

Molecular Mimicry of Pathogenicity of *Neisseria*

Indrani Sarkar

Salim Ali Centre for Ornithology and Natural History: Salim Ali Centre for Ornithology and Natural History

Prateek Dey

Salim Ali Center for Ornithology and Natural History

Saurabh Singh Rathore

Mahatma Gandhi Central University

Gyan Dev Singh

Central University of South Bihar

Ram Pratap Singh (✉ rampratapsingh81@gmail.com)

Central University of South Bihar <https://orcid.org/0000-0002-3238-5086>

Research Article

Keywords: Evolution, Host-Pathogen Interaction, *Neisseria*, Reverse Ecology, Secretomes

Posted Date: December 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1045528/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Neisseria, a genus from beta-proteobacteria class, is of potent clinical importance. This genus contains both pathogenic and commensal strains. Gonorrhoea and meningitis are two major diseases caused by pathogens belonging to this genus. With increased use of antimicrobial agents against these pathogens they have evolved the antimicrobial resistance (AMR) capacity making these diseases nearly untreatable. The set of anti-bacterial resistance genes (resistome) and genes associated with signal processing (secretomes) are crucial for the host-microbial interaction. With the virtue of whole genome sequences and computational biology it is now possible to study the genomic and proteomic riddles of *Neisseria* along with their comprehensive evolutionary and metabolic profiling.

We have studied relative synonymous codon usage, amino acid usage, reverse ecology, comparative genomics, evolutionary analysis and pathogen-host (*Neisseria*-human) interaction through bioinformatics analysis. Our analysis revealed the co-evolution of *Neisseria* genomes with the human host. Moreover, co-occurrence of *Neisseria* and humans has been supported through reverse ecology analysis. A differential pattern of evolutionary rate of resistomes and secretomes was evident among the pathogenic and commensal strains. Comparative genomics supported the presence of virulent genes in both pathogenic and commensal strains of select genus. Our analysis also indicated a transition from commensal to pathogenic *Neisseria* strains through the long run of evolution.

Introduction

Neisseria is a large microbial genus from Beta-proteobacteria class mainly colonizing in the mucosal surfaces of human and other animals like guinea pig and dog (Elias et al. 2015). Several strains of this genus have been isolated from brain, oral cavity, respiratory tract and uro-genital tract of humans (Elias et al. 2015). This genus consisted of both commensal and pathogenic strains. For instance *N. lactamica* (Snyder and Saunders, 2006), *N. polysaccharea* (Albenne et al. 2004) are reported to be commensal strains whereas *N. meningitis* (Stephens et al. 2007) and *N. gonorrhoeae* (Kellogg et al. 1963) are potent pathogens. *N. meningitis* has been reported to cause significant mortality and morbidity among both young adults and children worldwide through the spread of epidemic or sporadic meningitis and septicemia (Rouphael et al. 2012). A sexually transmitted infection (STD) resulting into infertility (if remain untreated) is caused by asymptomatic infection of *N. gonorrhoeae* ascending the genital tract and disseminates to distal tissues (McSheffrey and Gray-Owen, 2015). The global rate of both these diseases are increasing with the development of multidrug resistance capacity of the causative pathogens. Continuous exposure of broad spectrum and last-resort antimicrobial agents like Penicillin, Colistin and Carbapenems has developed the anti-microbial resistance (AMR) property among *Neisseria*. This has recently afforded these bacteria a 'superbug' status (McSheffrey and Gray-Owen, 2015). The evolution of *Neisseria genomes* as AMR strain is a journey of only ~80 years (Yang et al. 2020) and already it has become resistant towards most of the broad-spectrum antibiotics. Recently it was reported that, *N. gonorrhoeae* showed refusal to Sulfonamides, Penicillin, TEM-type β -lactamases, Tetracycline, Spectinomycin, Cephalosporins and Colistin (Yang et al. 2020). Ciprofloxacin-resistant *N. meningitidis* was found in 5-month-old boy in the United States (Taormina et al; 2021). These clinical reports have raised a concern and possible threat of untreatable Gonorrhoea and Meningitis in near future.

The virulent genes causing the disease are transmitted from pathogens to hosts. Whenever a gene is transferred from one organism (donor) to another (host), its codon usage which is adapted to the genomic context of the donor might not work optimally in the host genome hampering its expression (Grantham et al. 1980, Garcia-Vallve et al. 2003). The transferred gene must express itself properly to show its effect on the host and any mutation that leads to excess expressivity of that gene in the population will be favoured. This would increase the fixation probability of that transferred gene in the host population. Mostly the transferred genes undergo an amelioration process where their codon usage composition drastically shifts from the donor's genomic context towards recipients' genomic context (Amorós-Moya et al. 2010). Such genes are often present in pathogenic islands with medical significance and play pivotal roles in host-microbe interaction (Schmidt and Hensel, 2004). Since the genus *Neisseria* is composed of both obligate and facultative pathogens as well as

commensal bacteria, understanding the mechanisms that shapes the evolution of pathogenic and AMR genes will shed new light on the dynamics of their host- microbe interaction (Nakamura et al. 2004). Moreover, the commensal strains of *Neisseria* are not free living but are host dependent. A huge chunk of the pathogenicity and AMR related proteins are detected in those non-pathogenic strains (Calder et al. 2020). This has raised a question of their genomic evolution- whether there was a transmission from pathogens to commensalism or vice versa.

The advancement of next generation sequencing has availed us with the whole genome sequences of twenty-one different *Neisseria* species. Among them *N. animalis* (Clemence et al. 2018), *N. brasiliensis* (Mustapha et al. 2020), *N. canis* (Safton et al. 1999), *N. elongata glycolytica* (Andersen et al. 1995), *N. elongate* (Andersen et al. 1995), *N. gonorrhoeae* (Kellogg et al. 1963), *N. meningitidis* (Stephens et al. 2007), *N. weaver* (Andersen et al 1993), *N. zalophi* (Volokhov et al. 2018) and *N. zoodegmatis* (Heydecke et al. 2013) are potent pathogens. Some of these aforementioned species were previously considered as commensals (Calder et al. 2020) however, there are several clinical reports supporting them as pathogens. Two species *N. flavescens* and *N. mucosa* were reported to be opportunistic pathogens (Huang et al. 2014; Mechergui et al. 2014). In healthy humans, they remain as non-pathogens though in immune-compromised patients they behave as pathogens. *N. chenwenguii*, *N. cinerea*, *N. lactamica*, *N. musculi*, *N. polysaccharea*, *N. sicca*, *Neisseria sp. KEM232*, *N. subflava* and *Neisseria sp. 10022* are largely entitled as commensals (Calder et al. 2020; Kim and Seong, 2018). Although there are some reports that, *N. lactamica* and *N. sicca* being occasional pathogens, further confirmations are needed in those aspects (Hansman 1978; Gris et al. 1989). Complete genomes of all these aforementioned species are now available in the public domain database. Computational biology targeting codon usage, mRNA expression pattern, evolutionary analysis and network modelling have become popular tools to study differential microbial genomics as well as host-microbial interaction. In this study we have exploited these well-established methods to explore the genomic riddles of *Neisseria* and their interaction with the human host.

