A peer-reviewed version of this preprint was published in PeerJ on 5 March 2015.

View the peer-reviewed version (peerj.com/articles/806), which is the preferred citable publication unless you specifically need to cite this preprint.

Ezewudo MN, Joseph SJ, Castillo-Ramirez S, Dean D, del Rio C, Didelot X, Dillon J, Selden RF, Shafer WM, Turingan RS, Unemo M, Read TD. 2015. Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance. PeerJ 3:e806
https://doi.org/10.7717/peerj.806

Population structure of *Neisseria gonorrhoeae* based on whole genome data and its relationship with antibiotic resistance

Matthew Matthew Ezewudo, Sandeep Joseph, Santiago Castillo-Ramirez, Deborah Dean, Carlos Del Rio, Xavier Didelot, Jo-Anne Dillon, Richard F. Selden, William M. Shafer, Rosemary S. Turingan, Magnus Unemo, Timothy D. Read

Neisseria gonorrhoeae is the causative agent of gonorrhea, a sexually transmitted infection (STI) of major importance. As a result of antibiotic resistance, there are now limited options for treating patients. We collected whole genome sequences and associated metadata data on 76 *N. gonorrhoeae* strains from around the globe and searched for known determinants of antibiotics resistance within the strains. The population structure and evolutionary forces within the pathogen population were analyzed. Our results indicated a cosmopolitan gonoccocal population mainly made up of five subgroups. The estimated ratio of recombination to mutation (*r/m*=2.2) from our data set indicates an appreciable level of recombination occurring in the population. Strains with resistance phenotypes to more recent antibiotics (azithromycin and cefixime) were mostly found in two of the five population subgroups.

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

1 **Matthew N. Ezewudo[1], Sandeep J. Joseph[1], Santiago Castillo-Ramirez[2], Deborah Dean[3],**
2 **Carlos del Rio[1,4], Xavier Didelot[5], Jo-Anne Dillon[6], Richard F Selden[7], William M. Shafer[8,9],**
3 **Rosemary S. Turingan[7], Magnus Unemo[10], *Timothy D. Read[1]**

4 [1]Emory University School of Medicine, Division of Infectious Diseases, Atlanta, Georgia, USA
5 [2]National Autonomous University of Mexico, Cuernavaca, Mexico
6 [3]Children Hospital of Oakland Research institute, Oakland, California, USA
7 [4]Emory University School of Public Health, Atlanta, Georgia, USA
8 [5]Imperial College London, UK
9 [6] Department of Microbiology and Immunology, College of Medicine, Vaccine and Infectious
10 Disease Organization –International Vaccine Centre, University of , Saskatchewan, Saskatoon,
11 Saskatchewan, Canada
12 [7] NetBio, Waltham, MA USA
13 [8,9] Department of Microbiology and Immunology, Emory University School of Medicine, Atlanta,
14 Georgia and Laboratories of Bacterial Pathogenesis, Veterans Affairs Medical Center, Decatur,
15 Georgia, USA
16 [10]WHO Collaborating Centre for Gonorrhoea and other STIs, Örebro University Hospital,
17 Örebro, Sweden

18 ***Corresponding Author**: Timothy D. Read
19 Division of Infectious Diseases,
20 Emory University School of Medicine,
21 Atlanta, GA, USA
22 tread@emory.edu
23 +14047279706

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

24 ## Introduction

25 *Neisseria gonorrhoeae,* a Gram-negative bacterium, causes gonorrhea, a very prevalent disease
26 responsible for 106 million of an estimated 498 million new cases of curable non-viral sexually
27 transmitted infections (STIs), occurring globally every year (World Health Organization (WHO),
28 2012). The only effective option for treating the disease and stopping its spread has been the use
29 of antimicrobial therapy. Currently, there is no vaccine to prevent infection. Antimicrobial
30 treatment options have diminished over time due to the emergence of antimicrobial resistance
31 (AMR) to all of the classes of drugs previously used to treat gonorrhea and the paucity in the
32 development of newer antibiotics that could effectively eradicate the pathogen (Ohnishi et al.,
33 2011; Unemo & Shafer, 2014).

34 AMR evolution should be considered in the context of the genetic structure of the *N.*
35 *gonorrhoeae* population. Early work by O'Rourke *et al.* using electrophoretic analysis of
36 enzymes of the pathogen and serological methods suggested that gonococci have a non-clonal
37 sexual or panmictic population structure (O'Rourke & Stevens, 1993). More recent studies have
38 also suggested high rates of recombination within the *Neisseria* genus (Didelot & Maiden, 2010).
39 High levels of recombination could confound studies of the gonococcal populations, especially if
40 the studies are based on few genetic loci within strains as compared to the entire genomes. Recent
41 multi-genome studies have focused on either a restricted geographic region (Vidovic et al., 2014)
42 (genomes also included in present studies or on a small subset of the *N. gonorrhoeae* population
43 (Grad et al., 2014). Hence, there is a need for studies aimed at understanding the global *N.*
44 *gonorrhoeae* population structure at the whole genome scale.

PeerJ PrePrints

45 Past AMR studies using limited numbers of gonococcal strains from specific geographic regions
46 of the globe have mostly focused on a number of representative genes or genetic regions of the
47 genome to elucidate underlying mechanisms of antibiotic resistance (Hagman & Shafer, 1995;
48 Lindberg, Fredlund, Nicholas, & Unemo, 2007; Ohneck et al., 2011; Thakur, Levett, Horsman, &
49 Dillon, 2014; Tomberg, Unemo, Ohnishi, Davies, & Nicholas, 2013; Unemo & Shafer, 2014;
50 Unemo, Golparian, & Hellmark, 2014; Unemo, Golparian, & Nicholas, 2012; Zhao et al., 2009).
51 Extensive genome sequencing studies have yet to be conducted on a diverse collection of strains
52 from different geographical locations and collected over longer time periods. Our approach in
53 this study builds on recent multi-genome studies (Grad et al., 2014; Vidovic et al., 2014), with
54 the goal of using whole genome analysis to elucidate two processes: 1) the population structure
55 and dynamics of *Neisseria gonorrhoeae* and 2) the correlation between this population
56 differentiation and AMR evolution in gonococci. Our genome analysis of strains from multiple
57 sites across the world offers a geographic diversity of *N. gonorrhoeae* isolates, providing more
58 depth in genome-wide studies of this pathogen and identifying possible sub-populations
59 impacting AMR and evolution within the species.

60 Materials and Methods

61 *Neisseria gonorrhoeae* isolates

62 Sixty-one *N. gonorrhoeae* isolates of diverse origin were obtained. These included isolates from
63 the Gonococcal Isolation Surveillance Program (GISP) site covering Atlanta, Miami, New York
64 city and North Carolina in the United States (n=21), from Canada (primarily, Saskatchewan) and
65 Chile (Vidovic et al., 2014)) (n=24), and from WHO global collaborations; Sweden (n=7),
66 Norway (n=3), Japan (n=2), Austria (n=1), Pakistan (n=1), Philippines (n=1), and Australia
67 (n=1). Phenotypic determination of the mimum inhibitory concenterations (MICs) of all isolates

PeerJ PrePrints

68 was performed using the agar dilution method or the Etest method (bioMerieux), according to the
69 instructions from the manufacturer. The strains sequenced in this study were tested for resistance
70 to primarily three antibiotics, tetracycline, azithromycin and cefixime, with breakpoints for
71 resistance set at 2, 2.0, and 0.25 μg/mL, respectively, based on the CDC MIC (minimum
72 inhibitory concentration) breakpoints for testing in the GISP project. Antibiotic resistance
73 profiles of the Canadian strains have been previously reported (Vidovic et al., 2014). Details of
74 the different isolates with their NCBI accession numbers are presented in Table 1.

75 **Sequence generation and assembly**

76 The *N. gonorrhoeae* strains were shotgun sequenced (WGS) using the Illumina HiSeq™
77 instrument, utilizing libraries prepared from 5 μg of genomic DNA for each sample. The average
78 sequencing coverage was 225. The sequencing reads were filtered using the prinseq-lite
79 algorithm (Schmieder R. et al, 2011) to ensure only sequence reads with average phred score ≥30
80 were used. The reads for each project were then assembled *de novo*, using the velvet assembler
81 program (Zerbino & Birney, 2008). The optimal kmer length for each assembly prior to
82 assembly was determined using the velvet optimizer algorithm(Gladman & Seemann, 2012).
83 Data was deposited in the NCBI Sequence Read Archive public database (Accession #
84 SRA099559) (Table 1). For this study, we included an additional 14 draft genome sequences of
85 *N. gonorrhoeae* strains, downloaded from the NCBI draft genomes database (NCBI Bioproject
86 numbers: PRJNA55649, PRJNA55651, PRJNA55653, PRJNA55905, PRJNA46993,
87 PRJNA55657, PRJNA55655, PRJNA55659, PRJNA55661, PRJNA55663, PRJNA55665,
88 PRJNA55667, PRJNA55669, PRJNA55671, and the reference genome sequence Ref_FA_1090
89 (NC_002946.2)

PeerJ PrePrints

90 **Whole genome phylogeny and pangenome analysis**

91 The assembled genomes were annotated individually using the NCBI PGAP annotation pipeline
92 to give predicted proteome for each of the strains. The orthologs were determined by OrthoMCL
93 (Li, Stoeckert, & Roos, 2003), which uses bi-directional BLASTP scores of all the protein
94 sequences to perform Markov clustering in order to improve sensitivity and specificity of the
95 orthologs. For the OrthoMCL analysis, we used a BLASTP E-value cut-off of 1e-05, and
96 inflation Markov clustering parameter of 1.5. Core genes were defined as the orthologous genes
97 that are shared among all the *N. gonorrhoeae* strains used in this analysis

