene cluster of the strict core genome and to produce PED and MAP files. Those alignment locations that contained more than two SNP variations were excluded due to limitations in the GWAS software, but the presence of an N was included as missing data for that genome. The program PLINK (version 1.07)[38] was then used to perform the GWAS with the genomic location set to the haploid mitochondria in the input files. P-values were adjusted by PLINK for population stratification using a genomic control and for multiple testing using a conservative Bonferroni correction. Gene Ontology (GO) terms were determined by BLAST2GO [39].

## Betweenness centrality

For the betweenness centrality analysis we used methods similar to those used in Janies et al. [40]. Betweenness measures the centrality of a node by computing how many times a node of interest is on the shortest paths between any two other nodes normalized by the number of pairs of other nodes. In our analyses, we used the patients' borough and neighborhood of residence as nodes. High betweenness centrality indicates that the geographic location is important in the traffic of the pathogen. We used the subset of isolates with home addresses that could be geocoded plus an isolate from Paris, LNP27256 (GenBank: AVOU01000001.1) related to the outbreak [33]. When multiple isolates originated from the same patient, one isolate was selected as a representative. We then used the phylogenetic methods describe above to produce a tree based on the strict core genome for only these isolates [41]. When RAxML was run on the concatenated core genome, the earliest isolate, S00-1196, was used as an outgroup. This tree was then brought into PAUP\* (version 4.0a147) with the cases' home borough or neighborhood as units of geography. Borough was determined from the latitude and longitude of the home address using ArcGIS (<http://www.esri.com/software/arcgis/explorer/index.html>). New York City neighborhood was determined based on NYC Department of City Planning Neighborhood Tabulation Areas (NTA, Table A in [S1 File](#)). Isolates from outside NYC maintained their individual location for both analyses. PAUP\* was then used to calculate the ancestor descendent changes in geographic location at each branch of the phylogenetic tree. The directionality and number of each type of change was then measured. These data were then integrated to form the network of bacterial transmission using igraph and qgraph packages in R [42, 43].

## Ethical statement

The Department of Health and Mental Hygiene (DOHMH) Institutional Review Board determined the study to be public health surveillance that is non-research.

## Results

### Study population and isolate sequencing

A total of 102 isolates were submitted for WGS. Six isolates were excluded due to poor sequencing or assembly of their genomes and there were 11 repeat isolates (same patient, different specimen source) from 10 individuals leaving 85 unique patients (75 NYC residents and 64 invasive infections) in the analysis ([S1 Fig](#)). During the years 2003–2013, a total of 322 IMD cases were reported in NYC residents (incidence 0.4 per 100,000), of which 263 were culture positive (the universe of possible study isolates). The proportion of eligible invasive isolates in NYC residents included in the study by serogroup is as follows: 5/60 (8%) MenB, 52/86 (60%) MenC, 4/22 (18%) MenW, 3/89 (3%) MenY, and 0/6 (0%) nongroupable. Fourteen MenC patient isolates were included from each outbreak (Ob1 had 21 and Ob2 had 15 culture positive cases). The demographic characteristics of invasive MenC isolates from NYC residents included in the study were compared to MenC patients during the same interval whose isolates were unavailable (includes culture negative cases that met the case definition). Patients with MenC included in the study were more likely to be older, male and non-Hispanic ([Table 1](#)).

Our assembled draft genomes had between 218 and 1223 contigs with a mean contig number of 494 and read depth coverage ranging from 300 to 3000 ([Table B in S1 File](#)). The total genome length and gene count of these contigs are all within the expected size for *N. meningitidis*. The strict core genome, which explicitly left out orthologous clusters containing paralogs, was determined for the 96 genomes sequenced and 11 genomes from GenBank was found to



**Table 1. Characteristics of sequenced invasive MenC isolates compared to those that were unavailable (confirmed and probable IMD cases), New York City, 2003–2013.**

| Characteristic                 | Not Sequenced (n = 49) <sup>1</sup> | Sequenced (n = 52) |
|--------------------------------|-------------------------------------|--------------------|
| Age group                      |                                     |                    |
| 0–4 y                          | 3 (6.1)                             | 2 (3.9)            |
| 5–14 y                         | 3 (6.1)                             | 5 (9.6)            |
| 15–24 y                        | 11 (22.5)                           | 4 (7.7)            |
| 25–44 y                        | 20 (40.8)                           | 20 (38.5)          |
| 45–64 y                        | 7 (14.3)                            | 14 (26.9)          |
| ≥65 y                          | 5 (10.2)                            | 7 (13.5)           |
| Sex                            |                                     |                    |
| Female                         | 23 (46.9)                           | 19 (36.5)          |
| Male                           | 26 (53.1)                           | 33 (63.5)          |
| Race/ethnicity                 |                                     |                    |
| Non-Hispanic white             | 7 (14.3)                            | 14 (26.9)          |
| Non-Hispanic black             | 16 (32.7)                           | 23 (44.2)          |
| Hispanic                       | 16 (32.7)                           | 9 (17.3)           |
| Asian/Pacific Islander         | 2 (4.1)                             | 2 (3.9)            |
| Unknown/Other                  | 8 (16.3)                            | 4 (7.7)            |
| Borough of residence           |                                     |                    |
| Bronx                          | 7 (14.3)                            | 7 (13.5)           |
| Brooklyn                       | 27 (55.1)                           | 26 (50.0)          |
| Manhattan                      | 9 (18.3)                            | 13 (25.0)          |
| Queens                         | 5 (10.2)                            | 5 (9.6)            |
| Staten Island                  | 1 (2.0)                             | 1 (1.9)            |
| HIV                            | 12 (24.5)                           | 10 (19.2)          |
| MSM (Males, 2011–2013, age≥15) | 8/9 (88.9)                          | 13/15 (86.7)       |
| Died                           | 12 (24.5)                           | 22 (42.3)          |
| Outbreak 1 (2005–2006)         | 9/23 (39.1)                         | 14/23 (60.9)       |
| Outbreak 2 (2010–2013)         | 7/22 (31.8)                         | 14/22 (63.6)       |