Materials And Methods

Sequence retrieval and Phylogeny construction

Complete genome sequences of twenty-one *Neisseria* type strains were downloaded from Integrated Microbial Genomics (IMG) database (Markowitz et al. 2012). Each of these strains represented 21 different species of this genus. Necessary information about their habitat, pathogenicity, total number of protein coding genes, and KEGG Ontology (KO) were obtained from the same IMG database. Five housekeeping genes (*dnaA*, *ftsZ*, *secA*, *atpB*, *gyrB*) were used for generating MLSA phylogenetic tree. For Multi-locus sequence alignment (MLSA) phylogeny the translated protein sequences of aforementioned housekeeping genes were concatenated in *dnaA-ftsZ-secA-atpB-gyrB* order and were aligned in ClustalW (Thompson et al. 2003). The aligned file was imported in Mega X software (Kumar et al. 2018) and a 1000 bootstrap Neighbour-Joining (NJ) phylogeny was generated.

Prediction of *Neisseria* Resistome

The Comprehensive Antibiotic Resistance Database (CARD) was used for the prediction of antibiotic resistance genes from the studied strains (McArthur et al. 2013). We exploited the Resistance Gene Identifier (RGI) to predict the resistomes (set of antibiotic resistance proteins) from the complete protein sequence of the select *Neisseria* strains. We choose the 'perfect and strict hits only' and discarded the loose hits to minimise the false negative result in our analysis. High quality coverage 'excluding nudge' was set as parameter criteria for resistome prediction. To further validate the CARD result we used 50/50 blastP algorithm. Previous reports on *N. gonorrhoeae* mentioned nine proteins as antibiotic resistance (Unemo and Shafer 2014). Those proteins were used as a database and blastP was run against our considered proteomes. Proteins with best hits (98-100% similarities) were considered as resistance genes.

Prediction scheme for identification of *Neisseria* secretomes

Secretomes can be defined as a set of proteins involved in the cellular cross talk and communication with the surrounding environment as well as host. A multi-step prediction scheme as described by Cornejo-Granados et al. (2017) and Roy et al. (2018) was followed to predict the *Neisseria* secretomes. Four different prediction servers - SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP-4.1/>), SecretomeP 2.0 (<http://www.cbs.dtu.dk/services/SecretomeP/>), LipoP 1.0 (<http://www.cbs.dtu.dk/services/LipoP/>) and TatP1.0 (<http://www.cbs.dtu.dk/services/TatP/>) were first employed to extract the different types of signal peptides. For instance, classical secretory proteins were identified through SignalP 4.1. SecretomeP 2.0 identified the non-classical secretory components. Lipoprotein type of signal peptides (Lipo type) were predicted through LipoP 1.0. Proteins that were found to be “cytoplasmic” in LipoP 1.0 were discarded from further studies. The twin-arginine (RR/KR) signal peptides (TAT), which are involved in the TAT signalling pathway, were identified through TatP1.0. Positive sequences from all these four servers were fed into TMHMM (v2.0) (<http://www.cbs.dtu.dk/services/TMHMM/>) for predicting the number of transmembrane (TM) helices. Proteins with no TM helix were directly considered to be part of the secretome set. Proteins with one or more than one TM helices were further screened through Phobius signal peptide predictor (<https://phobius.sbc.su.se>). Sequences with positive Phobius result further enriched the secretome set of *Neisseria*.

Identification of potentially virulent secretory proteins

VirulentPred server (Garg and Gupta, 2008) was used to identify the potentially virulent proteins among the predicted secretome set of select *Neisseria*. Cascaded SVM classifier which is employed in the mentioned server has been reported to be highly accurate in predicting virulent genes among prokaryotes. This approach helped us in prediction of those *Neisseria* signal peptides which are directly associated with the infection cycle in the human host.

Reverse ecology analysis between *Neisseria* and *Homo sapiens*

In simple terms, reverse ecology is the study of ecology from the genomic data without any prior knowledge or assumption regarding the ecological interactions among considered organisms (Levy and Borenstein 2012). RevEcoR package of the R program was used to perform this analysis. RevEcoR takes account of the KEGG Ontology (KO) profiling of organisms and predicts the interaction among them based on ‘competition index’ and ‘Cooperation index’. The IMG database listed all the KO ids for each of the selected *Neisseria* strains and *Homo sapiens*. Total 8789 KO ids were there for *H. sapiens*. The reverse ecology was performed through RevEcoR package implemented in R software (ver 4.1.0) (Chambers, 2008). The effect of both pathogenic and non-pathogenic *Neisseria* was observed on their human host.

Codon usage analysis

The differential use of synonymous codons among different organisms is termed as codon usage bias (CUB). We calculated the relative synonymous codon usage (RSCU) of select *Neisseria* to explore their CUB pattern. Codon usage signature of a pathogen can be shaped by multiple determinants including the mutational pressure and translational selection constraint exerted by the host. Hence, pathogenic microbes are reported to co-evolve and mimic the codon usage pattern of respective hosts for efficient utilisation of host tRNA pool along with other resources (Butt et al. 2016). Adaptation within the host is important for pathogenic fitness and survival within the host body since high competence inside host body will increase the magnitude of the infection caused by the microbe (Butt et al. 2016). To assess the co-evolution pattern between *Neisseria* (both pathogenic and nonpathogenic) and *H. sapiens*, we compared the CUB from both. A similar pattern of CUB would indicate co-evolution between the studied pathogen and their respective host.

Along with CUB, other important factors related to codon usage indices like, GC, GC3, Fop (frequency of optimal codons), tRNA adaptation index (tAI), effective number of codons (ENC), codon adaptation index (CAI) were calculated for the select *Neisseria* strains. Spearman Rank correlation was calculated with SPSS ver 26.0 among aforementioned codon usage indices. This gave an overall idea about the genomic composition and codon usage pattern of *Neisseria*.

Prediction of protein energy cost

Biosynthesis of amino acids requires energy. Protein energy cost (PEC) can be defined as the total amount of energy (in terms of ATP or GTP) consumed for the biosynthesis of each amino acid (Akashi and Gojobori, 2002). We used Dambe software (Xia and Xie, 2001) to assess PEC for the select strains. PEC for the secretomes of *Neisseria* was further correlated with their CAI to reveal whether the secretomic energy consumption is directly or inversely related with their expression pattern.

Protein-Protein Interaction study

Human proteins related to gonorrhoea and meningitis were obtained from DisGenet server (Piñero et al 2016) with a cut-off value 0.7. The interaction between those genes and pathogenicity related *Neisseria* genes was obtained from STRING server (Snel et al. 2000). The enrichment analysis was continued till the PPI network showed significantly higher interaction than expected. The STRING network was exported to Cytoscape 3.8.2 (Smoot et al. 2011) for network generation.