98 The nucleotide sequences of all the core genes were concatenated together and whole-genome
99 nucleotide alignment was conducted using progressive MAUVE (Darling, Mau, Blattner, &
100 Perna, 2004). Similarly, whole genome amino-acid alignment was also generated by
101 concatenating the deduced amino-acid sequences of all the core genes generated using
102 MUSCLE(Edgar, 2004), and to form a super protein alignment. Homoplasious sites were
103 removed from the whole-genome nucleotide alignment using the Noisy software (Dress et al.,
104 2008). The protein alignments were filtered by GBLOCKS (Talavera & Castresana, 2007)using
105 default settings to remove regions that contained gaps or were highly diverged. A maximum
106 likelihood (ML) tree from the same data set was created using the GTR and JTT substitution
107 models for the nucleotide and protein alignment respectively and the GAMMA evolutionary
108 model (Stamatakis, 2014). The majority rule-consensus tree was generated from 200 bootstrap
109 replicates of the model. Linear regression of the root-to-tip distances against the year of isolation
110 was performed using the Path-O-Gen tool (http://tree.bio.ed.ac.uk/software/pathogen/).

111 **Multi-locus sequence typing (MLST) locus analysis**

PeerJ PrePrints

112 MLST is a genotyping tool for *Neisseria* based on sequencing of 7 core housekeeping
113 genes(Jolley & Maiden, 2010). There are currently close to 11,000 individual *Neisseria* sequence
114 profiles in the publicly available MLST database (http://pubmlst.org). We utilized a custom
115 python script mlstBLAST.py (http://sourceforge.net/projects/srst/files/mlstBLAST/) to perform a
116 BLAST search of these genes across all the strains in our data set and identified the sequence
117 type (ST) for each strain. Novel alleles of the locus and STs were submitted to the MLST
118 database. A phylogeny of the concatenated DNA sequences of all the *N. gonorrhoeae* STs in the
119 MLST public database was created using the neighbor joining distance matrix approach of the
120 PHYLIP(Felstein, 1989). Mean nucleotide distance for the sequence alignments and MLST genes
121 was computed using MEGA software (Tamura, et al 2013).

122 **Estimating population parameters and homologous recombination**

123 ClonalFrame(Didelot & Falush, 2007) utilizes a statistical framework to reconstruct the clonal
124 genealogy as well as identify the regions along the genomes that has been affected both by
125 recombination and mutation. The model uses a Bayesian approach to predict the phylogenetic
126 relationship in the sample set, given the whole genome sequence alignment data. The input
127 genome alignment data was the core genes (n = 1189) alignment generated from MAUVE. Four
128 independent ClonalFrame runs were performed for 40,000 iterations, with the first 20,000
129 discarded as burn-in. This allowed the model parameters to converge and each of the 4 runs were
130 checked for the consistency of the estimated parameters as well as the consistency of the
131 topology of the inferred clonal genealogies.

132 **Population structure analysis**

133 The program BAPS (Bayesian Analysis of Population Structure) version 5.3 (Corander &
134 Marttinen, 2006; Tang, Hanage, Fraser, & Corander, 2009)was used to infer the underlying

PeerJ PrePrints

135 population structure of the 76 *N. gonorrhoeae* strains in the sample set. SNPs from the core
136 MAUVE alignment, with gaps removed were converted to a BAPS input file, which is a
137 representation of all the polymorphic loci in the multi-sequence alignment. BAPS applied a
138 Bayesian model to predict the likelihood of a population structure given the input data and non-
139 parametric assumption approach to trace ancestry of the different individuals in the sample set.
140 For the mixture analysis we used the 'Clustering of individuals' approach. We ran a preliminary
141 analysis to evaluate the approximate number of genetically differentiated groups using a vector
142 from 2 to 40 K values, where K is the maximum number of groups. Given that 5 groups was the
143 K value with the best log likelihood, we ran a second analysis using from 3 to 7 K values and
144 again the best K value was 5 groups. We used the 'Admixture based on mixture clustering'
145 module for the admixture analysis. For the analysis; the minimum population and the admixture
146 coefficient for the individuals was then set to 5. For the reference individuals from each
147 population and the admixture coefficient for reference individuals we used the values as
148 described by Castillo-Ramírez et al (Castillo-Ramírez et al., 2012). In addition, population
149 structure analysis of the sample set using the fineSTRUCTURE tool(Lawson, Hellenthal, Myers,
150 & Falush, 2012) was performed. fineSTRUCTURE analysis was a two step process-1)
151 ChromoPainter algorithm was used to generate the co-ancestry matrix from the genome-wide
152 haplotype data using the linkage model. 2) The fineSTRUCTURE algorithm performed a model-
153 based clustering using a Bayesian MCMC approach to predict the likelihood of a population
154 structure given the input data and non-parametric assumption approach to trace ancestry of the
155 different individuals in the sample set. The fineSTRUCTURE approach was used to corroborate
156 the findings from the BAPS population structure analysis.

157 **Mapping the movement of DNA between *Neisseria gonorrhoeae* clades**

PeerJ PrePrints

We traced the flow of recombination between strains into five different subgroups in the phylogeny determined from the subgroups of the population defined by the BAPS analysis. We created a BLAST database of the whole genome sequence of all 76 strains in the sample set and included 14 whole genome sequences of all other *Neisseria* species that are present in the NCBI database. Next, we performed a BLASTN search for each of the genomic region within the strains identified by ClonalFrame to be under recombination, selecting the best hit within the sequences in the database we created, with an identity of >98%, to be the source of the recombined region. We also filtered off hits that had strains from similar subgroups as the source of the recombined region. We used the *migest* package (http://cran.r-project.org/web/packages/migest/) implemented in the R statistical language to create a circular representation of the matrix of relationship between the subpopulations identified by BAPS based on the purported recombination between strains in the different subgroups. We also supplied *migest* with the matrix from BAPS admixture analysis and recreated the circular flow of recombination across only the subpopulations as defined by BAPS.

Comparison of nucleotide substitution rates

Amino acid sequences were aligned using MUSCLE sequence aligner(Edgar, 2004). The amino acid sequence alignment was converted to nucleotide alignment based on the corresponding gene sequence using PAL2NAL(Suyama, Torrents, & Bork, 2006) and we implemented the YN00 method of the PAML package(Yang, 2007) to calculate the pairwise dN/dS ratios for the strains(Rocha et al., 2006). The contribution of each strain to the overall variation in the dN/dS rates across the sample set was estimated using ANOVA (Analysis of Variance) approaches.

Analysis of positive selection

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

180 For the analysis of positive selection within core genes of the strains in the sample set, we first
181 identified and removed core genes that have signals of homologous recombination using three
182 methods of Pairwise Homoplasy Index (PHI), Neighbor Similarity Score (NSS) and the
183 maximum χ^2 method. The three methods are implemented in the PhiPack package(Sawyer, 1989).
184 A window size of 50 nucleotides was used to run the methods in the package, and genes shown to
185 have significant probability of homologous recombination by a majority of the methods were not
186 used for the positive selection analysis. Next, we identified core genes under positive selection
187 using codeml of PAML tool version 4.7(Yang, 2007). We applied the branch-sites test for positive
188 selection Model A test 2 of the tool, to identify genes under positive selection population groups.
189 For each of the clades, we performed the Likelihood Ratio Test (LRT) for two hypotheses - the
190 null hypothesis is the existence of neutral selection as implemented in the null model versus the
191 alternative hypothesis implemented in the test model for positive selection. The LRT was
192 performed to a degree of freedom of 1, and we corrected for multiple testing using the False
193 discovery rate approach (FDR)(Benjamini & Hochberg, 1995). We further identified the Gene
194 Ontology (GO) terms and functional characterizations of the genes under positive selection (see
195 Table 2) and performed an enrichment test for functionality of these genes using the blast2go test
196 pipeline(Götz et al., 2011).

197 **Confirming known predictors of antibiotic resistance phenotype**

198 We downloaded from NCBI reference DNA sequences of resistance determinants that have been
199 shown in the literature to underlie the resistance phenotype we have observed in our sample set
200 (see Table 3), and performed a BLASTN search for each of these DNA sequence regions across
201 all the strains in the database of whole genome sequences. For convenience, the contigs for each

PeerJ PrePrints

202 assembly were ordered into one pseudocontig after tiling to the reference genome FA1090, using
203 the ABACUS tool (http://abacas.sourceforge.net/).
204 We selected the top hit (with identity match of 98% or more) for each sequence (strain) in the
205 database and parsed the alignment between the query and the subject sequence in the database for
206 the presence or absence of the underlying resistance genetic mutations as suggested in the
207 literature.