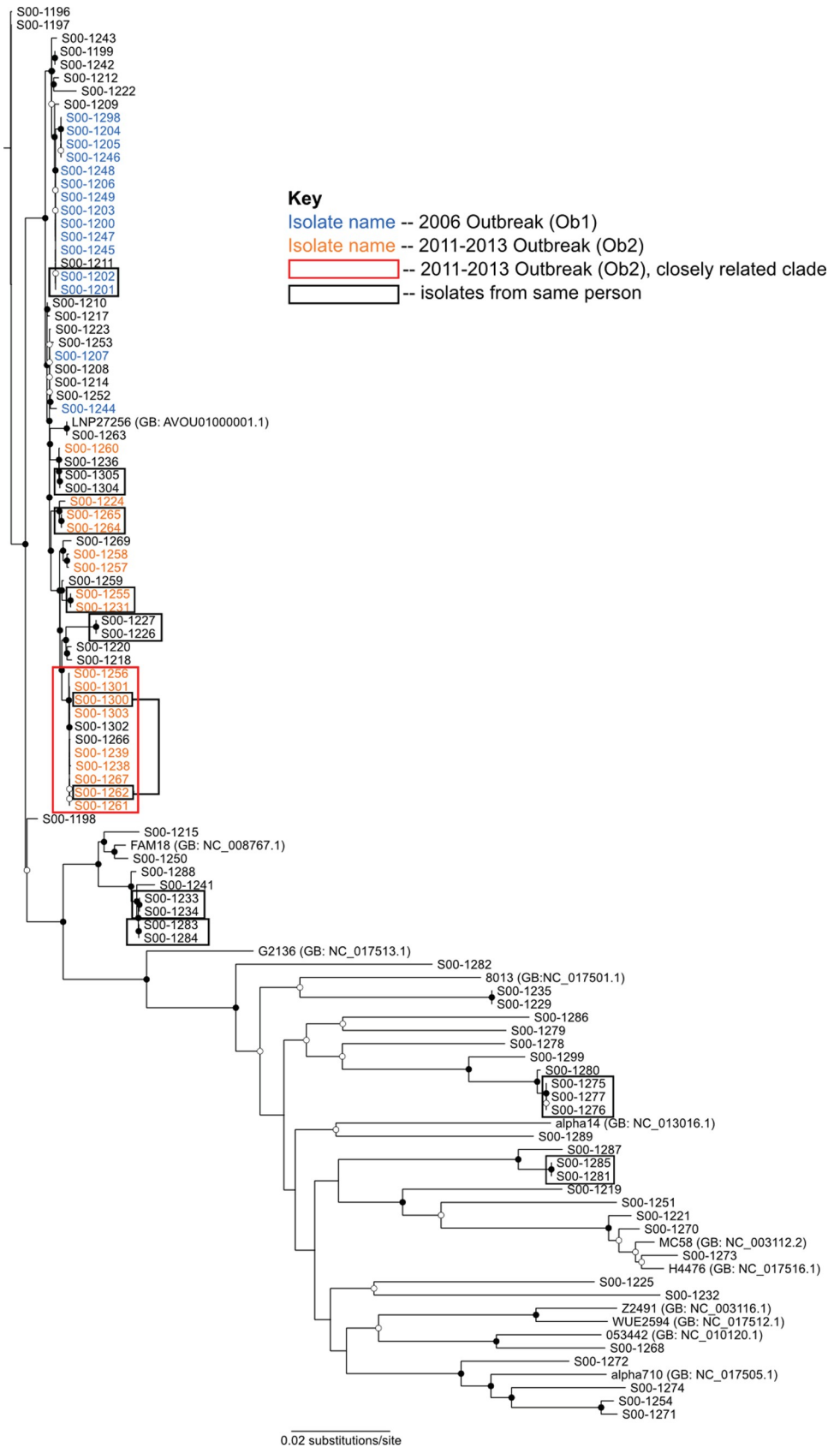
<sup>1</sup> Includes 15 culture negative, PCR positive cases

<https://doi.org/10.1371/journal.pone.0202615.t001>

contain 1,171 genes. When only the genomes sequenced in this study were considered for the core genome, it contains 1,373 genes. Since these were draft genomes, it was suggested that a better metric for determining the core genome would be to include gene clusters found in 95% of the genomes being analyzed [44]. Using this metric, the size of the core genome increased to 1,348 genes for 107 genomes and 1,545 genes for only the genomes sequenced in this study. The number of unique genes ranged from 7 to 376 (mean 33) per isolate (Table B in S1 File). The estimated *N. meningitidis* pan genome based on these isolates contained 8,453 genes.

### Core genome phylogeny

The phylogenetic relationships among the isolates was determined with maximum likelihood tree searches based on alignments of the strict core genome. From these analyses we determined that Ob1 isolates classified by epidemiology were mostly found on a single clade, with S00-1207 and S00-1244 as the exceptions (Fig 1, blue text). Likewise, the 14 Ob2 isolates mostly formed a separate, single clade, with only S00-1260 found on a different clade (Fig 1, orange text). One Ob2 subclade with dates ranging from August 2011 to February 2013 showed less diversity (Fig 1, 11 isolates, outlined in red) and included two patients who were not NYC



**Fig 1. Phylogenetic relationship of Nm isolates from around NYC 2003–2013.** The tree is based on core genome SNPs and rooted on the oldest isolate. Open circles represent nodes with bootstrap support  $\geq 50$ , while closed circles represent nodes with bootstrap support  $\geq 95$  based on 100 replicates. Those isolates that are part of the 2006 (Ob1) based on epidemiology outbreak have blue text. Those isolates that are part of the 2012 outbreak (Ob2) based on epidemiology have orange text, with the closely related subclade indicated by a red box. Isolates from the same individuals are indicated by black boxes.

<https://doi.org/10.1371/journal.pone.0202615.g001>

residents (S00-1266 and S00-1302). All of the patients in the subclade were male between the ages of 23 and 53 years; and of the nine NYC residents: seven were HIV-infected, three died, and seven reported drug use (methamphetamine, cocaine, and/or marijuana). The subclade contained 150 SNPs within the core genome, compared to 4,921 core genome SNPs in the entire outbreak, indicating that the isolates in the subclade were more closely related to one another than to the rest of the outbreak isolates. Of the eight isolates found within Ob2's clade that were not declared to be part of this outbreak based on epidemiology, six (S00-1269, S00-1259, S00-1227, S00-1226, S00-1302, and S00-1266) were collected during outbreak's time period, with S00-1220 and S00-1218 as the exceptions collected in 2008, before Ob2 began (Table 2 and Fig 1). Non-resident patients S00-1266, S00-1269, and S00-1302 were not known to be MSM and were not considered part of Ob2 at the time of collection. Due to the isolates' position in the tree, they may have shared a common exposure with outbreak cases or their social network. Isolate S00-1259 was obtained from a 65 year-old woman in March 2012. She reported no epidemiologic link to the outbreak and was not included as an Ob2 case. While