Comparative genomics

CMG Biotoools (Vesth et al. 2013) was used for the whole genomic comparison among select microbes. Blast metric was performed based on the 50/50 blastP program. This analysis was important from two aspects. First, it showed the inter-proteomic similarities among different species of *Neisseria*. Second, the duplication level within a proteome could also be revealed (Vesth et al. 2013).

Pan-Core genome analysis is important for comparative genomics analysis since it exposes the unique and shared protein set among studied organisms. Pan genome can be described as the total genomic set of all considered genomes whereas the core genome is the set shared among all investigated strains (Vesth et al. 2013). Along with the pan-core genome analysis a core genome-based phylogeny was also generated and compared with other phylogenies that we build in this study (as mentioned above).

Assessment of evolution of Neisseria

The ratio (ω) between the rate of non-synonymous substitutions per non-synonymous site (K_a) to the rate of synonymous substitutions per synonymous site (K_s) (Nekrutenko et al. 2002) was used as an evaluator for the evolutionary rate of *Neisseria*. We used the Codeml program included in the PAML software package (ver. 4.5) to assess the ω value (Yang et al. 1997). $\omega > 1$ indicated positive Darwinian selection and $\omega < 1$ stood for negative purifying selection. We divided our select strains into pathogenic and non-pathogenic sets. The evolutionary analysis was performed separately for pathogenic and non-pathogenic strains.

Results

Codon usage analysis of *Neisseria*

The genomic constitution of *Neisseria* revealed that, this genus is neither GC biased (GC% ranged from 48-52%) nor AT biased (AT ranged from 49-53%). A significantly negative correlation ($r = -0.67$; $p < 0.01$) between ENc and CAI was found among all studied strains. The value of ENc ranges from 20-62 and a negative correlation between ENc and CAI indicates the pivotal roles of codon usage indices other than compositional constraint. To evaluate the variation at the third position of the codons with the expression level of the genes, A3, C3, G3 and T3 was correlated with CAI. A positive correlation ($r = 0.86$; $p < 0.001$) between CAI and C3 was observed among all select strains. To assess the codon bias nature of *Neisseria*, Fop was correlated with CAI. A positive correlation between CAI and Fop ($r = 0.84$, $p < 0.001$) was found.

Evaluation of protein energy cost (PEC) among Neisseria

Spearman Rank correlation between PEC, CAI and Fop was calculated for both potentially highly expressed (PHX) genes and PLX (potentially lowly expressed). A positive correlation between CAI and Fop ($r=0.84$, $p<0.001$) along with negative correlation among PEC and CAI ($r=-0.76$, $p<0.001$) as well as Fop and PEC ($r=-0.68$, $p<0.001$) was obtained for PHX proteins. On the contrary, positive correlation was found between Fop and PEC ($r=0.34$, $p<0.01$) as well as CAI and PEC ($r=0.36$, $p<0.01$).

Host-microbe interaction strategy of Neisseria

While assessing the relative synonymous codon usage pattern (RSCU) among select *Neisseria* strains, we found fourteen codons (ATC, TAC, TTC, GCC, CTG, TCC, TGC, CAC, AAC, ACC, GGC, GTC, CCC, GAC) were optimally used in *Neisseria*. Interestingly those are optimally used in human also indicating towards a host-microbe interaction strategy for *Neisseria* in terms of codon-optimization and codon co-evolution. Further study on the host-pathogen interaction was performed using Reverse ecology analysis. The complementation indices among considered microbes were found to be very low. However, all select strains showed complementation (0.26-0.44) with humans.. A metabolic reconstruction was done on the basis of this reverse ecology analysis (Fig. 2c). There were 17677 edges in the constructed metabolic network directly connected to each other (blue circles). Only forty “seed” (red circles) were found in the metabolic map (Supplementary file 1). The term “seed” represents the exogenously acquired compounds required for the metabolism of an organism. “Edges”, on other hand, represents the compounds or chemical resources being used by all studied organisms.

Comparative genomics analysis among pathogenic and commensal Neisseria

A total of 10296 genes accumulated in the pan-genome set and 847 genes were in the core set. The blast matrix and pan-genomic dendrogram analysis revealed a dissimilarity of *N. musculi* NW831 from other *Neisseria* strains (Fig. 3a, 3b, 3c). Two pathogens *N. elongata* glycolytica ATCC 29315, *N. elongata* M15910 and one non-pathogen *Neisseria* sp. KEM232 were clustered together (cluster -I). They were sharing 52-66% proteomic similarities among each other. In another group (cluster -II), four pathogenic strains *N. weaveri* NCTC 13585, *N. zoodegmatidis* NCTC 12230, *N. zalophi* ATCC BAA-2455 and *N. canis* NCTC10296 were placed together with 47.7-57.5% proteomic similarities. In another cluster pathogenic *N. brasiliensis* N.177.16 was placed with *Neisseria* sp. 10022 (still not reported as pathogen) with 47.5% similarities (cluster-IV). The major cluster (cluster-III) of the dendrogram was a mixed bag consisting of both pathogenic and commensal strains. We subdivided this cluster into sub-clusters. In sub-cluster b, commensal strains like *N. cinerea* NCTC 10294, *N. lactamica* 020-06, *N. polysaccharea* M18661 were present. Unlikely, *N. meningitidis* LNP21362 and *N. gonorrhoeae* FQ01 were also present within this sub-cluster. The non-pathogens in this sub-cluster shared 64-65% similarities among their proteome set. *N. meningitidis* LNP21362 shared 60.7%, 61.7% and 62.7% proteomic identity with *N. polysaccharea* M18661, *N. lactamica* 020-06 and *N. cinerea* NCTC 10294 respectively. *N. gonorrhoeae* FQ01 was found to be 56.6% similar with *N. lactamica* 020-06, 53.1% identical with *N. cinerea* NCTC 10294 and 51.8% similar with *N. polysaccharea* M18661.

In the next subcluster (sub-cluster a) a hybrid (opportunistic pathogen) *Neisseria* strain (both pathogenic and commensal), *N. flavescens* ATCC 13120 grouped with commensal *N. subflava* ATCC 49275 (66.3% similarities). Similarly, the following sub-cluster contained one non-pathogen *N. sicca* FDAARGOS_260 and one opportunistic pathogen *N. mucosa* ATCC 19696 sharing 81.9% proteomic identity. Potent pathogen *N. animalis* NCTC 10212 was close to them within the same cluster (43.5-50.0% similarities). Another commensal strain *N. chenwenguii* 10023 was also placed within the same cluster (45.0-49.7% proteomic identity). *Neisseria* sp. KEM232 was described as a novel species of *N. chenwenguii* (Al Suwayyid et al. 2020) however, our pan genomic dendrogram and blast matrix showed that they were distantly related with only 30.1% proteomic similarities. To further investigate this issue, we performed Average Nucleotide Identity (ANI) score analysis. It has been suggested that, ANI score >95% between two genomes indicates they are the same species. The ANI score between *Neisseria* sp. KEM232 and *N. chenwenguii* 10023 was only 77.86%. Moreover, the ANI score between *Neisseria* sp. KEM232 and *N. elongata* glycolytica ATCC 29315 was 80.4%. The same score between *Neisseria* sp. KEM232 and *N.*

elongata M15910 were 80.6%. This result not only supported our blast matrix and pan-genomic dendrogram but also suggested a taxonomical reconsideration for *Neisseria* sp. KEM232. MLSA phylogeny showed an identical clustering pattern among considered strains (Fig. 3d).