208 Results and Discussion

209 Genome-wide homologous recombination in diverse *N. gonorrhoeae*

210 We sequenced 61 recent clinical isolates primarily from the US and Canada but also single
211 representatives from other countries, including Japan, Pakistan, Australia, Austria, Philippines,
212 Norway and Sweden, A preliminary phylogeographical analysis of the Canadian isolates (n=23)
213 was recently published (Vidovic et al., *2014*). For the analysis, we included the 14 *N.*
214 *gonorrhoeae* draft NCBI genome sequences (12 from the US and 2 from Europe) and the genome
215 sequence of the FA1090 *N. gonorrhoeae* reference strain. The 76 were assigned into 23
216 previously described MLST STs and four new STs (10931,10932,10933,10934). The genetic
217 diversity (measured as pairwise nucleotide distances of MLST loci) of the strains in this study
218 was about half that of the *N. gonorrhoeae* strains as a whole (0.001 substitutions per site in our
219 study compared to 0.002 in the large MLST set), and the strains were evenly distributed across
220 the different clades of a phylogeny of housekeeping genes of *N. gonorrhoeae* strains found in the
221 MLST database (see Fig S5). Alignment of the shotgun assembly to reference genome FA1090
222 (NC_002946.2), yielded 10,962 SNPs in the core region (conserved in all strains). The average
223 per nucleotide diversity in the core genome regions was 0.003.

PeerJ PrePrints

224 Homologous recombination is known to play a role in shaping bacteria populations(Didelot &
225 Maiden, 2010). The ClonalFrame tool (Didelot & Falush, 2007) detected 952 independent
226 recombination events, covering more than 50% of the reference genome. The average size of the
227 recombination regions identified was 360 base pairs. The estimate for the ratio of effects of
228 recombination and mutation (*r/m*) for our strain set was 2.2, a relatively high value for bacterial
229 species (Didelot & Maiden, 2010) (Fig S3, and quite similar to the *r/m* estimate of 1.9 based on
230 the whole genome alignment on a less genetically diverse group of *N. gonorrhoeae* strains
231 reported by Grad et al(Grad et al., 2014).

232 We constructed a maximum likelihood phylogeny of the core genome of the 76 strains using the
233 RAxML program (excluding regions identified as potentially recombinant) (Fig 1). This tree had
234 similar topology to the clonal frame that determined by the eponymous software (Fig S3). The
235 tree showed multiple clades but the there was no strong signal of genetic isolation by distance at
236 the continental scale. The rate of the molecular clock was estimated to be 8.93×10^{-6} mutations per
237 year based on the slope of the regression of the root-to-tip divergence with isolation dates (see Fig
238 S2). This value was similar to those obtained in other bacterial studies, ranging from 8.6×10^{-9} to
239 2.5×10^{-5} (Zhou et al., 2013). However, because the temporal signal was weak in the root-to-tip
240 analysis (Figure S2), we did not use these data for Bayesian phylogeny analysis using the BEAST
241 phylogeny tool (Drummond & Rambaut, 2007).

242 ***Neisseria gonorrhoeae* population structure and biogeography**

243 Given that recombination was frequent in these genomes, we sought to evaluate the genetic
244 substructure of the population. We used two complementary methods. BAPS (Tang et al., 2009)
245 predicts the likelihood of a population structure given the input data and uses a non-parametric

PeerJ PrePrints

246 assumption approach to trace ancestry. fineSTRUCTURE (Lawson et al., 2012), on the other
247 hand, uses similar methods of predicting population substructure, but to a finer detail and does
248 not assume a prior optimum number of subpopulations (K). The BAPS tool identified 5
249 subgroups within the *N. gonorrhoeae* population from the strains within the sample set (Fig 3).
250 As expected, members with the same subgroup ancestry generally were found near each other
251 when mapped on the ML phylogeny constructed using the nonrecombining portion of the
252 genome. On the other hand, fineSTRUCTURE analysis divided the 76 strains into 30 genetic
253 subgroups (Fig S4). However, every one of the fineSTRUCTURE subgroups contained members
254 from only one BAPS subgroup. Each of the five BAPS subgroups contained strains from multiple
255 continents based on geography or location of isolation (Fig 1). It was particularly interesting that
256 each BAPS cluster had at least one US strain and one candadian strain. The BAPS analysis
257 revealed a complex relationship between Group 3 and 5, with the latter group 5 separated into
258 two group 3 clades (Fig 1). Group 3 strains in clades closely related to group 5 showed
259 significant genetic import from group 5. It is possible that the extent of admixture occurring in
260 group 3 and 5 may have caused misidentification.

261 We assessed patterns of genetic drift effects in the population by estimating the pairwise
262 substitution rates between all the core gene orthologs for the strains and determining the mean
263 dN/dS ratio for each strain. The mean pairwise dN/dS ratios for each strain are shown in Fig 4.
264 There was significant variation in the mean dN/dS ratios among the strains (ANOVA p-value =
265 2.0e-16). The overall mean of the dN/dS estimate was 0.3184, similar to the 0.402 value
266 estimated for the bacterial pathogen *Chlamydia trachomatis* (Joseph et al., 2012). The mean
267 dN/dS ratio for strains from the Canadian region was 0.3279, which was above the overall mean
268 ratio, while that for strains collected in the US was 0.31708, which is below the overall pairwise

269 dN/dS mean ratio for the sample set. This was also a statistically significant difference (p-value
270 0.0018 for t test of means), suggesting a possible geographical effect within this subset of strains.

271 **Genetic admixture within *N. gonorrhoeae* and with other *Neisseria species***

272 In order to understand the flow of genetic information between the strains from five different
273 subgroups defined by the BAPS analysis (Fig 3) as well as strains from other *Neisseria* species,
274 we used two independent approaches. The first was to search each of the 952 recombination
275 regions identified by ClonalFrame for a best BLASTN match from another subgroup or *Neisseria*
276 species (We created a blast database of the 76 genomes from this study and representative strains
277 from the *Neisseria* genus) (Fig S3). In parallel, we also counted the occurrence of co-ancestry of
278 genetic markers revealed by the BAPS analysis. Both the BAPS and BLAST analyses suggested
279 that group 3 was the most common nexus of homologous recombination between other clades,
280 consistent with its basal phylogenetic status. In the BAPS-based network groups 1 and 2, and to a
281 greater extent, group 5, were primarily DNA donors to group 3 (Fig 2B). But this pattern was less
282 visible in the BLAST network (Fig 2A). It is notable that more than 90% of the recombination
283 with strains from other *Neisseria* species occured in groups 2 and 3. Group 5 stood out as a
284 significant source of genetic exchange into strains in group 3.
285 The genetic relatedness of the strains in the sample set or the purported sharing of genetic
286 materials across the different subgroups shown by the BAPS figure paralleled the pattern
287 revealed by the BLAST clonal frame analysis (p-value = 0.048, Mantel test for comparing the
288 distance matrices of the five populations between both methods). The exchange of genetic
289 materials from other *Neisseria* species was not accounted for in the BAPS admixture analysis.
290 Based on the BLAST analysis, the proportion of DNA transferred within *N. gonorrhoeae*
291 compared to arriving from *Neisseria* strains the species was 729 out of 849 intra-specific genetic
292 events. This finding is line with the “fuzzy species” concept of Fraser et al(Fraser, Hanage, &

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014
PeerJ PrePrints

293 Spratt, 2007): while *N. gonorrhoeae* is not sexually isolated, DNA flow seemed predominantly
294 through intra-specific exchanges.

295 **Genes under positive selection**

296 Of the 1189 core genes, we identified 352 genes as likely to contain past recombination histories
297 using the PHIPACK tests (Sawyer, 1989). Thirty-one genes within the subset of 837 non-
298 recombining core genes were found to be under positive selection using the tests implemented by
299 the PAML software (Materials and Methods). BAPS subgroup 5 had the highest number of core
300 genes under selection (14) followed by subgroup 3 (7). While we found no significant
301 enrichment of genes under positive selection in any of the functional classes in the Gene
302 Ontology (GO) database, the functions of the best match proteins from genes under positive
303 selection can be broadly classified to genes involved in DNA or RNA synthesis of gene
304 expression, membrane or transport proteins, and, to a lesser extent, genes involved in metabolic
305 pathways in the bacterial cell (Supplement Data S2 spreadsheet). Of the 352 genes found to have
306 signals of recombination, we found no significant enrichment of the genes in any of the
307 functional classes in the GO database. The functions of these genes could broadly be classified
308 into 2 groups: genes encoding membrane and transport proteins; and those involved in metabolic
309 pathways in the cell (Supplement Data S3).

310 In regard to antibiotic resistance and selection, the most interesting gene found to be under
311 positive selection was *porB* (Smith, Maynard Smith, & Spratt, 1995), which has been shown to
312 be involved in mechanisms of resistance to penicillins, macrolides, cephalosporins and
313 tetracyclines (Unemo & Shafer, 2014). *porB* exhibited signals of selection in subgroups 1, 3 and
314 5 - the groups that harbored most of the antibiotic resistant strains in our sample set (Fig 5).
315 *comA*, which encodes a membrane protein necessary for competence of *N. gonorrhoeae,* was also

PeerJ PrePrints

316 found to be under selection in a handful of strains that make up subgroup 5. This finding is of
317 interest in regards to the potential for DNA uptake in these strains, since they appear to be
318 primarily DNA donors, rather than recipients in genetic exchanges (Fig 2). Other genes
319 putatively under selection included a stress response gene, a gene encoding a chaperone protein
320 of the HscA family and a number of proteins: ribosyl transferase, RNA polymerase, and an *arsR*
321 family transcriptional regulator, which were all linked to gene expression. Genes under positive
322 selection in subgroup 3 were also mainly involved in gene expression or DNA metabolism,
323 including DNA helicase and tRNA pseudo uridine synthase. Most of the genes with known
324 functions, identified to be under positive selection in subgroup 1 were either membrane-
325 associated or transport proteins.