**Table 2. Comparative genomics of isolates from the same patient.**

| Repeat ID | Study ID | Year | Serogroup | Core Genome | Unique Genes | Genes with SNPs | Total SNPs in All Genes |
|-----------|----------|------|-----------|-------------|--------------|-----------------|-------------------------|
| 1         | S00-1201 | 2006 | C         | 2422        | 116          | 17              | 94                      |
| 1         | S00-1202 | 2006 | C         | 2422        | 124          | 17              | 94                      |
| 2         | S00-1262 | 2012 | C         | 2326        | 162          | 21              | 72                      |
| 2         | S00-1300 | 2012 | C         | 2326        | 183          | 21              | 72                      |
| 3         | S00-1264 | 2012 | C         | 2382        | 169          | 27              | 279                     |
| 3         | S00-1265 | 2012 | C         | 2382        | 143          | 27              | 279                     |
| 4         | S00-1226 | 2010 | C         | 2298        | 225          | 7               | 36                      |
| 4         | S00-1227 | 2010 | C         | 2298        | 188          | 7               | 36                      |
| 5         | S00-1231 | 2011 | C         | 2226        | 267          | 54              | 239                     |
| 5         | S00-1255 | 2011 | C         | 2226        | 266          | 54              | 239                     |
| 6         | S00-1304 | 2013 | C         | 2398        | 156          | 23              | 196                     |
| 6         | S00-1305 | 2013 | C         | 2398        | 202          | 23              | 196                     |
| 7         | S00-1275 | 2010 | Y         | 2224        | 166          | 28              | 76                      |
| 7         | S00-1276 | 2010 | Y         | 2224        | 177          | 28              | 76                      |
| 7         | S00-1277 | 2010 | Y         | 2224        | 169          | 28              | 76                      |
| 8         | S00-1233 | 2011 | C         | 2349        | 184          | 16              | 87                      |
| 8         | S00-1234 | 2011 | C         | 2349        | 182          | 16              | 87                      |
| 9         | S00-1281 | 2010 | W         | 2238        | 142          | 22              | 297                     |
| 9         | S00-1285 | 2010 | W         | 2238        | 213          | 22              | 297                     |
| 10        | S00-1283 | 2009 | W         | 2391        | 99           | 19              | 145                     |
| 10        | S00-1284 | 2009 | W         | 2391        | 130          | 19              | 145                     |
| Mean      |          |      |           | 2321        | 174          | 12              | 148                     |
| Std. Dev. |          |      |           | 75          | 44           | 23              | 91                      |

<https://doi.org/10.1371/journal.pone.0202615.t002>



the isolate's position on the tree suggests inclusion as part of Ob2, it is also possible that the case represents sporadic transmission.

### Clonal complexes and in-frame *aniA* gene

The majority (78%, 66/85) of unique patient isolates included in the study were MenC and 85% (56/66) of these were ST-11/ET-37 complex. Two MenW isolates also belonged to the ST-11/ET-37 complex. No other clonal complex accounted for more than 4% of isolates. For 6 isolates the clonal complex could not be determined.

Fifty-seven percent (55/96) of isolates and 58% of patients (49/85) contained an in-frame *aniA* gene coding for nitrite reductase. All multiple isolates were concordant. The earliest appearance of an in-frame *aniA* was in 2003 and the proportion by unique patient increased over the interval from 45% (17/38) in the years 2003–2008 to 68% (32/47) in the years 2009–2013. The proportion of isolates with an intact *aniA* gene increased from 14% (2/14) patients for Ob1 to 86% (12/14) for Ob2. All 11 Ob2 subclade isolates contained the in-frame *aniA* gene. Data on clonal complex and *aniA* gene are included in the provided data set.

### Comparison of isolates from the same person

Nine patients had two isolates collected from blood and cerebrospinal fluid (CSF), and 1 patient had 2 blood isolates with different collection dates and 1 CSF isolate (S00-1275, S00-1276, S00-1277). One additional patient isolate, obtained from the rectum, had two colony morphologies on regrowth and both isolates were sequenced (S00-1233, S00-1234). Phylogenetic analysis showed that isolates from the same patient were more similar to one another than they are to isolates from other patients (Fig 1).

The strict core genome for each set of isolates from the same patient consisted of an average of 2320 (stdev  $\pm 74$ ) genes (Table 2). When comparing the isolates to the other(s) from the same patient, each isolate had an average of 174 (stdev  $\pm 43$ ) genes that were not found in the other(s). The alignments of the core genome were examined for the presence of SNPs. Each group had an average of 148 (stdev  $\pm 90$ ) SNP across the entire core genome. If a gene cluster contained SNPs, it was usually less than five SNPs. The few genes that did contain more than 10 SNPs were those that are known to have a high degree of variation in pilus related genes. No gene cluster had SNPs for every group of isolates.

### Genome wide association studies

Genome Wide Association Studies were performed using only the genomes sequenced for the study using two metrics: (1) the presence or absence of genes; and (2) the presence of SNPs in the strict core genome. Due to software limitations, only SNPs with two variants could be considered. We examined 82,597 SNPs, while having to exclude 3,541 SNP locations from the analysis. We tested for the association of SNPs or genes to both Ob1 and Ob2 (Table B in S1 File). The analysis for associations with Ob2 excluded S00-1260 due to its phylogenetic distance from the rest of the Ob2 isolates.

We found the presence of five genes and the absence of one to be significantly associated with Ob1 ( $p < 0.05$ , Table B in S1 File). Ob1 also had 30 SNPs significantly associated with it ( $p < 0.05$ , Table B in S1 File). GO terms indicated that these, mostly synonymous, SNPs are found in genes of a wide variety functions.

Ob2 isolates have one gene encoding a protein associated with the outbreak clade which was annotated as a hypothetical protein ( $p = 0.0193$ , Table B in S1 File). One SNP in a TonB-dependent receptor was found to be significantly associated with Ob2 ( $p = 0.04994$ , Table B in S1 File). This SNP is non-synonymous and results in a change from threonine to isoleucine.

All 8 isolates that were not epidemiologically considered part of Ob2, but were present on the same clade also possess this change. Through a BLAST search, we found that this SNP is not unique to this clade of *N. meningitidis*.