Evolutionary analysis of *Neisseria*

The rate of evolution among protein coding genes varies tremendously. Evolutionary analysis based upon k_a/k_s (or ω) value revealed a differential decoration among diverse sets of genes. It was found that PHX genes (mean ω 0.07) were less evolved ($p < 0.001$) and more conserved than PLX genes (mean ω 0.32). The dN/dS analysis of *Neisseria* secretomes revealed their faster evolution than PHX in all investigated strains ($p < 0.001$). Moreover, the pathogenic secretomes (mean ω 0.21) were evolving at a faster rate ($p < 0.001$) than commensal strains (mean ω 0.14).

Discussion

1. Genomic constitution of *Neisseria* is multi-factorial and complex

The GC/AT content of organisms is one of the most highly variable traits (Botzman and Margalit, 2011). Variation in nucleotides can be observed in protein coding genes, non-protein coding genes, synonymous sites as well as non-synonymous sites of genomes (Reis et al. 2004). The genomic constitution of *Neisseria* revealed that this genus is neither GC biased nor AT biased. Previously it has been reported that, mutational biases determine the nucleotide composition which means GC biased mutation pattern resulted in GC rich organisms and AT biased mutation developed AT rich organisms (Hershberg and Petrov, 2010). However, recently the in-effectivity of mutational bias in directing the nucleotide composition has been revealed. This happens in organisms which are both AT and GC rich (Hildebrand et al. 2010). This also happened in the case of *Neisseria*. In such organisms, mutations are more likely to happen from G/C towards A/T due to the rapid deamination of cytosine to thymine (C > T/U) (Bohlin et al. 2017). However, increased AT content indicates genomic disability (Yakovchuk et al. 2006). In such situations, another selectively neutral pressure termed as 'amelioration' acts as a major force which even out the differences in base composition (Lawrence and Ochman, 1997). Such cases are beneficial for pathogens with an AT rich host (human for our case) (Bohlin 2011). Thus, the genomic organization of *Neisseria* does not directly depend solely upon its nucleotide variation. Instead, it is a multi-factorial process that is resulted through a complex combination of both neutral and selective processes. This was also validated by the significantly negative correlation ($r = -0.67$; $p < 0.01$) between ENc and CAI. The value of ENc ranges from 20-62 and a negative correlation between ENc and CAI indicates the pivotal roles of codon usage indices other than compositional constraint. To determine those factors, we used the Spearman rank correlation method. A positive correlation ($r = 0.86$; $p < 0.001$) between CAI and C3 was observed among all select strains. Preference towards Cytosine (C) in AT/GC unbiased organisms further strengthens the role of mutational pressure on these genomes. The third position of the codon is a hotspot for random mutation without a drastic effect on amino acid usage due to the redundancy of codons. Thus, it can be posulated that, deamination of C to Uracil (U) in *Neisseria* not only increases the AT richness of the genome but also helps this genus to aptly use the host translational machinery using the human tRNA pool. The presence of active cytidine deaminase responsible for C>U mutation has been reported in pathogens like *E. coli* and *S. typhimurium* (Henderson and Paterson, 2014). Such activity may also be present in *Neisseria* with a pivotal role in their genomic constituency. This arrangement of codon usage indices was found to be consistent among the secretomes and pathogenicity related genes present in *Neisseria*. Moreover, no significant difference was found in this pattern between the pathogenic and non-pathogenic *Neisseria* strains.

2. Optimal codons are mounting the quantity of energy economic amino acids in *Neisseria*

Another factor found to play an important role in *Neisseria* genomes was Fop. A positive correlation between CAI and Fop ($r = 0.84$, $p < 0.001$) supported the previous statement. This value indicated higher usage of optimal codons in potentially

highly expressed genes than lowly expressed genes. Twenty-nine codons were found to be optimal codons. Among them nineteen were GC rich codons and fifteen ended with Cytosine (C). This leads to an important aspect regarding the amino acid usage of select genomes. GC rich codons code for more energy economic amino acids rather than AT rich codons (Bohlin et al. 2017). Thus, the higher usage of such GC rich optimal codons in PHX genes indicated less biosynthetic energy cost for respective translated proteins. To further assess these findings the correlation between PEC, CAI and Fop was calculated for both PHX and PLX. A positive correlation between CAI and Fop ($r=0.84$, $p<0.001$) along with negative correlation among PEC and CAI ($r=-0.76$, $p<0.001$) as well as Fop and PEC ($r=-0.68$, $p<0.001$) was obtained for PHX proteins. On the contrary, positive correlation was found between Fop and PEC ($r=0.34$, $p<0.01$) as well as CAI and PEC ($r=0.36$, $p<0.01$). This result indicated that, although the *Neisseria* genomes are not generally biased towards either AT or GC rich codons (Fig. 1a), natural selection is discriminating among the synonymous codons and preferring GC rich codons in PHX genes. This enhances the translational elongation rates as well as reduces misincorporation of amino acids during protein synthesis (Akashi and Gojobori, 2002). Previously, Akashi and Gojobori (2002) reported a relation between the protein energy cost (PEC) and tRNA adaptation in differentially expressed genes. Hence, we correlated PEC and tAI. We found a negative correlation ($r=-0.75$, $p<0.001$) between them among PHX. This further validated the translational efficiency of potentially highly expressed genes along with an indication that afterwards these genes will be translated into energy economic proteins with higher expression level. An overall amino acid usage (Fig. 1b) calculation indicated alanine, valine, glycine, serine, asparagine, proline, glutamic acids and threonine were highly used in *Neisseria*. The overall usage of costly aromatic amino acids like phenylalanine, tyrosine and tryptophan were comparatively lower than the aforementioned amino acids. No significant difference was found in this pattern between the pathogenic and commensal *Neisseria* strains.

