Supplemental Text

Protocol for searching for known mutations

The resultant blast output was parsed for resistance genes mentioned in the ontology to confer antibiotic resistance, using our custom python script *blast_read.py.* We also trimmed the output file removing resistance genes that are in the core genome or present in all the strains, so we ended up with flexible genes which are present in only a few strains but have similarity with known resistance genes in other species which could have been horizontally transferred to the respective *N. gonorrhoeae* strains.

I grouped the strains in our sample set based on their antibiotics resistance phenotypes, for each of the groups of antibiotics (macrolides, cephalosporins, tetracycline) we used in determining the MICs for the strains. Each group is further divided into the strains that have the known mutation mutations (determined through the bioinformatics alignment search) underlying resistance to the respective antibiotic and those that do not have such mutations.

We used the NCBI blast tools and command line prompts in Unix to create a nucleotide blast database. This database is a multi-sequence fasta file of the WGS shotgun sequences of all the strains in our sample set assembled using the velvetg assembler. The contigs for each assembly was ordered into one pseudo contig after tiling to the reference genome FA1090, using the abacas.pl perl script.

Next we obtained reference NCBI sequence data of genomic regions encompassing resistance variants that have be shown in the literature to underlie the resistance phenotype we have observed in our sample set, and we performed a blast search for each of these sequences across all the strains in our database.

We picked the top hit for each sequence (strain) in the database and parsed the alignment between the query and the subject sequence in the database for the presence or absence of the underlying resistance genetic mutations as suggested in the literature using our custom perl script *parse.pl*. The results of our bioinformatics searches are shown in Supplement Data S3.

A number of the strains that have reduced susceptibility to the antibiotics drugs in the clinical data set do not possess the expected mutations within their corresponding resistance loci as suggested in previous literature. These results, which are shown in Table 4 suggested the possibility of epistatic interactions between either known variant sites or between known variants and novel loci to give rise to the resistance phenotypes.

Pangenome Analysis

The OrthoMCL module basically performs an all by all BLAST query for all the genes across all the strains define in our sample set. Orthologs and paralogs are defined as the best match for a given gene in another strain or in the same strain respectively. The best matches are grouped in clusters representing similar genes from across the different strains and forming the pangenome.

The result from the analysis maps well with the expected distribution of the pan genome into the core and accessory genes. The extended core genes that comprises of genes that have homologues in each of the strains, generally represent genes that are functionally relevant to the species. The accessory genes on the other hand are unique to each strain. A plot of the distribution of the clusters of orthologous genes show that majority of the genes in the metagenome are either accessory or part of the extended core genes. See Fig S1."

Our custom scripts for this project are in the Github repository: https://github.com/Read-Lab-Confederation/Neisseria_gonorrhoea_Population_Study

Supplemental tables

Strain Name	Accession Number	Total Reads Generated	Coverage	Contig Size
CH811	ATKL00000000	11,828,471	222	2,131,365
GC1-182	ATPK00000000	9,107,034	181	2,131,681
SK708	ATPE00000000	13,658,702	267	2,323,855
SK1902	ATPF00000000	3,400,264	293	2,179,833
SK6987	ATPG00000000	32,137,175	627	2,073,250
SK7461	ATPH00000000	7,960,584	155	2,146,692
SK7842	ATPI00000000	11,864,731	232	2,075,916
SK8976	ATPJ00000000	17,315,251	332	2,137,529
SK12684	ATPK00000000	17,965,229	357	2,082,127
SK14515	ATPM00000000	13,167,390	253	2,091,380
SK15454	ATPN00000000	15,461,938	305	2,133,996
SK16259	ATPO00000000	37,489,469	766	2,095,257
SK16942	ATPP00000000	16,855,240	324	2,090,527

SK17973	ATPQ00000000	25,041,823	486	2,085,242
SK22871	ATPR00000000	38,611,166	757	2,178,173
SK23020	ATPS00000000	9,662,895	191	2,145,793
SK28355	ATPU00000000	15,334,822	303	2,089,170
SK29344	ATPV00000000	12,367006	246	2,082,525
SK29471	ATPW00000000	22,462,999	428	2,081,649
SK32402	ATPX00000000	12,166,255	242	2,144,766
SK33414	ATPY00000000	12,282,205	235	2,229,613
SK36809	ATPZ00000000	18,031,339	343	2,150,262
SK39420	ATQA00000000	23,575,321	456	2,143,503
ALB0303	ATQB00000000	2,655,237	66	2,105,613
ALB0403	ATQC00000000	6,895,994	173	2,131,051
ATL0103	ATQD00000000	4,901,910	122	2,093,463
ALB0102	ATQE00000000	14,019,188	606	2,153,543
ATL0105	ATQF00000000	7,614,080	191	2,118,098
ATL0108	ATQG00000000	6,103,184	152	2,159,213
ATL0117	ATQH00000000	3,839,056	95	2,173,569