### Betweenness centrality

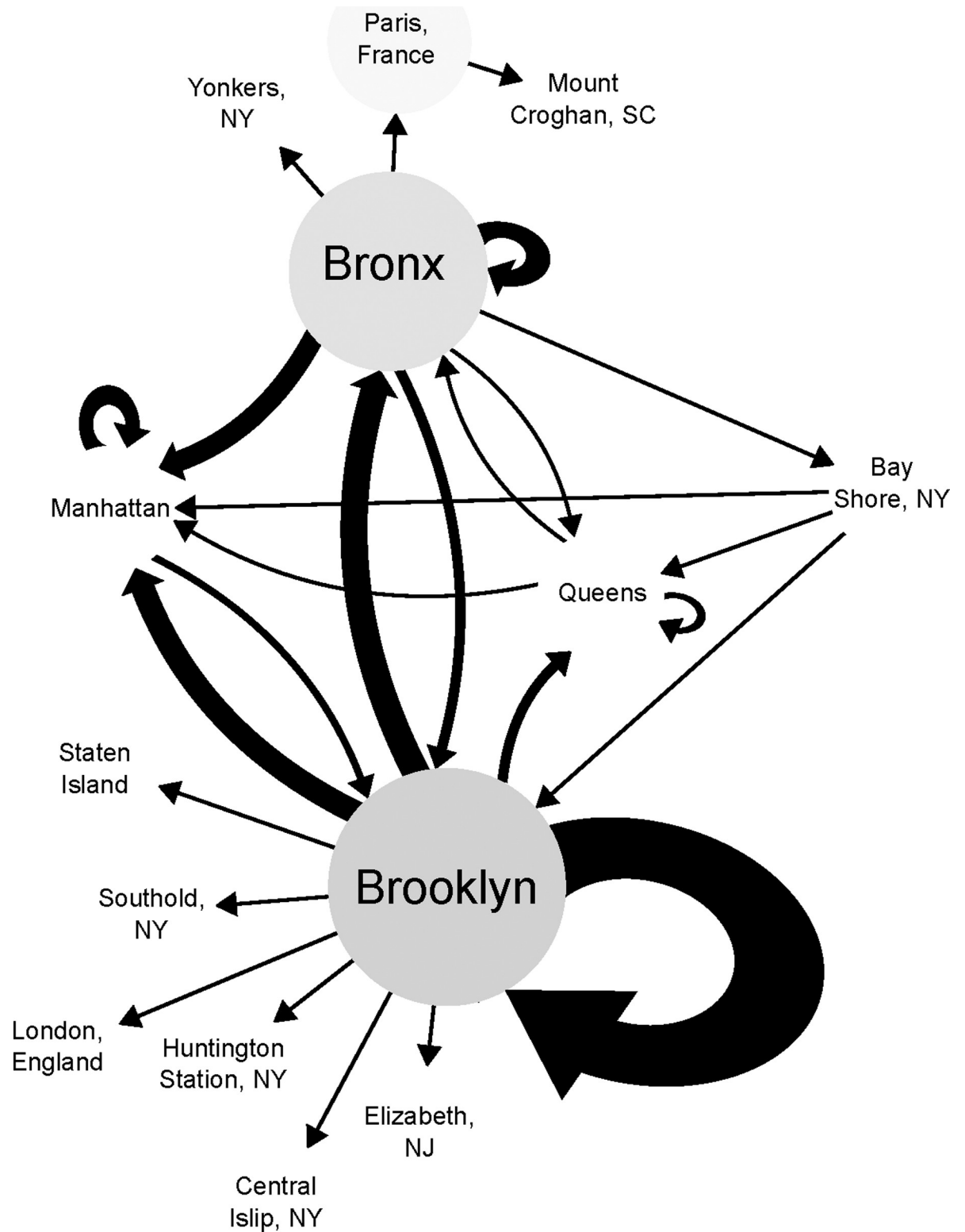
As detailed in the methods section, we optimized the home location of the patients on the phylogeny and converted the result of all the ancestor descendent changes to a network to determine how *N. meningitidis* moved around the city over the 11-year period. We then applied betweenness centrality as a measure the importance of a place in the traffic of the pathogen across geography. Seventy-eight unique patient isolates (92%), with a strict core genome of 1320 genes, were included in the betweenness centrality analysis. At the NYC borough resolution, not all of the boroughs were connected to each other nor were all the connections bidirectional (Fig 2). Brooklyn had the highest betweenness centrality as the only borough connected to all other boroughs and the one with the most connections to non-NYC locations. Brooklyn also had the most intra-borough transmission events (Fig 2, indicated by the arrow that loops back). The Bronx had the second highest betweenness centrality of the boroughs. The Bronx was connected to all the boroughs except Staten Island. We observed bidirectional movement between the Bronx, Brooklyn and Queens, and connections to three non-NYC locations.

While betweenness centrality at the level of the borough allowed us to see an additive summary of *N. meningitidis* movement over 11 years, it did not elucidate the details of transmission events for the two outbreaks. When the betweenness centrality analysis was examined at the level of neighborhood, the arrows represented much more granular transmission events between locations elucidating the transmissions specific to each outbreak (Fig 3). Most transmission events at the level of neighborhood were single events. However, there were some multiple transmissions between or within locations: (1) two transmissions occurred between Stuyvesant Heights (BK35) and Bedford (BK75); (2) three transmissions occurred within Stuyvesant Heights (BK35); and (3) two transmissions occurred within Crown Heights North (BK61) (Fig 3). All three of these multiple transmission events were from Ob1 and only two non-Brooklyn neighborhoods were involved with spread which establishes the geographic concentrations of this outbreak in Brooklyn (Fig 3, colored blue). The two non-Brooklyn neighborhoods involved in Ob1 were the Bronx neighborhood of Co-Op City (BX13) and the Manhattan neighborhood of Carnegie Hill (MN40). The transmission to Co-Op City (BX13) was a dead end for further transmission (Fig 3, colored blue). From Carnegie Hill (MN40), two transmissions occurred to two different Brooklyn neighborhoods that were disconnected from the rest of the Ob1 outbreak (Fig 3, colored blue).

For Ob2 Midtown/Midtown South (MN17) is the most important neighborhood for transmission (Fig 3, colored red and orange). The closely related Ob2 clade originates in this neighborhood before moving to Brownsville (BK81) from which the transmission fans out to multiple boroughs.