326 **Analysis of known genetic predictors for AMR phenotypes**

327 A substantial amount of research effort over the past 10 years has been devoted to understanding
328 the genetic basis of drug resistance in *N. gonorrhoeae* (Garvin et al., 2008; MD et al., 2014;
329 Unemo & Shafer, 2011; Veal, Nicholas, & Shafer, 2002; World Health Organization (WHO),
330 2012). Since there is a an increasing interest in the direct attribution of resistance phenotypes
331 based on genome sequencing, we attempted to ascertain how knowledge of existing variants
332 could be applied to the *N. gonorrhoeae* genomes in this study. We searched for variants known to
333 underlie resistance to 3 antibiotics classes within our study (Table 3). In terms of subgroup
334 distribution, tetracycline resistance was found in each of the 5 population subgroups,
335 azithromycin resistance was present in only 2 of the strains tested (SK36809 and MUNG8) and
336 restricted to subgroup 2 (Fig 5), and cefixime resistance was found in subgroup 1 and subgroup 2.
337 We identified genes responsible for resistance to the drugs tested in this works using literature
338 searches and the CARD antimicrobial resistance database (McArthur et al., 2013).

PeerJ PrePrints

339 The *tetM* resistant determinant, which confers high-level resistance to tetracycline, is borne on
340 plasmids and is transferred either through conjugation or transformation (Knapp, Johnson,
341 Zenilman, Roberts, & Morse, 1988; Morse, Johnson, Biddle, & Roberts, 1986; Turner, Gough, &
342 Leeming, 1999). It was found in only 5 of the 10 strains with high-level resistance to tetracyline
343 (MIC equal or greater than 16 µg/ml). Strain SK1902, one of the 5 strains with the *tetM*
344 determinant, had a significantly higher MIC (> 256 µg/ml) than the rest (see attached Supplement
345 S3). Other strains with reduced susceptibility or chrosomally-mediated resistance to tetracycline,
346 i.e without the *tetM* determinant do have other corresponding chromosomal mutations on one or
347 more of the resistance loci: *mtrR* (including its promoter), *penB*, *rpsJ*. Only one strain (ATL0508)
348 within the sample set exhibits resistance to tetracycline in the laboratory, without the presence of
349 any of the known resistance determinants of the tetracycline resistance phenotype.

350 Different "mosaic" *penA* alleles are thought to have developed from recombination with portions
351 of DNA transferred horizontally from commensal *Neisseria* and/or *N. meningitidis* and underly
352 decreased susceptibility or resistance to cephalosporins by preventing their binding action on the
353 encoded mosaic PBP2 (Ameyama et al., 2002). The mosaic *penA* XXXIV (Ohnishi, 2011; Grad
354 2014; Unemo & Shafer, 2014) had the best positive predictive value of all the known resistance
355 determinants we searched for within our dataset, being present in 6/7 of the strains resistant to
356 cefixime. This result echoed the observations made by Grad et al (Grad et al., 2014)in their
357 epidemiologic study of *N. gonorrhoeae* strains. The other loci (i.e., mutations in the *mtrR*,
358 *mtrCDE* operon promoter region and *penB* gene) also proven to enhance the MICs of
359 cephalosporins (Unemo & Shafer, 2011; Warner, Shafer, & Jerse, 2008) did not have a similar
360 predictive property within strains in our data set. These variants were seen in 2 out of 7 and 3 out
361 of 7 cefixime resistant strains, respectively. MUNG17 is the only strain in the sample set that has

PeerJ PrePrints

362 an elevated MIC (0.38µg/mL) to cefixime that we could not find any of the known resistance

363 determinants within its genome sequence (See attached supplement Data S1).

364 Resistance to azithromycin can be mediated by mutations in the previously mentioned *penB* and

365 *mtr* operon genes as well as mutations found in the 4 different alleles of the 23S rRNA gene that

366 inhibits protein synthesis (Chisholm et al., 2009; Palmer, Young, & Winter, 2008; Starnino,

367 Stefanelli, N*eisseria gonorrhoeae* Italian Study Group, 2009). The 23S rRNA mutation allele was

368 found in one (SK36809) of the two strains with the azithromycin resistance phenotype. The other

369 azithromycin resistant strain, MUNG8, did not have the 23S rRNA resistance determinant or any

370 of the other known mutations in the *mtrR* or *penB* loci (See attached Supplement Data S1).

371 Conclusions

372 Our study suggested that *N. gonorrhoeae* globally is made up of at least five genetic

373 subpopulations. That individual strains from the subpopulations are from diverse geographical

374 locations confirms the cosmopolitan nature of the pathogen. This suggested a population

375 structure with multiple waves of rapid international expansion. Subgroup 3 strains may be the

376 nexus for gene exchange within the species. Groups 1 and 2 might be the most recently branched

377 and contain a higher proportion of resistant isolates. Given the importance of the antibiotic

378 resistant phenotype, these may be emerging lineages that are expanding within *N. gonorrhoeae.*

379 Our analysis confirms earlier studies that showed an appreciable effect of recombination within

380 the population. This could be playing a role in the evolution of AMR in the bacterium, as strains

381 with resistance phenotypes to currently used antibiotics are mostly within similar population

382 subgroupings.

PeerJ PrePrints

383 Although most of the known predictors that underlie the observed resistance phenotypes were
384 accounted for in the strains we studied, they could not explain some of the phenotypes of several
385 strains. These findings suggested that a broader genome search of a large number of whole
386 genomes from strains of this pathogen could yield candidate novel variants that may explain
387 some of the "missing" antibiotic resistance phenotypes we have observed.
388 In general, large genome sequencing studies examining a high number of temporally and
389 geographically diverse *N. gonorrhoeae* isolates are essential to elucidate the evolution and
390 diversity of the *N. gonorrhoeae* species as well as associations between genomic content,
391 antibiotic resistance and clinical outcome of treatment.

392 **Acknowledgements**

393 Genome sequencing was performed at the Emory Genomics Center. We wish to thank Tauqeer
394 Alam at Emory University, for advice on phylogenetic methods, and Sinisa Vidovic and Sidarath
395 Dev from the International Vaccine center at the University of Saskatchewan, Canada for their
396 help in antimicrobial susceptibility determination and preliminary genomic assessments. We also
397 thank the Broad Institute for pre-publication release of genomic data (Bioproject PRJNA55649-
398 PRJNA55649) used in this study.

399 References

400 Ameyama, S., Onodera, S., Takahata, M., Minami, S., Maki, N., Endo, K., et al. (2002). Mosaic-like
401 structure of penicillin-binding protein 2 Gene (penA) in clinical isolates of Neisseria gonorrhoeae
402 with reduced susceptibility to cefixime. *Antimicrobial Agents and Chemotherapy*, *46*(12), 3744–3749.
403 Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful
404 approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57(1), 289 -300.
405 Castillo-Ramírez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W. P., Westh, H., et al. (2012).
406 Phylogeographic variation in recombination rates within a global clone of methicillin-resistant
407 Staphylococcus aureus. *Genome biology*, *13*(12), R126–R126. doi:10.1186/gb-2012-13-12-r126

PeerJ PrePrints

408 Chisholm, S. A., Neal, T. J., Alawattegama, A. B., Birley, H. D. L., Howe, R. A., & Ison, C. A. (2009).
409 Emergence of high-level azithromycin resistance in Neisseria gonorrhoeae in England and Wales. *The*
410 *Journal of antimicrobial chemotherapy*, *64*(2), 353–358. doi:10.1093/jac/dkp188
411 Corander, J., & Marttinen, P. (2006). Bayesian identification of admixture events using multilocus
412 molecular markers. *Molecular ecology*.
413 Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved
414 genomic sequence with rearrangements. *Genome Research*, *14*(7), 1394–1403.
415 doi:10.1101/gr.2289704
416 Didelot, X., & Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data.
417 *Genetics*, *175*(3), 1251–1266. doi:10.1534/genetics.106.063305
418 Didelot, X., & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in*
419 *Microbiology*, *18*(7), 315–322. doi:10.1016/j.tim.2010.04.002
420 Dress, A. W. M., Flamm, C., Fritzsch, G., Grünewald, S., Kruspe, M., Prohaska, S. J., & Stadler, P. F.
421 (2008). Noisy: identification of problematic columns in multiple sequence alignments. *Algorithms for*
422 *molecular biology : AMB*, *3*, 7. doi:10.1186/1748-7188-3-7
423 Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees.
424 *BMC evolutionary biology*, *7*, 214. doi:10.1186/1471-2148-7-214
425 Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space
426 complexity. *BMC Bioinformatics*, *5*, 113–113. doi:10.1186/1471-2105-5-113
427 Felstein, J. (1989). PHYLIP-Phylogeny Inference Package. *Cladistics*, *5*(2), 163–166. doi:10.1111/j.1096-
428 0031.1989.tb00562.x
429 Fraser, C., Hanage, W. P., & Spratt, B. G. (2007). Recombination and the nature of bacterial speciation.
430 *Science*, *315*(5811), 476–480. doi:10.1126/science.1127573
431 Garvin, L. E., Bash, M. C., Keys, C., Warner, D. M., Ram, S., Shafer, W. M., & Jerse, A. E. (2008).
432 Phenotypic and Genotypic Analyses of Neisseria gonorrhoeae Isolates That Express Frequently
433 Recovered PorB PIA Variable Region Types Suggest that Certain P1a Porin Sequences Confer a
434 Selective Advantage for Urogenital Tract Infection. *Infection and Immunity*, *76*(8), 3700–3709.
435 doi:10.1128/IAI.00265-08
436 Gladman, S., & Seemann, T. (2012, December 22). Velvet Optimiser. *vicbioinformatics.com*. Retrieved
437 May 30, 2014, from http://www.vicbioinformatics.com/software.velvetoptimiser.shtml
438 Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.-A., et al. (2011).
439 B2G-FAR, a species-centered GO annotation repository. *Journal of Gerontology*, *27*(7), 919–924.
440 doi:10.1093/bioinformatics/btr059
441 Grad, Y. H., Kirkcaldy, R. D., Trees, D., Dordel, J., Harris, S. R., Goldstein, E., et al. (2014). Genomic
442 epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a
443 retrospective observational study. *The Lancet Infectious Diseases*, *14*(3), 220–226.
444 doi:10.1016/S1473-3099(13)70693-5
445 Hagman, K. E., & Shafer, W. M. (1995). Transcriptional control of the mtr efflux system of Neisseria
446 gonorrhoeae. *Journal of bacteriology*, *177*(14), 4162–4165.
447 Jolley, K. A., & Maiden, M. C. J. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the
448 population level. *BMC Bioinformatics*, *11*, 595. doi:10.1186/1471-2105-11-595
449 Joseph, S. J., Didelot, X., Rothschild, J., de Vries, H. J. C., Morré, S. A., Read, T. D., & Dean, D. (2012).
450 Population genomics of Chlamydia trachomatis: insights on drift, selection, recombination, and
451 population structure. *Molecular Biology and Evolution*, *29*(12), 3933–3946.
452 doi:10.1093/molbev/mss198
453 Knapp, J. S., Johnson, S. R., Zenilman, J. M., Roberts, M. C., & Morse, S. A. (1988). High-level
454 tetracycline resistance resulting from TetM in strains of Neisseria spp., Kingella denitrificans, and
455 Eikenella corrodens. *Antimicrobial Agents and Chemotherapy*, *32*(5), 765–767.
456 Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using
457 dense haplotype data. *PLoS genetics*, *8*(1), e1002453. doi:10.1371/journal.pgen.1002453
458 Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic
459 genomes. *Genome Research*, *13*(9), 2178–2189. doi:10.1101/gr.1224503