ATL0121	ATQI00000000	2,866,176	66	2,132,852
ATL0125	ATQJ00000000	1,950,942	84	2,137,996
ATL0508	ATQK00000000	5,671,170	242	2,144,219
ATL0513	ATQL00000000	5,062,764	219	2,101,891
MIA0202	ATQM00000000	7,846,178	195	2,123,083
MIA0309	ATQN00000000	7,335,289	181	2,167,271
MIA0310	ATQO00000000	8,780,248	380	2,175,084
MIA0510	ATQP00000000	3,771,306	169	2,139,760
MIA0515	ATQQ00000000	5,031,480	220	2,143,669
MIA0516	ATQR00000000	1,046,410	45	2,145,374
NOR0306	ATQS00000000	2,800,862	121	2,137,378
NYC0507	ATQT00000000	9,298850	233	2,128,515
NYC0513	ATQU00000000	8,453,765	210	2,130,515
MU_NG1	ATQV00000000	3,936,042	98	2,155,632
MU_NG3	TQW00000000	4,154,753	103	2,067,028
MU_NG4	ATQX00000000	3,267,675	81	2,104,187
MU_NG5	ATQY00000000	6,312,531	231	2,146,163

MU_NG6	ATQZ00000000	2,199,521	97	2,169,805
MU_NG8	ATRA00000000	4,824,331	121	2,081,555
MU_NG9	ATRB00000000	1,743,836	44	2,122,832
MU_NG12	ATRC00000000	3,749,254	93	2,128,018
MU_NG14	ATRD00000000	3,285,030	82	2,126,112
MU_NG15	ATRE00000000	4,842,133	121	2,127,213
MU_NG17	ATRF00000000	3,781,066	94	2,110,292
MU_NG18	ATRG00000000	4,075,444	102	2,108,750
MU_NG19	ATRH00000000	3,778,235	94	2,076,860
MU_NG20	ATRI00000000	1,273,718	32	2,054,548
MU_NG21	ATRJ00000000	3,730,151	94	2,167,411
MU_NG23	ATRK00000000	3,814,011	95	2,120,184
MU_NG25	ATRL00000000	2,068,833	95	2,135,960
MU_NG26	ATRM00000000	4,615,586	114	2,154,378

Table S1. Names, NCBI accession number and location of isolation of strains of *N. gonorrhoeae* as well as the number of illumina sequence reads, read coverage and assembled contigs size of genomes isolated and sequenced as part of our sample set.