3. RSCU pattern indicated towards co-evolution *Neisseria* for better host adaptation

Previous studies have shown the relation between codon adaptation and ecological preferences (Peden 1998). A relation between the codon adaptation and co-evolution has also been drawn. To assess the co-evolutionary pattern between *Neisseria* and their host *Homo sapiens*, their RSCU pattern was exploited. We found fourteen codons (ATC, TAC, TTC, GCC, CTG, TCC, TGC, CAC, AAC, ACC, GGC, GTC, CCC, GAC) were optimally used in both *Neisseria* and human. Moreover, ~96% pathogenic island genes in *Neisseria* were under the PHX category. This suggested elevated translational efficiency of those genes in the host body. The translational selection pressure towards these fourteen most adapted codons aided the microbes to live in the host environment and efficiently utilize their metabolic resources (Botzman and Margalit, 2011). Thus, the codon usage is playing a pivotal role in enhancing the cellular fitness of *Neisseria* within the host body mostly by mimicking the codon usage pattern of humans.

4. Co-existence of *Neisseria* with human host

The genus *Neisseria* composed of both pathogenic and non-pathogenic commensal bacteria. According to the ecological principles, co-existence can be ruled either via competition or complementation (Carr and Borenstein 2012; Levy and Borenstein 2012). The reverse ecology analysis among select *Neisseria* and their host (human) revealed inter-species specific and intra-species-specific competition among members of *Neisseria* (Fig. 2a, 2b). The pathogenic strains were exerting more competition on commensal strains. Both types of strains were found to exert a moderate competition against humans dictating an efficient distribution of host-derived resources among pathogenic and commensal *Neisseria* (Fig. 2a). However, the competition exerted by humans on *Neisseria* was diminutive. This has turned humans into the perfect host for this microbial genus.

The complementation indices among considered microbes were found to be very low. Thus, co-existence of different *Neisseria* strains in a small niche can be least expected. This also explains the broad range of distribution (for example, brain, oral cavity, respiratory system, reproductive system, urinary tract etc.) of *Neisseria* within the human body. However, all select strains showed complementation (0.26-0.44) with humans. This metabolic reconstruction clearly (Fig. 2c)

depicted that, large number of resources are shared and utilized efficiently between humans and *Neisseria*. This suggested the co-inhabitation of *Neisseria* within the human body is ecologically favorable.

5. Differential evolutionary pattern indicated transition from commensalism to pathogenicity among *Neisseria*

The rate of evolution among protein coding genes varies tremendously. Evolutionary analysis based upon ka/ks (or ω) value revealed a differential decoration among diverse sets of genes. It was found that PHX genes were less evolved ($p < 0.001$) and more conserved than PLX genes. The 'knock-out rate' prediction proposed that most of the PHX genes are essential or housekeeping genes with important functionality (Hust and Smith, 1999). These essential genes evolve more slowly than other non-essential genes (Wilson et al. 1977). Similar results were also found previously in *Escherichia coli*, *Helicobacter pylori* and even in *Neisseria meningitidis* (Jordan et al. 2002). Moreover, secretomes of pathogens continuously struggle with the host immune system and try to beat it which resulted in their faster evolution (Ehrlich et al. 2008; Saha et al. 2019). This differential evolutionary pattern for pathogens indicated the possibility for emergence of pathogenicity from commensalism among *Neisseria*.

Another aspect of our ka/ks analysis was based on pathogenicity related (PI) genes. We found a set of pathogenic genes were present in non-pathogenic *Neisseria* strains which was unexpected. Few studies (Calder et al. 2020; Lu et al. 2019; Clemence et al. 2018) on *Neisseria* have also reported these surprising results where the potent virulent genes of *N. meningitidis* and *N. gonorrhoea* were found in nonpathogenic *N. lactemia* (Snyder and Saunders, 2006). However, no clear explanation for this result is still stated. Hence, we calculated the evolutionary rates of PI genes from both pathogenic and non-pathogenic strains to reveal whether a transition from pathogenicity to commensalism has occurred during evolution of *Neisseria* or vice-versa. The PI genes were highly ($p < 0.001$) evolved in pathogenic strains rather than their nonpathogenic counterparts. The ω value of pathogenic PI genes ranged from 0.32-0.45 whereas the same for non-pathogenic strains ranged from 0.05-0.08. Their difference was statistically significant ($p < 0.001$). Hence the transition from commensalism to pathogenicity in *Neisseria* is evident from this result. This type of transition was previously reported in *Mycobacterium avium* complex (Saha et al. 2019). Moreover, nine protein coding genes have been reported to be associated with antimicrobial resistance for *N. gonorrhoea*. Orthologs of those genes were found in all considered strains. Evolution analysis among them predicted their higher evolution in pathogens rather than nonpathogens. The mean ka/ks value for each of the nine genes were lowest when only non-pathogenic strains were studied. The rate of evolution increased when we considered both pathogen and non-pathogens (strains from cluster III from pan-genomic dendrogram) together and the value was highest after only pathogens were considered (Fig. 4).

This supported our aforementioned hypothesis for transition from commensalism to pathogenicity among *Neisseria*. With the emergence of pathogenicity this genus became exposed to both narrow- as well as broad-spectrum antibiotics and in the long run their anti-microbial resistance property evolved.

6. PPI study of *Neisseria*-Human interaction

Protein-protein interaction (PPI) analysis has become a major tool in system biology with its ability to handle a broad range of data related to biological processes, cell signaling and developmental strategies (Rao et al. 2014). In this study we have studied the PPI network among *N. gonorrhoea* (N_gon), *N. meningitidis* (N_men) and *Homo sapiens*. The PI protein related PPI network of considered pathogenic strains have been given in Fig. 5a and 5b. The COG based clustering of both the networks showed "cellular processing and signaling" category (red circle) contained most connected proteins. Those proteins were also connected with others associated with "information storage and processing" (yellow circles) and "Metabolism" (blue circles) categories. Few proteins (crimson circles) were proteins with uncharacterized COG category and their connectedness was less than other proteins. Overall, the PPI score was $1.0e-16$. Similar pattern of clustering was observed for NM where the pink circles were protein for "cellular processing and signaling", green circles were for "information storage and processing", yellow circles were for "Metabolism" and red circles were unknown categories. The

PPI enrichment score for *N. meningitis* was $1.0e-15$. These values for both the PPI networks indicated a stable and promising interaction among the pathogenic proteins.

Another aspect of this study was to analyze the human-*Neisseria* interaction. The human PPI network associated with Gonorrhoea and Meningitis were predicted. A huge number of proteins with tight inter-connection were found to be linked directly or indirectly with both these disorders. Twenty human proteins were found to be directly associated with Gonorrhoea having DSI (disease-significant index) more than 0.7. Their KEGG enrichment analysis revealed their functionality with oocyte meiosis, cell cycle, Epstein-Barr virus infection, dopaminergic synapse, acrosomal vesicle formation, Hippo signaling pathway, long-term depression, sphingolipid signaling pathway, p53 signaling pathway, FoxO signaling pathway and autophagy (Fig. 5c). Ten potent human proteins were found to be directly related to Meningitis with DSI value more than 7. KEGG enrichment of those proteins revealed their pivotal role in tryptophan metabolism, prion diseases, complement and coagulation cascades, Systemic lupus erythematosus (SLE), Seleno-compound metabolism, amoebiasis and axon development (Fig. 5d). The PPI analysis among NG and human revealed acrosomal vesicle formation, Hippo signaling pathway, Epstein-Barr virus infection, long-term depression and p53 signaling pathway related proteins interacted with NG PI proteins with P-value $1.0e-16$. NG causing Gonorrhoea, a sexually transmitted disorder (STD) is thus interacting with human proteins that are directly related to the development of urogenital tract, oocyte meiosis, placenta and sperm formation and development (Soncin and Parast 2020; Caini et al. 2014). The same analysis with NM and human proteins revealed a strong biological interaction (P-value $1.0e-16$) between NM PI protein and human proteins related to prion diseases, axon development, tryptophan metabolism, SLE and blood brain barrier formation. Clinical reports have been found that patients with SLE and prion diseases are more prone to Meningitis (Al Mahmeed et al. 2020; Batra et al. 2016). Thus, the PPI network analysis further established the complex machinery of Human-*Neisseria* interaction.