PeerJ PrePrints

460 Lindberg, R., Fredlund, H., Nicholas, R., & Unemo, M. (2007). Neisseria gonorrhoeae isolates with
461 reduced susceptibility to cefixime and ceftriaxone: association with genetic polymorphisms in penA,
462 mtrR, porB1b, and ponA. *Antimicrobial Agents and Chemotherapy*, *51*(6), 2117–2122.
463 doi:10.1128/AAC.01604-06
464 McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., et al. (2013). The
465 comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3348–
466 3357. doi:10.1128/AAC.00419-13
467 MD, Y. H. G., MD, R. D. K., PhD, D. T., PhD, J. D., PhD, S. R. H., PhD, E. G., et al. (2014). Genomic
468 epidemiology of. *The Lancet Infectious Diseases*, *14*(3), 220–226. doi:10.1016/S1473-
469 3099(13)70693-5
470 Morse, S. A., Johnson, S. R., Biddle, J. W., & Roberts, M. C. (1986). High-level tetracycline resistance in
471 Neisseria gonorrhoeae is result of acquisition of streptococcal tetM determinant. *Antimicrobial Agents*
472 *and Chemotherapy*, *30*(5), 664–670.
473 O'Rourke, M., & Stevens, E. (1993). Genetic structure of Neisseria gonorrhoeae populations: a non-clonal
474 pathogen. *Journal of general microbiology*, *139*(11), 2603–2611.
475 Ohneck, E. A., Zalucki, Y. M., Johnson, P. J. T., Dhulipala, V., Golparian, D., Unemo, M., et al. (2011). A
476 novel mechanism of high-level, broad-spectrum antibiotic resistance caused by a single base pair
477 change in Neisseria gonorrhoeae. *MBio*, *2*(5). doi:10.1128/mBio.00187-11
478 Ohnishi, M., Golparian, D., Shimuta, K., Saika, T., Hoshina, S., Iwasaku, K., et al. (2011). Is Neisseria
479 gonorrhoeae Initiating a Future Era of Untreatable Gonorrhea?: Detailed Characterization of the First
480 Strain with High-Level Resistance to Ceftriaxone. *Antimicrobial Agents and Chemotherapy*, *55*(7),
481 3538–3545. doi:10.1128/AAC.00325-11
482 Palmer, H. M., Young, H., & Winter, A. (2008). Emergence and spread of azithromycin-resistant Neisseria
483 gonorrhoeae in Scotland. *Journal of antimicrobial* chemotherapy, 62(3), 490 -494.
484 Rocha, E. P. C., Smith, J. M., Hurst, L. D., Holden, M. T. G., Cooper, J. E., Smith, N. H., & Feil, E. J.
485 (2006). Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of*
486 *theoretical biology*, *239*(2), 226–235. doi:10.1016/j.jtbi.2005.08.037
487 Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Molecular Biology and Evolution*, *6*(5),
488 526–538.
489 Smith, N. H., Maynard Smith, J., & Spratt, B. G. (1995). Sequence evolution of the porB gene of
490 Neisseria gonorrhoeae and Neisseria meningitidis: evidence of positive Darwinian selection.
491 *Molecular Biology and Evolution*, *12*(3), 363–370.
492 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
493 phylogenies. *Bioinformatics*. doi:10.1093/bioinformatics/btu033
494 Starnino, S., Stefanelli, P., Neisseria gonorrhoeae Italian Study Group. (2009). Azithromycin-resistant
495 Neisseria gonorrhoeae strains recently isolated in Italy. *The Journal of antimicrobial chemotherapy*,
496 *63*(6), 1200–1204. doi:10.1093/jac/dkp118
497 Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence
498 alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(Web Server issue),
499 609–612. doi:10.1093/nar/gkl315
500 Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and
501 ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, *56*(4), 564–577.
502 doi:10.1080/10635150701472164
503 Tang, J., Hanage, W. P., Fraser, C., & Corander, J. (2009). Identifying currents in the gene pool for
504 bacterial populations using an integrative approach. *PLoS computational biology*.
505 Thakur, S. D., Levett, P. N., Horsman, G. B., & Dillon, J.-A. R. (2014). Molecular epidemiology of
506 Neisseria gonorrhoeae isolates from Saskatchewan, Canada: utility of NG-MAST in predicting
507 antimicrobial susceptibility regionally. *Sexually transmitted infections*, *90*(4), 297–302.
508 doi:10.1136/sextrans-2013-051229
509 Tomberg, J., Unemo, M., Ohnishi, M., Davies, C., & Nicholas, R. A. (2013). Identification of amino acids
510 conferring high-level resistance to expanded-spectrum cephalosporins in the penA gene from
511 Neisseria gonorrhoeae strain H041. *Antimicrobial Agents and Chemotherapy*, *57*(7), 3029–3036.
512 doi:10.1128/AAC.00093-13

PeerJ PrePrints

513 Turner, A., Gough, K. R., & Leeming, J. P. (1999). Molecular epidemiology of tetM genes in Neisseria
514 gonorrhoeae. *Sexually transmitted infections*, *75*(1), 60–66.
515 Unemo, M., & Shafer, W. M. (2011). Antibiotic resistance in Neisseria gonorrhoeae: origin, evolution, and
516 lessons learned for the future. *Annals of the New York Academy of Sciences*, *1230*(1), E19–E28.
517 doi:10.1111/j.1749-6632.2011.06215.x
518 Unemo, M., & Shafer, W. M. (2014). Antimicrobial resistance in Neisseria gonorrhoeae in the 21st
519 century: past, evolution, and future. *Clinical Microbiology Reviews*, *27*(3), 587–613.
520 doi:10.1128/CMR.00010-14
521 Unemo, M., Golparian, D., & Hellmark, B. (2014). First three Neisseria gonorrhoeae isolates with high-
522 level resistance to azithromycin in Sweden: a threat to currently available dual-antimicrobial
523 regimens for treatment of gonorrhea? *Antimicrobial Agents and Chemotherapy*, *58*(1), 624–625.
524 doi:10.1128/AAC.02093-13
525 Unemo, M., Golparian, D., & Nicholas, R. (2012). High-level cefixime-and ceftriaxone-resistant Neisseria
526 gonorrhoeae in France: novel penA mosaic allele in a successful international clone causes treatment
527 failure. *Antimicrobial agents and Chemotherapy 56(3) 1273 - 1280*
528 Veal, W. L., Nicholas, R. A., & Shafer, W. M. (2002). Overexpression of the MtrC-MtrD-MtrE efflux
529 pump due to an mtrR mutation is required for chromosomally mediated penicillin resistance in
530 Neisseria gonorrhoeae. *Journal of bacteriology*, *184*(20), 5619–5624.
531 Vidovic, S., Caron, C., Taheri, A., Thakur, S. D., Read, T. D., Kusalik, A., & Dillon, J.-A. R. (2014). Using
532 Crude Whole-Genome Assemblies of Neisseria gonorrhoeae as a Platform for Strain Analysis: Clonal
533 Spread of Gonorrhea Infection in Saskatchewan, Canada. *Journal of Clinical Microbiology*, *52*(10),
534 3772–3776. doi:10.1128/JCM.01502-14
535 Warner, D. M., Shafer, W. M., & Jerse, A. E. (2008). Clinically relevant mutations that cause derepression
536 of the Neisseria gonorrhoeaeMtrC-MtrD-MtrE Efflux pump system confer different levels of
537 antimicrobial resistance and in vivofitness. *Molecular Microbiology*, *70*(2), 462–478.
538 doi:10.1111/j.1365-2958.2008.06424.x
539 World Health Organization (WHO). (2012). Global action plan to control the spread and impact of
540 antimicrobial resistance in *Neisseria gonorrhoea*.
541 Yang, Z. Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and*
542 *Evolution*, *24*(8), 1586–1591. doi:10.1093/molbev/msm088
543 Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn
544 graphs. *Genome Research*, *18*(5), 821–829. doi:10.1101/gr.074492.107
545 Zhao, S., Duncan, M., Tomberg, J., Davies, C., Unemo, M., & Nicholas, R. A. (2009). Genetics of
546 chromosomally mediated intermediate resistance to ceftriaxone and cefixime in Neisseria
547 gonorrhoeae. *Antimicrobial Agents and Chemotherapy*, *53*(9), 3744–3751. doi:10.1128/AAC.00304-
548 09
549 Zhou, Z., McCann, A., Litrup, E., Murphy, R., Cormican, M., Fanning, S., et al. (2013). Neutral Genomic
550 Microevolution of a Recently Emerged Pathogen, Salmonella enterica Serovar Agona. *PLoS genetics*,
551 *9*(4), e1003471. doi:10.1371/journal.pgen.1003471