Supplemental Figures

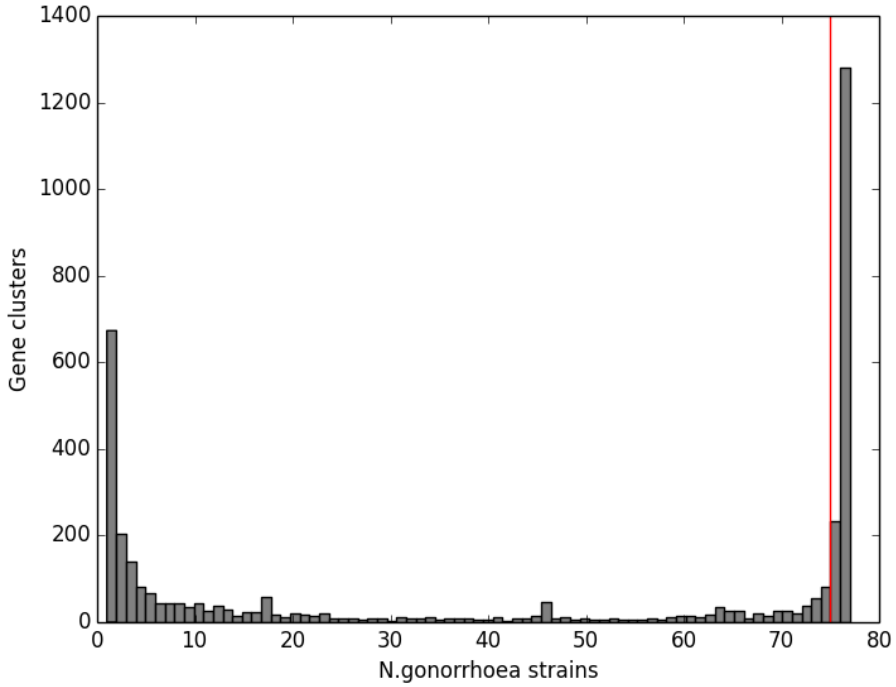

Figure S1**.** Representation of the pan-genome of strains of *N. gonorrhoeae* in the sample set. Each bar is a count of the number of genes (technically gene clusters) found in n genomes (n=81). Area to the right of the red line represents the core genes; to the left are the non-core genes.

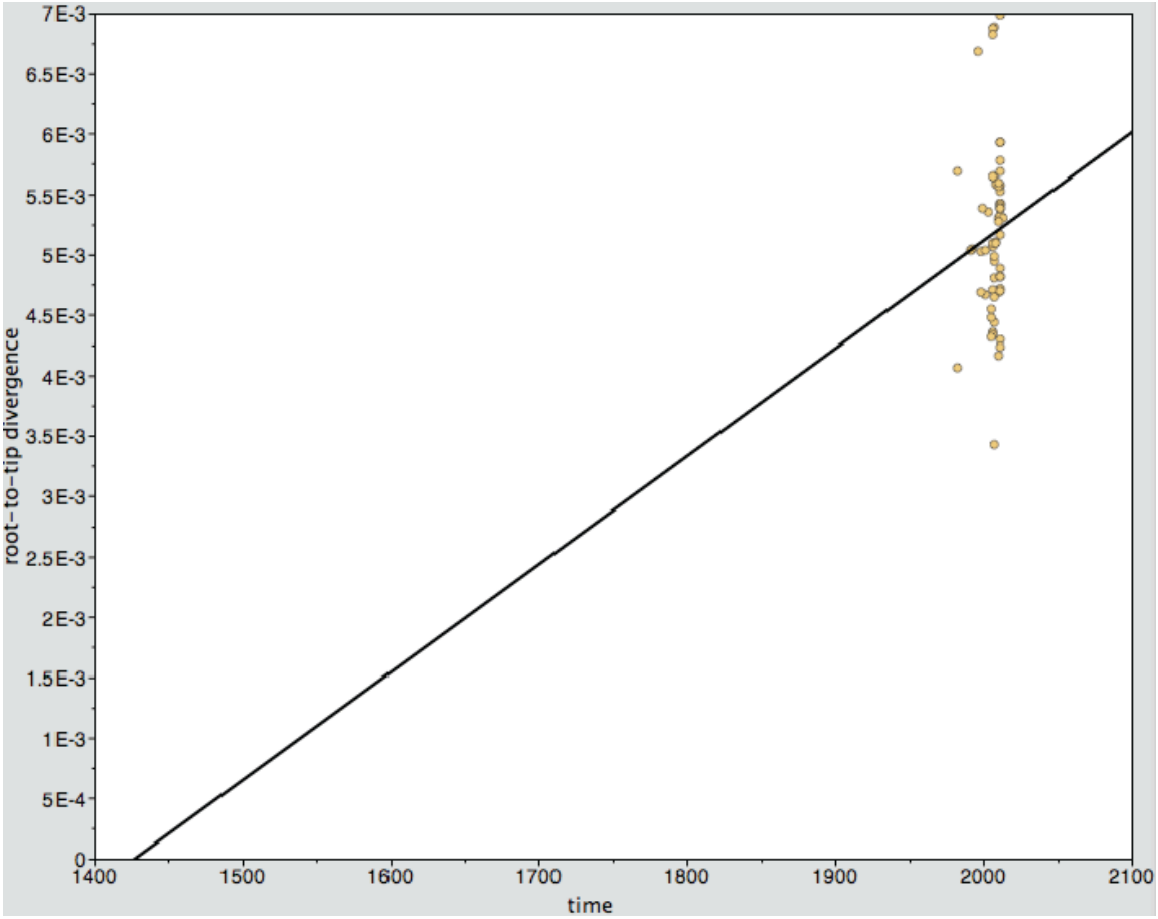

Figure S2. Root-to-tip plot as a measure of molecular rate of change in strains of *N. gonorrhoeae* species over time produced using the Path-oGen software