Conclusion

This study investigates different genomic and proteomic aspects of *Neisseria* along with their interaction with humans. The codon usage analysis revealed that this genus is neither biased towards GC rich codons nor towards AT rich codons. Similarities between *Neisseria* and human in terms of synonymous codon usage analysis indicated towards co-evolution of microbes and hosts. Moreover, CAI, tAI and Fop were found to be major indices governing the codon usage of *Neisseria*. The amino acid usage study showed the preference of energy economic amino acids among *Neisseria*. The reverse ecology analysis supported the co-occurrence of *Neisseria* and humans. The complementary effect of humans on *Neisseria* was evident from this analysis. Reverse ecology-based networking showed a strong metabolic interaction between human and considered *Neisseria* strains. Comparative analysis revealed considerable proteomic similarities between pathogenic and commensal strain. This also supported previous reports for presence of virulent genes in non-pathogenic *Neisseria* strains. The pan genomic dendrogram suggested a taxonomic reconsideration of *Neisseria* sp. KEM232. The evolutionary analysis supported the less evolved and more conserved nature of potentially highly expressed genes. Moreover, the higher evolutionary rate of both secretomes and resistomes among pathogenic *Neisseria* proposed the transition of commensal to pathogenicity in *Neisseria*. The human-pathogen interaction was studied mainly for *N. gonorrhoea* and *N. meningitis*. A strong biological interaction was established between the host and pathogen. The GO enrichment analysis and KEGG pathway study indicated most of the interacting proteins were associated with biological processes like cell signaling, cell cycle and developmental pathways. Pathogenicity related genes of *N. meningitis* was found to interact with human proteins associated with tryptophan metabolism, prion diseases, complement and coagulation cascades, Systemic lupus erythematosus (SLE), Seleno-compound metabolism, amoebiasis and axon development. Virulent genes of *N. gonorrhoea* interacted with human proteins related to development of urogenital tract, oocyte meiosis, placenta and sperm formation.

Thus, the genomic and evolutionary study on *Neisseria* revealed a considerable similarity with the genomic pattern of their host, human indicating a codon co-evolution strategy taken up by this genus.

Declarations

Conflict of Interest

Authors declare no conflict of interest.

Authors' Contribution

RPS conceived the idea. IS and PD performed all analysis. RPS, IS, PD, SSR, GDV wrote the manuscript. All authors have agreed to the final version of the manuscript.

References

1. Akashi, H., & Gojobori, T. (2002). Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*, 99(6), 3695-3700.
2. Al Mahmeed, N., El Nekidy, W. S., Langah, R., & Nusair, A. R. (2020). Meningitis as the initial manifestation of systemic lupus erythematosus. *IDCases*, 21, e00904.
3. Al Suwayyid, B. A., Rankine-Wilson, L., Speers, D. J., Wise, M. J., Coombs, G. W., & Kahler, C. M. (2020). Meningococcal disease-associated prophage-like elements are present in *Neisseria gonorrhoeae* and some commensal *Neisseria* species. *Genome biology and evolution*, 12(2), 3938-3950.
4. Albenne, C., Skov, L. K., Mirza, O., Gajhede, M., Feller, G., d'Amico, S., & Remaud-Simeon, M. (2004). Molecular basis of the amylose-like polymer formation catalyzed by *Neisseria polysaccharea* amylosucrase. *Journal of biological chemistry*, 279(1), 726-734.
5. Amorós-Moya, D., Bedhomme, S., Hermann, M., & Bravo, I. G. (2010). Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Molecular biology and evolution*, 27(9), 2141-2151.
6. Andersen, B. M., Steigerwalt, A. G., O'Connor, S. P., Hollis, D. G., Weyant, R. S., Weaver, R. E., & Brenner, D. J. (1993). *Neisseria weaveri* sp. nov., formerly CDC group M-5, a gram-negative bacterium associated with dog bite wounds. *Journal of clinical microbiology*, 31(9), 2456-2466.
7. Andersen, B. M., Weyant, R. S., Steigerwalt, A. G., Moss, C. W., Hollis, D. G., Weaver, R. E., & Brenner, D. J. (1995). Characterization of *Neisseria elongata* subsp. *glycolytica* isolates obtained from human wound specimens and blood cultures. *Journal of clinical microbiology*, 33(1), 76-78.
8. Batra, V., Khararjian, A., Wheat, J., Zhang, S. X., Crain, B., & Baras, A. (2016). From suspected Creutzfeldt-Jakob disease to confirmed histoplasma meningitis. *Case Reports*, 2016, bcr2016214937.
9. Bohlin, J. (2011). Genomic signatures in microbes—properties and applications. *The Scientific World Journal*, 11, 715-725.
10. Bohlin, J., Eldholm, V., Pettersson, J. H., Brynildsrud, O., & Snipen, L. (2017). The nucleotide composition of microbial genomes indicates differential patterns of selection on core and accessory genomes. *BMC genomics*, 18(1), 1-11.
11. Botzman, M., & Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome biology*, 12(10), 1-11.
12. Butt, A. M., Nasrullah, I., Qamar, R., & Tong, Y. (2016). Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerging microbes & infections*, 5(1), 1-14.
13. Caini, S., Gandini, S., Dudas, M., Bremer, V., Severi, E., & Gherasim, A. (2014). Sexually transmitted infections and prostate cancer risk: a systematic review and meta-analysis. *Cancer epidemiology*, 38(4), 329-338.
14. Calder, A., Menkiti, C. J., Çağdaş, A., Santos, J. L., Streich, R., Wong, A., ... & Snyder, L. A. (2020). Virulence genes and previously unexplored gene clusters in four commensal *Neisseria* spp. isolated from the human throat expand the *Neisserial* gene repertoire. *Microbial Genomics*, 6(9).