PeerJ PrePrints

552 **Tables and Figures**

Strain Name	Location	Date	MLST	Azithromycin (MIC)	Cefixime (MIC)	Tetracycline (MIC)
CH811	Chile	1982	1583	0.25	0.008	2
GC1-182	Canada	1982	1583	0.5	0.008	4
SK708	Canada	2006	1594	1	0.016	0.5
SK1902	Canada	2006	10935	0.25	0.002	256
SK6987	Canada	2006	10010	1	0.016	4
SK7461	Canada	2008	1901	0.5	0.032	8
SK7842	Canada	2006	10010	1	0.016	8
SK8976	Canada	2006	1594	0.06	0.004	2
SK12684	Canada	2006	31129	0.5	0.016	8
SK33414	Canada	2007	1928	0.25	0.008	4
SK14515	Canada	2005	1893	0.25	0.016	2
SK15454	Canada	2007	1585	0.06	0.004	2

PeerJ PrePrints

SK16259	Canada	2007	1893	0.125	0.008	4
SK16942	Canada	2005	1893	0.125	0.016	2
SK17973	Canada	2006	1893	1	0.016	8
SK22871	Canada	2007	8122	0.125	0.004	4
SK23020	Canada	2006	1901	0.25	0.125	16
SK28355	Canada	2007	1893	0.25	0.016	4
SK29344	Canada	2007	10010	0.125	0.008	4
SK29471	Canada	2005	1893	0.25	0.016	2
SK32402	Canada	2007	8153	0.5	0.016	4
SK36809	Canada	2007	8126	2	0.008	8
SK39420	Canada	2008	1585	0.5	0.016	0.5
ALB0303	USA	2011	1588	0.03	0.015	16
ALB0403	USA	2011	1901	1	0.125	4
ATL0103	USA	2011	10931	0.5	0.015	0.25
ALB0102	USA	2011	1901	0.25	0.06	2
ATL0105	USA	2011	1588	0.06	0.015	0.25
ATL0108	USA	2011	1584	0.03	0.015	0.25
ATL0117	USA	2011	10932	0.125	0.015	16
ATL0121	USA	2011	1902	0.5	0.03	1
ATL0125	USA	2011	1901	0.25	0.015	1
ATL0508	USA	2011	1585	0.06	0.015	16
ATL0513	USA	2011	1893	0.25	0.03	2
MIA0202	USA	2011	1901	0.5	0.03	2
MIA0309	USA	2011	1931	0.125	0.015	16
MIA0310	USA	2011	1584	0.03	0.015	16
MIA0510	USA	2011	1901	1	0.03	2
MIA0515	USA	2011	1901	0.25	0.03	16
MIA0516	USA	2011	1901	0.5	0.06	8

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

NOR0306	USA	2011	1583	0.25	0.015	2
NYC0507	USA	2011	1901	0.25	0.06	2
NYC0513	USA	2011	1901	0.25	0.06	4
MUNG1	Canada	1991	10934	0.125	<0.016	0.25
MUNG3	Japan	2003	7363	0.25	0.5	2
MUNG4	Japan	1996	1590	0.5	0.25	4
MUNG5	Philippines	1992	1901	0.25	<0.016	1
MUNG6	Australia	2001	10008	0.125	<0.016	16
MUNG8	USA	2001	8127	2	<0.016	0.5
MUNG9	Sweden	2010	1901	0.5	1	2
MUNG12	Norway	2010	1901	0.5	0.25	4
MUNG14	Norway	2010	1901	0.5	0.25	4
MUNG15	Austria	2011	1901	0.25	1	2
MUNG17	Sweden	2010	1892	1	0.5	2
MUNG18	Norway	2010	10933	0.125	<0.016	2
MUNG19	Sweden	2010	1580	>256	<0.016	2
MUNG20	Sweden	2013	7363	0.25	0.5	2
MUNG21	Pakistan	2008	1902	1	0.032	2
MUNG23	Sweden	1998	1585	0.064	<0.016	0.125
MUNG25	Sweden	1998	1901	0.125	<0.016	0.5
MUNG26	Sweden	1999	1584	0.064	<0.016	0.5

Table 1 Location and date of collection of the *N. gonorrhoeae* strains, including Sequence Types (ST) and MICs of the different strains to the antibiotics azithromycin, cefixime and tetracycline. The MIC breakpoint value for azithromycin resistance is 2 ug/mL, for cefixime 0.25 ug/mL, and for tetracycline 2 ug/mL, based on the CDC breakpoints for antibiotic testing.

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

Gene	Clades Present	Gene ID (reference genomeFA 1090)
PorB	1,3,5	NGO1812
Acetate kinase 2	5	NGO1521
Primosomal replication protein	3	NGO0582
DNA Helicase	3,5	NGO1196
Hypothethical protein	5	NGO0880
Hypothethical Protein	5	NGO1847
Hypothetical protein	5	NGO1948
ComA	5	NGO0276
Chaperone Protein HscA	5	NGO0829
tRNA-ribosyltransferase	5	NGO0294
RNA polymerase Subunit β	5	NGO1850
ArsR family transcriptional regulator	5	NGO1562
Hypothethical protein	5	NGO0165
PriB	5	NGO0582
ABC transporter subunit	3	NGO2088
Hypothethical protein	3	NGO1984
tRNA pseudouridine synthase B	3	NGO0642
Prolyl endopeptidase	1	NGO0026
Apo-lipoprotein N-acyltransferase	1	NGO0289
Sodium dependent transporter	1	NGO2096
Phage associated protein	1	NGO1012
Hypothetical Protein	4	NGO0914

557 **Table 2.** Core genes of *N. gonorrhoeae* under positive selection in the different clades of the phylogeny of
558 strains in the sample set

PeerJ PrePrints

PeerJ PrePrints

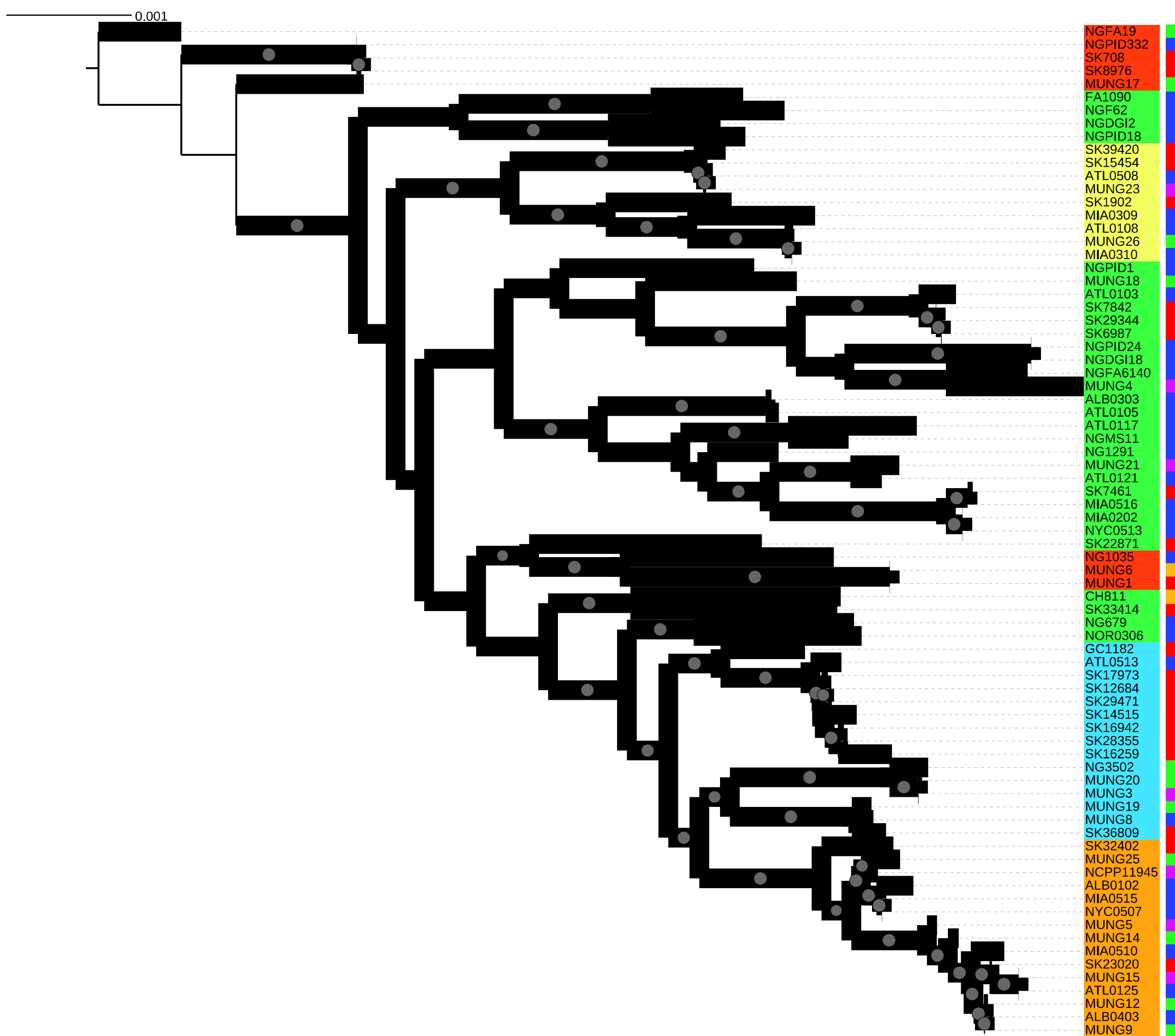
Gene name	**FA1090 Reference locus_tag/Gene Bank ID**	**Genetic Mutations**	**Resistance Phenotype**	**References**
mtrR	NGO1366	G45D, A39T (glycine and aspartate substitutions)	Decreased susceptibility to macrolides and beta-lactams	PMID: 18761689
mtrCDE promoter	NGO1366	Single nucleotide deletion on reference genome position 1327932	Decreased susceptibility to macrolides and beta-lactams	PMID: 18761689