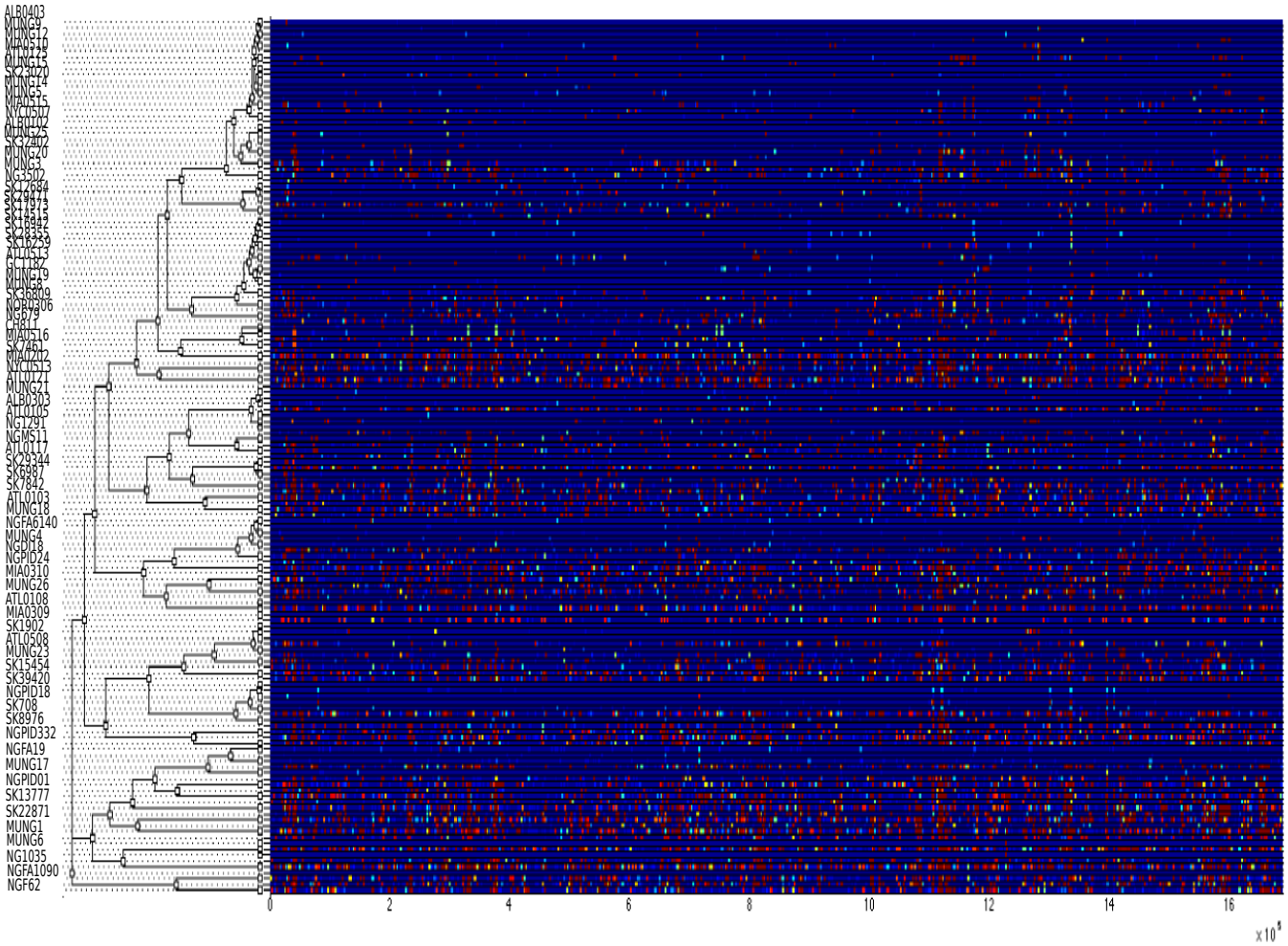

Figure S3. Heat map of pairwise recombination between strains of *N. gonorrhoeae* in the sample set. Putative recombination hotspots are on positions 8, 11 and 16 MB on the x-axis of the plot.