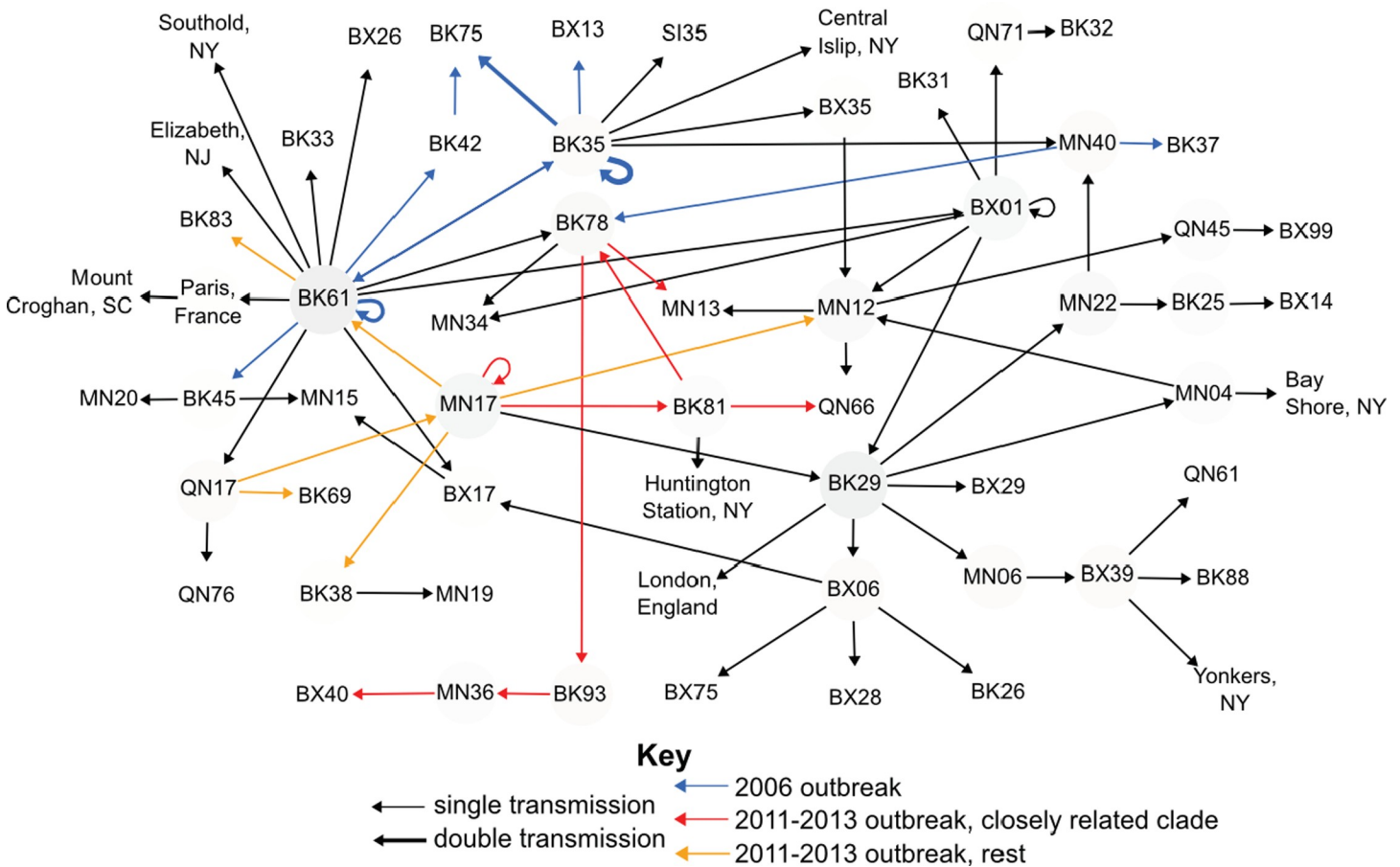
### Discussion

We identified 102 *N. meningitidis* isolates from 2003–2013 and successfully performed WGS on 96 of them to uncover connections among patients not found during case investigations. Our *de novo* genome assembly metrics were comparable to that of other researchers using similar sequencing methods. Multiple studies have estimated that the core genome of *N. meningitidis* is around 1300–1600 genes [45,46, 47], which is consistent with what we found for the genomes analyzed here. As expected, the pan genome for this group is larger than what has been seen in previous studies. This is because *N. meningitidis* has an open pan genome with an



**Fig 2. Betweenness and the movement of *Nm* around NYC boroughs.** Locations with higher betweenness values have a darker colored circle. The arrows indicate the direction of movement with the size of the arrowheads and thickness of arrow scaling with the amount of movement between those locations, with a larger size indicating greater movement.

<https://doi.org/10.1371/journal.pone.0202615.g002>



**Fig 3. Betweenness and the movement of Nm around NYC neighborhoods.** Locations with higher betweenness values have a darker colored circle. Most arrows represents single transmission event between neighborhoods with the 3 exceptions that have slightly thicker arrows: (1) two transmissions happened between BK35 (Stuyvesant Heights) and BK75 (Bedford), (2) two transmissions happened within BK61 (Crown Heights North), (3) three transmissions happened within BK35 (Stuyvesant Heights). Arrows colored blue represent those transmissions that were part Ob1. Ob2 is designated with arrows colored either red for the closely related subclade and orange for the rest of the outbreak isolates. The neighborhood codes are all the official designation from New York City Neighborhood Tabulation Areas. A key for those referenced is in Table A in S1 File. Those from the same borough start with the same letters; BX:Bronx, BK:Brooklyn, MN:Manhattan, SI:Staten Island, QN:Queens.

<https://doi.org/10.1371/journal.pone.0202615.g003>

estimated 38 new genes added to its pan genome with each newly sequenced isolate [46]. Consistent with this estimate, our genomes had an average of 33 unique genes.

The core genome phylogenetic analysis provided evidence that the epidemiologically determined Ob1 and Ob2 were separate outbreaks caused by different, yet related, strains. The overall shape of the phylogenetic tree indicates that while many of the isolates in NYC have highly similar core genomes, including those from Ob1 and Ob2, there are a number of isolates that are very different, having greater similarity to isolates sequenced as parts of other studies. This could be due to the introduction of *N. meningitidis* strains from other parts of the world through immigration and tourism.

Included in the Ob2 clade are two isolates (S00-1218 and S00-1220) from 2008 indicating that isolates with highly similar core genomes were present in NYC prior to the start of this outbreak. These isolates could be missing important genetic elements outside of the core genomes that conferred the Ob2 strain added virulence or afforded it a competitive advantage. Alternatively, and more likely, patient behaviors and social network dynamics differed for those isolates, preventing them from triggering an outbreak in 2008.

Our analysis also identified isolates that should have been considered part of an outbreak based on genetic relatedness that were not epidemiologically categorized as such. For example, S00-1211 is located with the Ob1 isolates on the phylogenetic tree, however, the patient had onset more than six months after the end of the outbreak, was not known to use drugs, and did not live in Brooklyn. The patient resided in a shelter and her onset was close in time to two other IMD cases associated with shelters and substance abuse (neither isolate was available for WGS), suggesting that it is likely the patient came in contact with a person carrying the Ob1 lineage strain.

During Ob2, one case of serogroup C IMD, indistinguishable by PFGE from Ob2 isolates, occurred in a woman (S00-1259) who denied contact with the MSM community and was therefore not considered part of the outbreak. The position of S00-1259 on the phylogenetic tree suggests that she was infected by the Ob2 strain. Two non-resident male cases (S00-1266 and S00-1302) were also not considered part of the outbreak. One patient died and was not known to family members to be MSM and it was later learned that the other individual reported MSM in the year prior to his IMD. Both of these patient isolates lie within the Ob2 subclade and therefore were likely part of the MSM outbreak. One other MSM outbreak-associated patient isolate (S00-1260), which was not related by PFGE criteria, appears in a different part of the phylogenetic tree than the Ob2 strains and we now consider this to have been a sporadic case and unrelated to Ob2.

We found that the amount of within host diversity of *N. meningitidis* was relatively minor by comparing isolates from different specimen sources from the same patient. Same person isolates cluster more closely together in the phylogenetic analysis than they do with isolates from different patients (Fig 1). Same person isolates also had fewer SNPs across their core genomes.

Furthermore, these results support that our use of phylogenetics to determine transmissions events is not likely to be complicated by within host evolution or variation [45]. The total number of unique genes in each isolate is much higher than the number of unique gene for those isolates when compared to the entire OrthoMCL output indicating that most of these genes are present in the Nm pangenomes. The same patient isolate pair S00-1262 and S00-1300 lie within the highly similar Ob2 clade (red box, Fig 1) which serves to emphasize how closely related the Ob2 isolates are to one another.