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

penB	NGO1812	G101K, A102D (glycine and alanine substitutions)	Decreased susceptibility to third-generation cephalosporins	PMID: 17420216
Mosaic *penA*	NGO1542	Mosaic pattern amino acid substitutions form position 294 to end of gene	Decreased susceptibility to third-generation cephalosporins	PMID: 20028823
rpsJ	NGO1841	V57M	Decreased susceptibility to tetracycline	PMID:16189114
23S rRNA	AF450080	C2611T(Cystine toThronine substitution)	Decreased suscptibility to Azithromycin	PMID: 12183262
tetM	N/A	Horizontally transferred determinant on plasmids	Resistance to tetracycline (MIC >= 16µg/ml)	PMID: 21349987

559 **Table 3**. Known antibiotic resistance determinants, their PubMed reference ID and their associated
560 resistance phenotypes in *N. gonorrhoeae*. Cephalosporin antibiotics include cefixime
561 and ceftriaxone, while macrolides include erythromycin and azithromycin.

PeerJ PrePrints

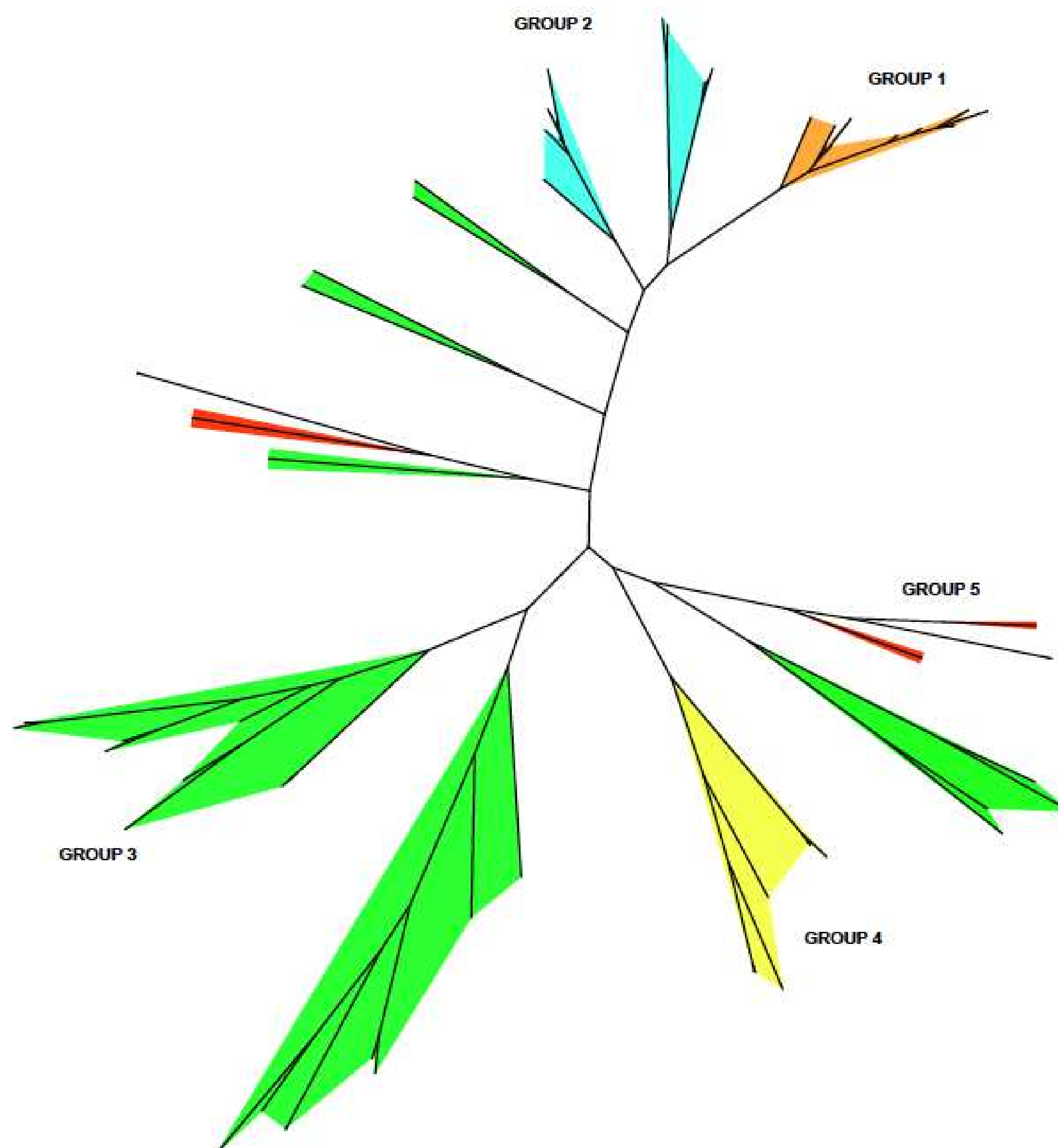


562

A

PeerJ PrePrints

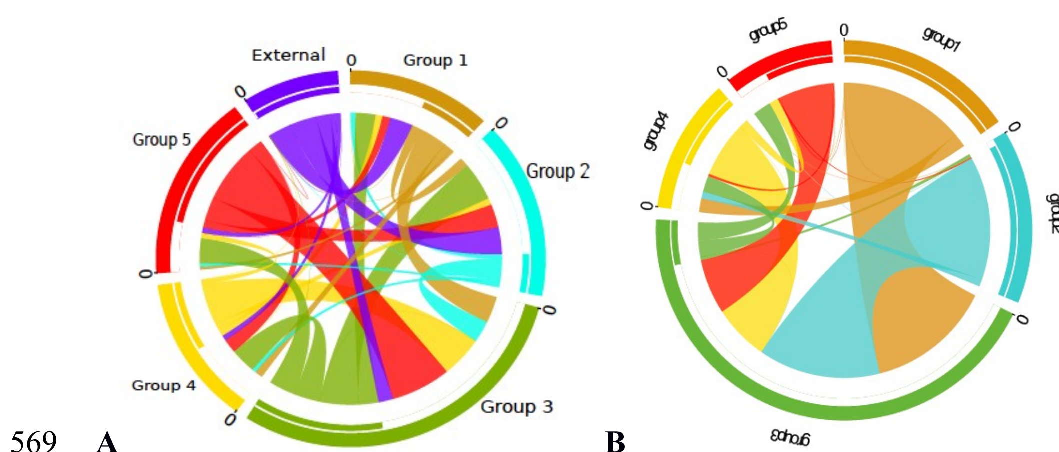
PeerJ PrePrints



563 B

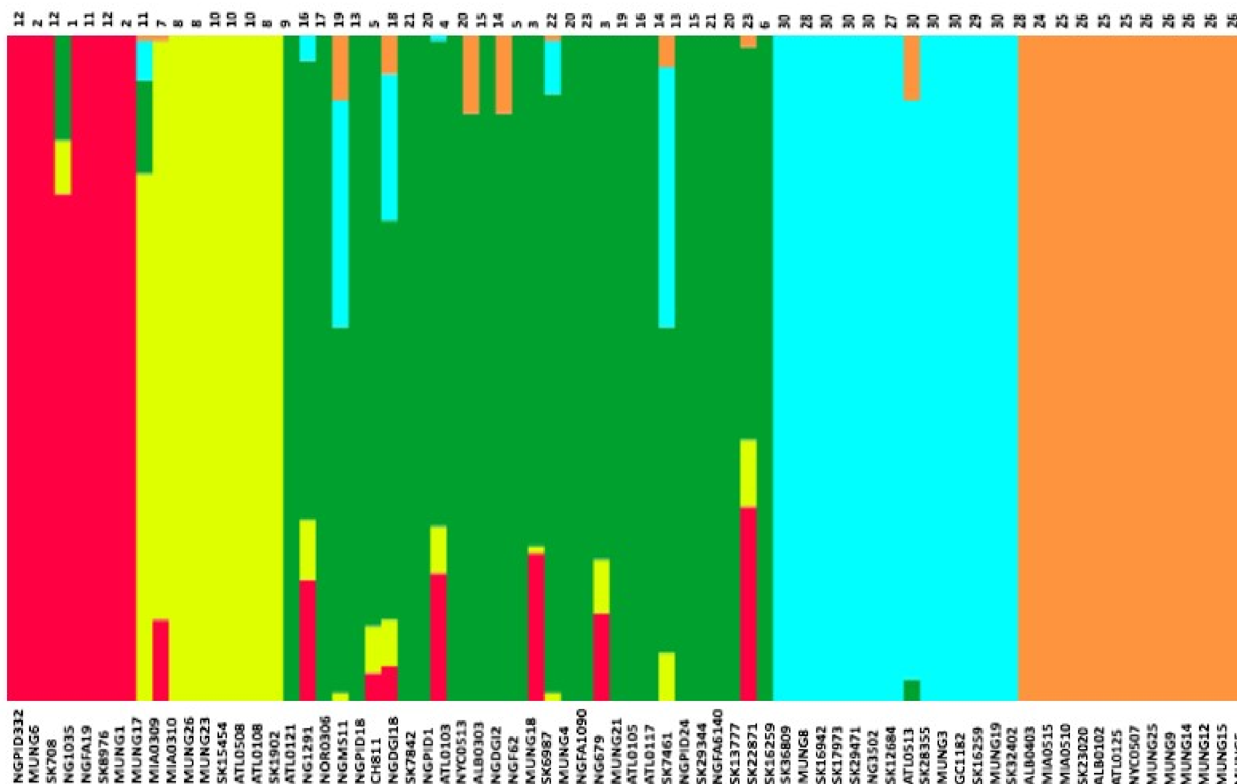
564 **Figure 1A** (Top). Maximum Likelihood phylogeny of sequenced strains of *N. gonorrhoeae* Branches with
565 boostrap value > 80% for branches are indicated. Taxa are highlighted based on 5 different subgroups
566 defined by BAPS. Annotations next to the leaves are colored based on location of isolation; Canada is
567 colored red, US blue, Europe green, Asia purple and the lone strains from Australia (MUNG6) and Chile
568 (CH811) is colored brown. **Figure 1B** Unrooted phylogeny based on the same tree.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014