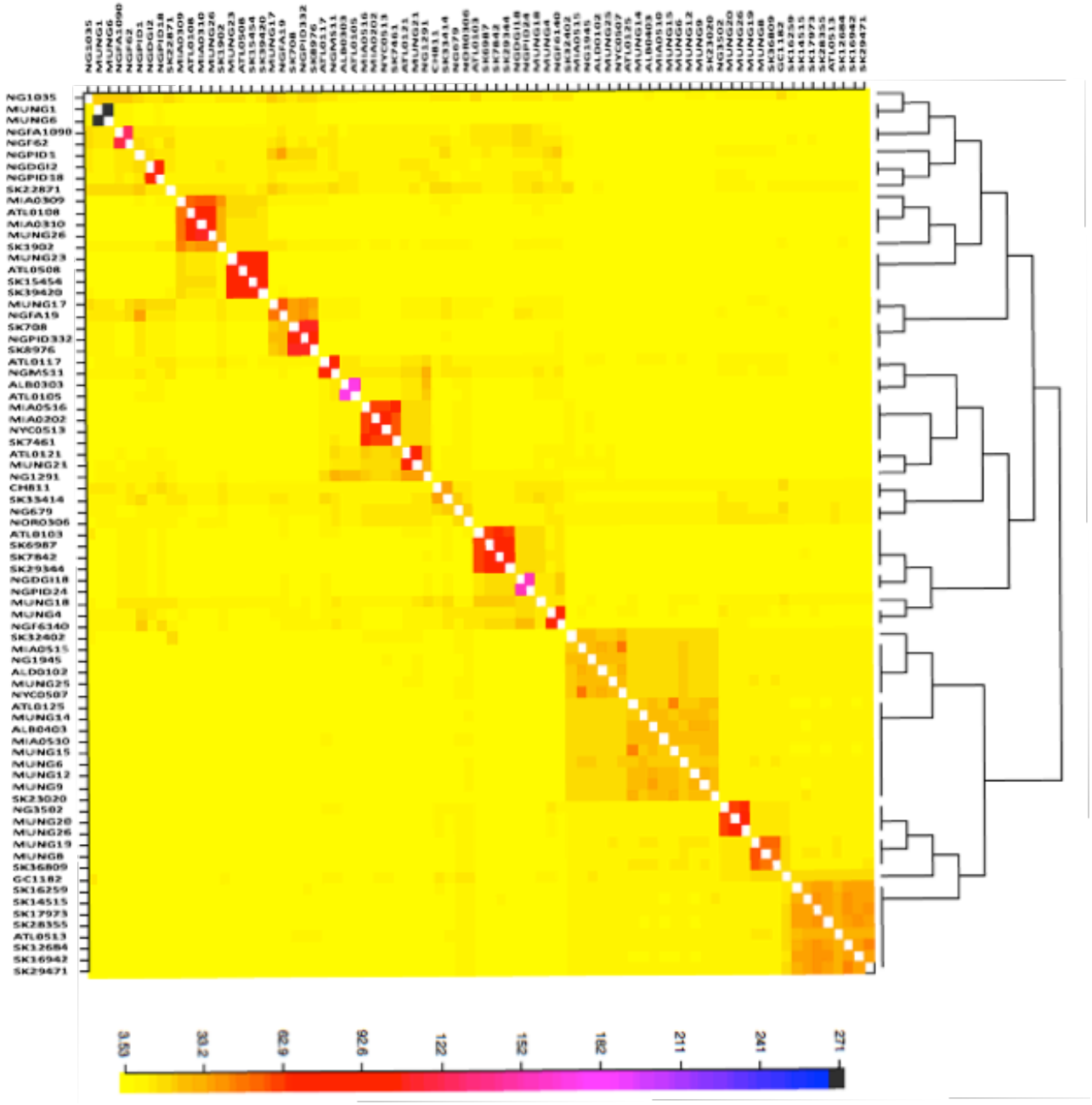

Figure S4 fineSTRUCTURE heatmap of *N. gonorrhoeae* population structure representing pairwise relationship between the strains in the sample set.

Figure S5 Phylogenetic representation of *N. gonorrhoeae* individual Sequence Types obtained from the MLST public database (http://pubmlst.org). The blue colored spots represent majority of the sequence types present in our sample set.