Both outbreaks were caused by MenC of clonal complex ST-11/ET-37. This strain of Nm is known for its epidemic potential, accounted for 15% of all Nm isolates in Europe in 2011, and is not limited to MenC [48]. MenW strains of this clonal complex were responsible for outbreaks following the Hajj pilgrimages of 2001 and 2002 and subsequent outbreaks in many countries [49, 50]. The recently described nongroupable US Nm urethritis clade is also a member of this clonal complex [22]. Our inability to identify specific markers of virulence among subsets of patients included in this study is in line with current knowledge and understanding of Nm virulence factors [48].

We were able to document the appearance of an intact *aniA* gene in a NYC IMD patient from 2003, and that the proportion of patient isolates with this polymorphism increased from Ob1 to Ob2. The public health implications of this adaptation are not fully understood, however, two recent observations are of concern: 1) the emergence in the US Midwest of *N. meningitidis* urethritis in heterosexual men; and 2) a NYC newborn with *N. meningitidis* conjunctivitis suspected to have been acquired by vertical transmission (DOMHM unpublished data). The *N. meningitidis* strain in these patients contained an intact *aniA* gene [22].

We were limited to the core genome for the SNP GWAS analysis and therefore were unable to look for SNPs in genes that were present in some, but not all, genomes examined. This may have restricted the identification of SNPs important to the epidemiology of the outbreak. Due to bacterial clonal reproduction, GWAS analyses are complicated by phylogenetic



relationships and linkage disequilibrium. For this reason, we included in our analysis correction for population stratification using a genomic control and for multiple testing using the more conservative Bonferroni correction. Although we found a number of SNPs that were associated with Ob1, they varied widely in their functions and could not be correlated with any known association with outbreak transmission. Given that Ob2 occurred in the MSM community, we anticipated that there would be one or more identifiable genetic components to explain the focal epidemiology. Unfortunately, the GWAS analysis did not provide data beyond what has been found in other studies; that genomic distinctions between carriage, non-invasive, and invasive isolates of *N. meningitidis* are not easily determined by gene content and may be polygenic [46].

We used a transmission network analysis to evaluate how *N. meningitidis* moved among NYC communities during and between two outbreaks. Isolates from different specimen sources in the same patient were more similar to each other than other isolates in the analysis, demonstrating that within host evolution or variation is not a complicating factor in using the phylogenetics to determine transmission events [47].

Network analysis could assist in resource allocation, such as expanded chemoprophylaxis or vaccination that may abate or interrupt future outbreaks. This methodology can be particularly effective with granular isolate location data, as that is more likely to produce an accurate representation of the movement of bacterial lineages across a population in an area. For example, this concept is demonstrated by the isolate from the Bay Shore, NY patient, which connects to the rest of the locations in different ways in the borough level analysis compared to the neighborhood level analysis. Both the Brooklyn MenC outbreaks began with cases outside of Brooklyn (Ob1: S00-1200 and Ob2: S00-1224) and serve as additional examples [2,3]. Whether circulating Nm strains, and the individual immunity they provoke in carriers, are localized to neighborhoods or boroughs is not known. It is conceivable that the routine use of network analyses could identify a borough level jump of an Nm strain representing an elevated risk of initiating an outbreak.

Brooklyn was identified as an important hub for all *N. meningitidis* infections over the 11 years examined and in particular for the spread of Ob1. There were two transmission events (MN40 to BK37 and MN40 to BK78, Fig 3) within the neighborhood network analysis that were part of the Ob1 outbreak based on epidemiological data but are disconnected from the rest of the outbreak. This is in part because the epidemiological categorization of which isolates are part of an outbreak is imperfect as discussed above. It is also possible for a number of reasons that we are missing isolates that were part of both outbreaks. Carriers of Nm who do not develop meningococcal disease serve as intermediaries between clinical cases, however, they do not get cultured and links in the transmission chain go undiscovered. Similarly, meningococcal cases who are not NYC residents but may work or travel within the five boroughs, are often not reported to DOHMH. One neighborhood in Manhattan was identified as a hub for to the spread of Ob2. Evidence suggests that the Ob2 strain was transmitted to Europe and caused outbreaks in Germany and France [33]. The disconnection in our network emphasizes the difficulty with performing these types of analyses as a single city. The availability of worldwide transportation allows individuals to move from country to country potentially spreading disease. Perhaps there is no greater recent example than the 2014–5 Ebola virus outbreak in West Africa that spread to Europe and the United States. This only serves to emphasize the importance of countries and institutions sharing data in real-time to obtain the most accurate characterization of global disease movement.

Two Nm outbreaks occurred in NYC in the 21<sup>st</sup> century and both were centered in Brooklyn. Whole genome sequencing of *N. meningitidis* isolates, phylogenetic trees, and betweenness analyses of transmission networks were used to characterize the relationship among cases and



illuminate transmission. These tools add an important dimension to our understanding of outbreaks of IMD and have the potential to inform prevention efforts by identifying high transmission areas. One such approach might be to afford two or more IMD cases occurring in space and time in a Brooklyn neighborhood closer scrutiny to uncover links not immediately detected through routine investigations. Additionally, Nm carriage studies in MSM have either been completed (New York City) or are underway (Los Angeles) and will add to our knowledge of the evolution of this serial killer pathogen.

## Supporting information

### S1 Fig. Flow chart of *Neisseria meningitidis* isolates included in the study.

(TIF)

**S1 File. Study data.** New York City Communicable Disease Surveillance data was used for this study and contains confidential elements. A reduced data set is provided to prevent identification of patients is provided along with data dictionary.

(XLSX)

## Acknowledgments

We thank Mike Antwi and Marie Dorsinville who assisted in data collection. We also thank Marcelle Layton, Demetre Daskalakis, and James Hadler for their critical review of the manuscript. This publication made use of the PubMLST website (<https://pubmlst.org/>) developed by Keith Jolley (Jolley & Maiden 2010, BMC Bioinformatics, 11:595) and sited at the University of Oxford. The development of that website was funded by the Wellcome Trust.

## Author Contributions

**Conceptualization:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Henry S. Gibbons, John Kornblum, Paula Del Rosso, Daniel A. Janies, Don Weiss.

**Data curation:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Henry S. Gibbons, Danielle Bloch, John Kornblum, Paula Del Rosso, Daniel A. Janies, Don Weiss.

**Formal analysis:** Ifeoma Ezeoke, Madeline R. Galac, Alvin T. Liem, Andrew Kilianski, Danielle Bloch, Daniel A. Janies, Don Weiss.

**Funding acquisition:** Henry S. Gibbons, John Kornblum, Daniel A. Janies, Don Weiss.

**Investigation:** Ifeoma Ezeoke, Ying Lin, Alvin T. Liem, Andrew Kilianski, Danielle Bloch, John Kornblum, Paula Del Rosso, Don Weiss.