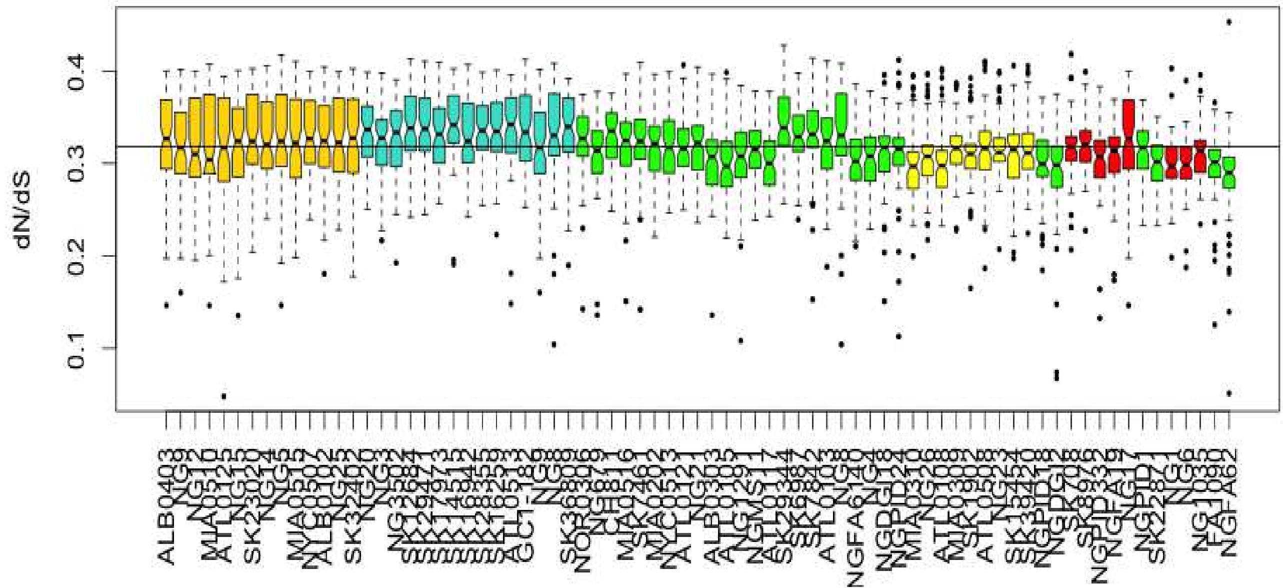
A **B**

Figure 2. (**A**) Recombination pattern traced from BLAST results of similarity of recombined regions between the subgroups defined by BAPS of *N. gonorrhoeae*. The clade showed as external represents strains from other Neisseria species. **(B)** Exchange of genetic materials among subgroups within the sample set as defined by BAPS admixture analysis. Colored base sub-sectors of the circle for each subgroup in the diagram represents outflow of genetic material while blank or white colored sub-sectors represent inflow of genetic materials to the subgroups



576 **Figure 3.** Population subgroups from strains or *N. gonorrhoeae* in the sample set defined by BAPS. The
577 names for each strain in the different sub group is indicated at the bottom of the plot on the x –axis, while
578 the fineSTRUCTURE group labels for each strain is indicated on top of the plot.

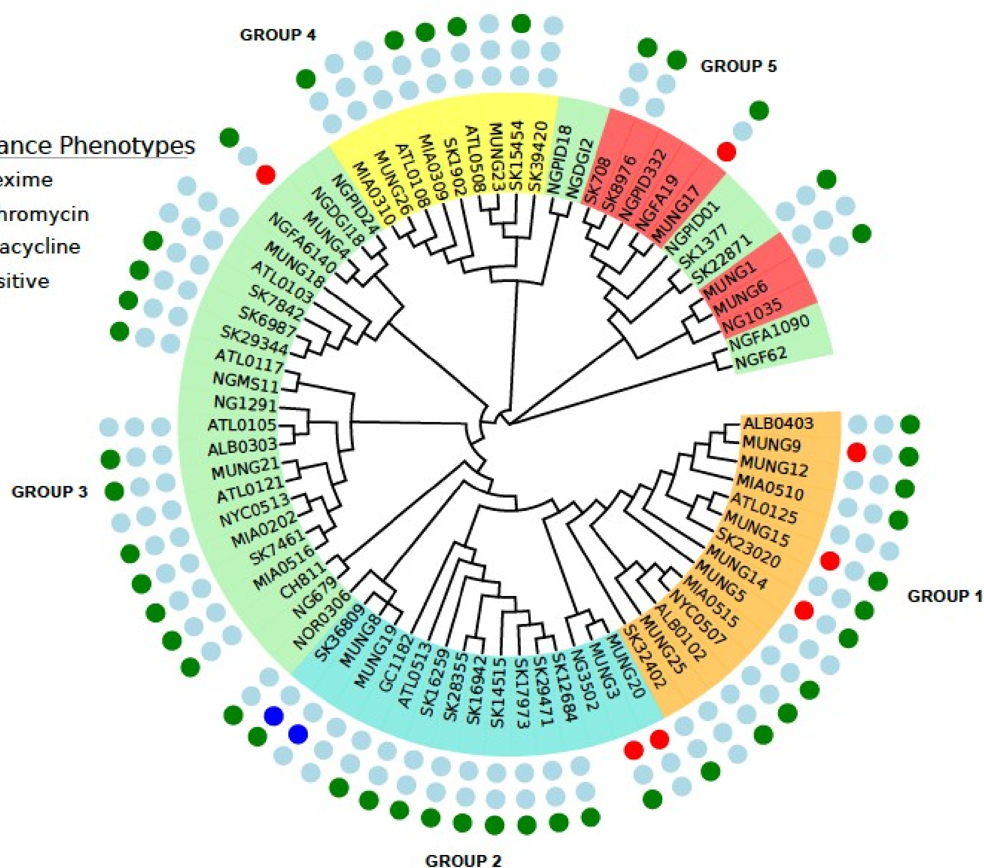
PeerJ PrePrints



579 **Figure 4.** Boxplot of mean dN/dS ratio pair-wise comparison of core genes of each of the strains of *N.*
580 *gonorrhoeae* in the sample set. The boxplot is colored by subgroups within the Neisseria population,
581 defined by the BAPS tool.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

PeerJ PrePrints



582 **Figure 5.** Representation of antibiotic resistance profile of *N. gonorrhoeae* strains across different
583 subgroups of the population. The topology is identical to the ML tree in Figure 1.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014

PeerJ PrePrints

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.636v1 | CC-BY 4.0 Open Access | rec: 26 Nov 2014, publ: 26 Nov 2014