15. Carr, R., & Borenstein, E. (2012). NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. *Bioinformatics*, 28(5), 734-735.
16. Chambers, J. (2008). *Software for data analysis: programming with R*. Springer Science & Business Media.
17. Clemence, M. E. A., Maiden, M. C. J., & Harrison, O. B. (2018). Characterization of capsule genes in non-pathogenic *Neisseria* species. *Microbial genomics*, 4(9).
18. Cornejo-Granados, F., Zatarain-Barrón, Z. L., Cantu-Robles, V. A., Mendoza-Vargas, A., Molina-Romero, C., Sánchez, F., ... & Ochoa-Leyva, A. (2017). Secretome prediction of two *M. tuberculosis* clinical isolates reveals their high antigenic density and potential drug targets. *Frontiers in microbiology*, 8, 128.
19. Ehrlich, G. D., Hiller, N. L., & Hu, F. Z. (2008). What makes pathogens pathogenic. *Genome biology*, 9(6), 1-7.
20. Elias, J., Frosch, M., & Vogel, U. (2015). *Neisseria*. *Manual of clinical microbiology*, 635-651.
21. Garcia-Vallve, S., Guzmán, E., Montero, M. A., & Romeu, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic acids research*, 31(1), 187-189.
22. Garg, A., & Gupta, D. (2008). VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC bioinformatics*, 9(1), 1-12.
23. Grantham, R., Gautier, C., Gouy, M., Mercier, R., & Pave, A. (1980). Codon catalog usage and the genome hypothesis. *Nucleic acids research*, 8(1), 197-197.
24. Gris, P., Vincke, G., Delmez, J. P., & Dierckx, J. P. (1989). *Neisseria sicca* pneumonia and bronchiectasis. *European Respiratory Journal*, 2(7), 685-687.
25. Hansman, D. (1978). Meningitis caused by *Neisseria lactamica*. *New England Journal of Medicine*, 299(9).
26. Henderson, J. F., & Paterson, A. R. P. (2014). *Nucleotide metabolism: an introduction*. Academic Press.
27. Hershberg, R., & Petrov, D. A. (2010). Evidence that mutation is universally biased towards AT in bacteria. *PLoS genetics*, 6(9), e1001115.
28. Heydecke, A., Andersson, B., Holmdahl, T., & Melhus, Å. (2013). Human wound infections caused by *Neisseria animaloris* and *Neisseria zoodegmatis*, former CDC Group EF-4a and EF-4b. *Infection ecology & epidemiology*, 3(1), 20312.
29. Hildebrand, F., Meyer, A., & Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. *PLoS genetics*, 6(9), e1001107.
30. Huang, L., Ma, L., Fan, K., Li, Y., Xie, L., Xia, W., ... & Liu, G. (2014). Necrotizing pneumonia and empyema caused by *Neisseria flavescens* infection. *Journal of thoracic disease*, 6(5), 553.
31. Hurst, L. D., & Smith, N. G. (1999). Do essential genes evolve slowly?. *Current biology*, 9(14), 747-750.
32. Jordan, I. K., Rogozin, I. B., Wolf, Y. I., & Koonin, E. V. (2002). Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome research*, 12(6), 962-968.
33. Kellogg Jr, D. S., Peacock Jr, W. L., Deacon, W. E., Brown, L., & Pirkle, C. I. (1963). *Neisseria gonorrhoeae* I: virulence genetically linked to clonal variation. *Journal of bacteriology*, 85(6), 1274-1279.
34. Kim, E. M., & Seong, C. N. (2018). Complete genome sequence of *Neisseria* sp. KEM232 isolated from a human smooth surface caries. *Korean Journal of Microbiology*, 54(1), 81-83.
35. Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, 35(6), 1547.
36. Lawrence, J. G., & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *Journal of molecular evolution*, 44(4), 383-397.
37. Levy, R., & Borenstein, E. (2012). Reverse ecology: from systems to environments and back. In *Evolutionary systems biology* (pp. 329-345). Springer, New York, NY.

38. Lu, Q. F., Cao, D. M., Su, L. L., Li, S. B., Ye, G. B., Zhu, X. Y., & Wang, J. P. (2019). Genus-wide comparative genomics analysis of *Neisseria* to identify new genes associated with pathogenicity and niche adaptation of *Neisseria* pathogens. *International journal of genomics*, 2019.
39. Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., & Kyrpides, N. C. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic acids research*, 40(D1), D115-D122.
40. McArthur, A. G., Waglechner, N., Nizam, F., Yan, A., Azad, M. A., Baylay, A. J., & Wright, G. D. (2013). The comprehensive antibiotic resistance database. *Antimicrobial agents and chemotherapy*, 57(7), 3348-3357.
41. McSheffrey, G. G., & Gray-Owen, S. D. (2015). *Neisseria gonorrhoeae*. In *Molecular Medical Microbiology* (pp. 1471-1485). Academic Press.
42. Mechergui, A., Achour, W., Baaboura, R., Ouertani, H., Lakhal, A., Torjemane, L., & Hassen, A. B. (2014). Case report of bacteremia due to *Neisseria mucosa*. *Apmis*, 122(4), 359-361.
43. Mustapha, M. M., Lemos, A. P. S., Griffith, M. P., Evans, D. R., Marx, R., Coltro, E. S., & Sacchi, C. T. (2020). Two cases of newly characterized *Neisseria* species, Brazil. *Emerging infectious diseases*, 26(2), 366.
44. Nakamura, Y., Itoh, T., Matsuda, H., & Gojobori, T. (2004). Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nature genetics*, 36(7), 760-766.
45. Nekrutenko, A., Makova, K. D., & Li, W. H. (2002). The KA/KS ratio test for assessing the protein-coding potential of genomic regions: an empirical and simulation study. *Genome research*, 12(1), 198-202.
46. Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., & Furlong, L. I. (2016). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic acids research*, gkw943.
47. Rao, V. S., Srinivas, K., Sujini, G. N., & Kumar, G. N. (2014). Protein-protein interaction detection: methods and analysis. *International journal of proteomics*, 2014.
48. Reis, M. D., Savva, R., & Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic acids research*, 32(17), 5036-5044.
49. Roupheal, N. G., & Stephens, D. S. (2012). *Neisseria meningitidis*: biology, microbiology, and epidemiology. *Neisseria meningitidis*, 1-20.
50. Roy, A., Sen, A., Chakrobarty, S., & Sarkar, I. (2018). Comprehensive profiling of functional attributes, virulence potential and evolutionary dynamics in mycobacterial secretomes. *World Journal of Microbiology and Biotechnology*, 34(1), 1-19.
51. Safton, S., Cooper, G., Harrison, M., Wright, L., & Walsh, P. (1999). *Neisseria canis* infection: a case report. *Commun Dis Intell*, 23(8), 221.
52. Saha, M. S., Pal, S., Sarkar, I., Roy, A., Mohapatra, P. K. D., & Sen, A. (2019). Comparative genomics of *Mycobacterium* reveals evolutionary trends of *M. avium* complex. *Genomics*, 111(3), 426-435.
53. Schmidt, H., & Hensel, M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clinical microbiology reviews*, 17(1), 14-56.
54. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., & Ideker, T. (2011). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3), 431-432.
55. Snel, B., Lehmann, G., Bork, P., & Huynen, M. A. (2000). STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic acids research*, 28(18), 3442-3444.
56. Snyder, L. A., & Saunders, N. J. (2006). The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as 'virulence genes'. *BMC genomics*, 7(1), 1-11.