**Methodology:** Ifeoma Ezeoke, Madeline R. Galac, Alvin T. Liem, Andrew Kilianski, Henry S. Gibbons, Danielle Bloch, John Kornblum, Paula Del Rosso, Daniel A. Janies, Don Weiss.

**Project administration:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Henry S. Gibbons, Danielle Bloch, Paula Del Rosso, Daniel A. Janies, Don Weiss.

**Resources:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Henry S. Gibbons, John Kornblum, Paula Del Rosso, Don Weiss.

**Software:** Ifeoma Ezeoke, Madeline R. Galac, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Danielle Bloch.

**Supervision:** Ying Lin, Alvin T. Liem, Andrew Kilianski, Henry S. Gibbons, John Kornblum, Daniel A. Janies, Don Weiss.

**Validation:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Danielle Bloch, Don Weiss.

**Visualization:** Madeline R. Galac, Alvin T. Liem.

**Writing – original draft:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Danielle Bloch, Don Weiss.

**Writing – review & editing:** Ifeoma Ezeoke, Madeline R. Galac, Ying Lin, Alvin T. Liem, Pierce A. Roth, Andrew Kilianski, Henry S. Gibbons, Danielle Bloch, John Kornblum, Paula Del Rosso, Daniel A. Janies, Don Weiss.

## References

- Centers for Disease Control and Prevention. Summary of Notifiable Infectious Diseases and Conditions, 2015. *MMWR Morb Mortal Wkly Rep.* 2015; 64: 116.
- Weiss D, Stern EJ, Zimmerman C, Bregman B, Yeung A, Das D, et al. Epidemiologic Investigation and Targeted Vaccination Initiative in Response to an Outbreak of Meningococcal Disease among Illicit Drug Users in Brooklyn, New York. *Clin Inf Dis.* 2009; 48: 894–901. <https://doi.org/10.1086/597257> PMID: 19231975
- Kratz MM, Weiss D, Ridpath A, Zucker JR, Geevarughese A, Rakeman J, et al. Community-Based Outbreak of *Neisseria meningitidis* Serogroup C Infection in Men who Have Sex with Men, New York City, New York, USA, 2010–2013. *Emerg Infect Dis.* 2015; 21(8):1379–86. <https://doi.org/10.3201/eid2108.141837> PMID: 26197087
- Tsang RSW, Kiefer L, Law DKS, Stoltz J, Shahin R, Brown S, et al. Outbreak of Serogroup C Meningococcal Disease Caused by a Variant of *Neisseria meningitidis* Serotype 2a ET-15 in a Community of Men Who Have Sex with Men. *J Clin Microbiol.* 2003; 41(9):4411–4. <https://doi.org/10.1128/JCM.41.9.4411-4414.2003> PMID: 12958279
- Schmink S, Watson JT, Coulson GB, Jones RC, Diaz PS, Mayer LW, et al. Molecular Epidemiology of *Neisseria meningitidis* Isolates from an Outbreak of Meningococcal Disease among Men Who Have Sex with Men, Chicago, Illinois, 2003. *J Clin Microbiol.* 2007; 45(11):3768–70. <https://doi.org/10.1128/JCM.01190-07> PMID: 17728467
- Nanduri S, Foo C, Ngo V, Jarashow C, Civen R, Schwartz B, et al. Outbreak of Serogroup C Meningococcal Disease Primarily Affecting Men Who Have Sex with Men—Southern California, 2016. *Morb Mortal Weekly Rep.* 2016; 65: 939–940. <https://doi.org/10.15585/mmwr.mm6535e1> PMID: 27606798
- Folaranmi TA, Kretz CB, Kamiya H, MacNeil JR, Whaley MJ, Blain A, et al. Increased Risk for Meningococcal Disease Among Men Who Have Sex With Men in the United States, 2012–2015. *Clin Inf Dis.* 2017; 65 (5): 756–763. <https://doi.org/10.1093/cid/cix438> PMID: 28505234
- Marcus U, Vogel U, Schubert A, Claus H, Baetzing-Feigenbaum J, Hellenbrand W, et al. A cluster of invasive meningococcal disease in young men who have sex with men in Berlin, October 2012 to May 2013. *Eurosurveillance.* 2013; 18(28): pii = 20523.
- Aubert L, Taha MK, Boo N, Le Strat Y, Deghmane AE, Sanna A, et al. Serogroup C invasive meningococcal disease among men who have sex with men and in gay-oriented social venues in the Paris region: July 2013 to December 2014. *Eurosurveillance.* 2015; 20(3): pii = 21016
- Gutierrez-Fernandez J, Medina V, Hidalgo-Tenorio C, Abad R. Two Cases of *Neisseria meningitidis* Proctitis in HIV-positive Men Who Have Sex With Men. *Emerg Inf Dis.* 2017; 23 (3): 542–543. <https://doi.org/10.3201/eid2303.161039>
- Read RC. *Neisseria meningitidis*; clones, carriage and disease. *Clin Microbiol Infect.* 2014; 20: 391–395. <https://doi.org/10.1111/1469-0691.12647> PMID: 24766477
- Quan G, Gilbert M, David ST, Rahim T, Adie K, Shaw C, et al. Usefulness of pulsed-field gel electrophoresis in tracking two outbreaks of invasive meningococcal disease serogroup C in British Columbia. *Can J Infect Dis Med Microbiol.* 2007; 18(6): 363–367. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2533566/pdf/jidmm18363.pdf> PMID: 18978987
- Lavezzo E, Toppo S, Franchin E, Di Camillo B, Finotello F, Falda M, et al. Genomic comparative analysis and gene function prediction in infectious diseases: application to the investigation of a meningitis outbreak. *BMC Infect Dis.* 2013; 13: 554. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4225559/pdf/1471-2334-13-554.pdf> <https://doi.org/10.1186/1471-2334-13-554> PMID: 24252229

14. Walker TM, Ip LC, Harrell RH, Evans JT, Kapatai G, Dedicoat J, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis*. 2012; 13: 137–146. [https://doi.org/10.1016/S1473-3099\(12\)70277-3](https://doi.org/10.1016/S1473-3099(12)70277-3) PMID: 23158499
15. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles JC, et al. The Origin of the Haitian Cholera Outbreak Strain. *N Eng J Med*. 2011; 364: 33–42. <https://doi.org/10.1056/NEJMoa1012928#t=article>
16. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program, Henderson DK, et al. Tracking a Hospital Outbreak of Carbapenem-Resistant *Klebsiella pneumoniae* with Whole-Genome Sequencing. *Sci Transl Med*. 2012; 4(148): 148ra116. <https://doi.org/10.1126/scitranslmed.3004129> PMID: 22914622
17. Lapiere P, Nazarian E, Zhu Y, Wroblewski D, Saylor A, Passaretti T, Hughes S, et al. Legionnaires' Disease Outbreak Caused by Endemic Strain of *Legionella pneumophila*, New York, USA, 2015. *Emerg Inf Dis*. 2017; 23(11): <https://doi.org/10.3201/eid2311.170308> PMID: 29047425.
18. Kleta S, Hammerl JA, Dieckmann R, Malorny B, Borowiak M, Halbedel S, et al. Molecular Tracing to Find Source of a Protracted Invasive Listeriosis Outbreak, Southern Germany, 2012–2016. *Emerg Inf Dis*. 2017; 23(10): 1680–1683. <https://doi.org/10.3201/eid2310.161623> PMID: 28930013.
19. Lucidarme J, Scott KJ, Ure R, Smith A, Lindsay D, Stenmark B, et al. An international invasive meningococcal disease outbreak due to a novel and rapidly expanding serogroup W strain, Scotland and Sweden, July to August, 2015. *Eurosurveillance*. 2016; 21:pii = 30395.
20. Taha MK, Claus H, Lappmann M, Veyrier FJ, Otto A, Becher D, et al. Evolutionary Events Associated with an Outbreak of Meningococcal Disease in Men Who Have Sex with Men. *PLOS ONE*. 2017; 11(5): e0154047. <https://doi.org/10.1371/journal.pone.0154047> PMID: 27167067
21. Bazan JA, Turner AN, Kirkcaldy RD, Retchless AC, Kretz CB, Briere E, et al. Large Cluster of *Neisseria meningitidis* Urethritis in Columbus, Ohio, 2015. *Clin Infect Dis*. 2017; 65(1): 92–99. <https://doi.org/10.1093/cid/cix215> PMID: 28481980
22. Tzeng YL, Bazan JA, Turner AN, Wang X, Retchless AC, Read TD, et al. Emergence of a new *Neisseria meningitidis* clonal complex 11 lineage 11.2 clade as an effective urogenital pathogen. *PNAS*. 2017; 114: 4237–4242. <https://doi.org/10.1073/pnas.1620971114> PMID: 28373547
23. Council of State and Territorial Epidemiologists. Position statement 14-ID-06. Revision of the National Surveillance Case Definition for Meningococcal Disease. 2014; 2017 (May 26).
24. Heckenberg S, Brouwer M, Van de Beek D. Bacterial meningitis. *Handb Clin Neuro*. 2014; 121: 1361–75. <https://doi.org/10.1016/B978-0-7020-4088-7.00093-6>
25. Guarner J, Greer PW, Whitney A, Shieh WJ, Fischer M, White EH, et al. Pathogenesis and diagnosis of human meningococcal disease using immunohistochemical and PCR assays. *Am J Clin Pathol*. 2004; 122:754–64. <https://doi.org/10.1309/A7M2-FN2T-YE6A-8UFX> PMID: 15491972
26. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18:821–9. <https://doi.org/10.1101/gr.074492.107> PMID: 18349386
27. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008; 9:75. <http://www.ncbi.nlm.nih.gov/pubmed/18261238>. <https://doi.org/10.1186/1471-2164-9-75> PMID: 18261238
28. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep*. 2015; 5:8365. <http://www.ncbi.nlm.nih.gov/pubmed/25666585> <http://www.ncbi.nlm.nih.gov/pubmed/25666585> <https://doi.org/10.1038/srep08365> PMID: 25666585
29. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003; 13:2178–89. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=403725&tool=pmcentrez&rendertype=abstract> <https://doi.org/10.1101/gr.1224503> PMID: 12952885
30. Chen F, Mackey AJ, Stoeckert CJ, Roos DS. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res*. 2006; 34:D363–8. <https://doi.org/10.1093/nar/gkj123> PMID: 16381887
31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 421. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
32. Veyrier FJ, Hong E, Deghmane A-E, Taha M-K. Draft Genome Sequence of a *Neisseria meningitidis* Serogroup C Isolate of Sequence Type 11 Linked to an Outbreak among Men Who Have Sex with Men. *Genome Announc*. 2013; 1:20522. <http://www.ncbi.nlm.nih.gov/pubmed/24092789>.

33. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010; 11: 595. <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-595> <https://doi.org/10.1186/1471-2105-11-595> PMID: 21143983
34. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013; 30:772–80. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3603318&tool=pmcentrez&rendertype=abstract>
35. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
36. Kück P, Meusemann K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* 2010; 56:1115–8. <http://www.ncbi.nlm.nih.gov/pubmed/20416383> <https://doi.org/10.1016/j.ympev.2010.04.024> PMID: 20416383
37. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22:2688–90. <http://www.ncbi.nlm.nih.gov/pubmed/16928733> <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
38. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81:559–575. <https://doi.org/10.1086/519795> PMID: 17701901
39. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005; 21:3674–6. <http://www.ncbi.nlm.nih.gov/pubmed/16081474> <https://doi.org/10.1093/bioinformatics/bti610> PMID: 16081474
40. Janies DA, Pomeroy LW, Krueger C, Zhang Y, Senturk I, Kaya K, et al. Phylogenetic visualization of the spread of H7 influenza A viruses. *Cladistics*. 2015; 1–13.
41. Swofford DL. PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods). 2002; Sinauer Associates, Sunderland, Massachusetts.
42. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Syst.* 2006; 1695:1–9.
43. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network Visualizations of Relationships in Psychometric Data. 2012; *J. Stat. Softw.* 48. <http://www.jstatsoft.org/v48/i04/>.
44. Bratcher HB, Corton C, Jolley K a, Parkhill J, Maiden MC. A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. *BMC Genomics*. 2014; 15:1138 <https://doi.org/10.1186/1471-2164-15-1138> PMID: 25523208
45. Schoen C, Tettelin H, Parkhill J, Frosch M. Genome flexibility in *Neisseria meningitidis*. *Vaccine*. 2009; 27 Suppl 2:B103–11.
46. Schoen C, Blom J, Claus H, Schramm-Glück A, Brandt P, Müller T, et al. Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:3473–8. <https://doi.org/10.1073/pnas.0800151105> PMID: 18305155
47. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat. Rev. Microbiol.* 2016; 14:150–162. <https://doi.org/10.1038/nrmicro.2015.13> PMID: 26806595
48. Waśko I, Hryniewicz W, Skoczyńska A. Significance of Meningococcal Hyperinvasive Clonal Complexes and their Influence on Vaccine Development.
49. Yezli S. The threat of meningococcal disease during Hajj and Umrah mass gatherings: A comprehensive review. *Travel Med Infect. Dis.* 2018; 24: 51–58. *Pol. J. Microbio.* 2015; 64: 313–3321. <https://doi.org/10.1016/j.tmaid.2018.05.003> PMID: 29751133
50. Mustapha MM, March JW, Harrison LH. Global epidemiology of Group W meningococcal disease (1970–2015): Multifocal emergence and persistence of hypervirulent sequence type (ST)-11 clonal complex. *Vaccine* 2016; 34: 1515–1523. <https://doi.org/10.1016/j.vaccine.2016.02.014> PMID: 26876439