57. Soncin, F., & Parast, M. M. (2020). Role of Hippo signaling pathway in early placental development. *Proceedings of the National Academy of Sciences*, 117(34), 20354-20356.
58. Stephens, D. S., Greenwood, B., & Brandtzaeg, P. (2007). Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *The Lancet*, 369(9580), 2196-2210.
59. Taormina, G., Campos, J., Sweitzer, J., Retchless, A. C., Lunquest, K., McNamara, L. A., ... & Hanisch, B. (2021). β -Lactamase-Producing, Ciprofloxacin-Resistant *Neisseria meningitidis* Isolated From a 5-Month-Old Boy in the United States. *Journal of the Pediatric Infectious Diseases Society*, 10(3), 379-381.
60. Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2003). Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*, (1), 2-3.
61. Unemo, M., & Shafer, W. M. (2014). Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clinical microbiology reviews*, 27(3), 587-613.
62. Vesth, T., Lagesen, K., Acar, Ö., & Ussery, D. (2013). CMG-biotools, a free workbench for basic comparative microbial genomics. *PloS one*, 8(4), e6012
63. Volokhov, D. V., Amselle, M., Bodeis-Jones, S., Delmonte, P., Zhang, S., Davidson, M. K., ... & Chizhikov, V. E. (2018). *Neisseria zalophi* sp. nov., isolated from oral cavity of California sea lions (*Zalophus californianus*). *Archives of microbiology*, 200(5), 819-828.
64. Wilson, A. C., Carlson, S. S., & White, T. J. (1977). Biochemical evolution. *Annual review of biochemistry*, 46(1), 573-639.
65. Xia, X., & Xie, Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *Journal of heredity*, 92(4), 371-373.
66. Yakovchuk, P., Protozanova, E., & Frank-Kamenetskii, M. D. (2006). Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic acids research*, 34(2), 564-574.
67. Yang, F., Yan, J., & van der Veen, S. (2020). Antibiotic resistance and treatment options for multidrug-resistant gonorrhea. *Infectious Microbes & Diseases*, 67-76.
68. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences*, 13(5), 555-556.

Tables

Table 1: Overall genomic features of *Neisseria* strains considered for this analysis. The column P/OP/NP stands for Pathogens/opportunistic pathogens/nonpathogens.

taxon_oid	Genome Name / Sample Name	Short names used	Genome Size (Mb)	Gene Count	GC	RNA Count	KO Count	P/OP/NP	Signal Peptide Count
2896579436	<i>Neisseria animalis</i> NCTC 10212	N_ani	2.2	2214	51.86	82	1268	P	170
2896584165	<i>Neisseria brasiliensis</i> N.177.16	N_bra	2.6	2720	49.23	77	1499	P	229
2877225389	<i>Neisseria canis</i> NCTC10296	N_can	2.6	2516	49.73	76	1421	P	226
2765235962	<i>Neisseria chenwenguii</i> 10023	N_10023	2.5	2391	54.03	79	1359	NP	241
2814123399	<i>Neisseria cinerea</i> NCTC 10294	N_cin	1.8	1795	50.9	80	1146	NP	173
2627853762	<i>Neisseria elongata glycolytica</i> ATCC 29315	N_elo_gly	2.3	2235	54.26	79	1320	P	250
2879704437	<i>Neisseria elongata</i> M15910	N_elo	2.5	2615	53.91	79	1372	P	289
2877235250	<i>Neisseria flavescens</i> ATCC 13120	N fla	2.2	2310	49.05	82	1346	OP	230
2869919688	<i>Neisseria gonorrhoeae</i> FQ01	N_gon	2.3	2543	52.39	71	1313	P	182
649633075	<i>Neisseria lactamica</i> 020-06	N_lac	2.2	2049	52.28	77	1239	NP	169
2630968873	<i>Neisseria meningitidis</i> LNP21362	N_men	2.1	2066	51.82	80	1299	P	162
2874897852	<i>Neisseria mucosa</i> ATCC 19696	N_muc	2.7	2664	51.1	76	1470	OP	255
2912369489	<i>Neisseria muscoli</i> NW831	N_mus	2.9	3236	53.23	70	1415	NP	217
2879694896	<i>Neisseria polysaccharea</i> M18661	N_pol	2.0	2149	52.13	76	1196	NP	178
2811995355	<i>Neisseria sicca</i> FDAARGOS_260	N_sic	2.8	2634	50.95	84	1475	NP	265
2775507059	<i>Neisseria</i> sp. 10022	N_10022	2.5	2860	49.22	71	1155	Unknown*	204
2765235961	<i>Neisseria</i> sp. KEM232	N_kem	2.4	2301	58.52	56	1294	NP	260

2877215797	<i>Neisseria subflava</i> ATCC 49275	N_sub	2.2	2158	49.48	75	1317	NP	227
2773857968	<i>Neisseria weaveri</i> NCTC 13585	N_wea	2.2	2006	49	18	1289	P	198
2896586886	<i>Neisseria zalophi</i> ATCC BAA-2455	N_zal	2.4	2329	44.56	72	1369	P	237
2765235960	<i>Neisseria zoodegmatidis</i> NCTC 12230	N_zoo	2.6	2363	50.94	74	1406	P	247

* Still not confirmed as pathogen or commensal.

Figures

Figure 1

(a) Heatmap of overall codon usage among 21 *Neisseria* strains. The color code has been indicated in the figure. Both AT and GC rich codons are preferred. At third position C is preferred than other nucleotides. (b) Heatmap of overall amino acid usage among 21 *Neisseria* strains. The color code has been indicated in the figure. Energy economic amino acids were more preferred than aromatic energy costly amino acids.

Figure 2

(a) Heatmap of the competition index among 21 *Neisseria* and Human. (b) Heatmap of the complementation index among 21 *Neisseria* and Human. (c) Metabolic reconstruction among select 21 *Neisseria* strains and human based upon 'seed set'.

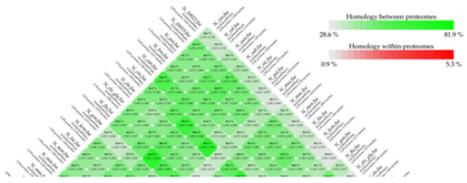


Figure 3

(a) Blast matrix analysis showing proteomic comparison among select *Neisseria* strains. (b) Pan-Core genome analysis among select *Neisseria* strains. The blue circle represents the pan-genomes and red circle represents core genome. (c) Pan genomic dendrogram using relative Manhattan distance among 21 *Neisseria* strains. (d) MLSA based phylogeny with 1000 bootstrap and NJ algorithm. *E.coli* used as out-group.

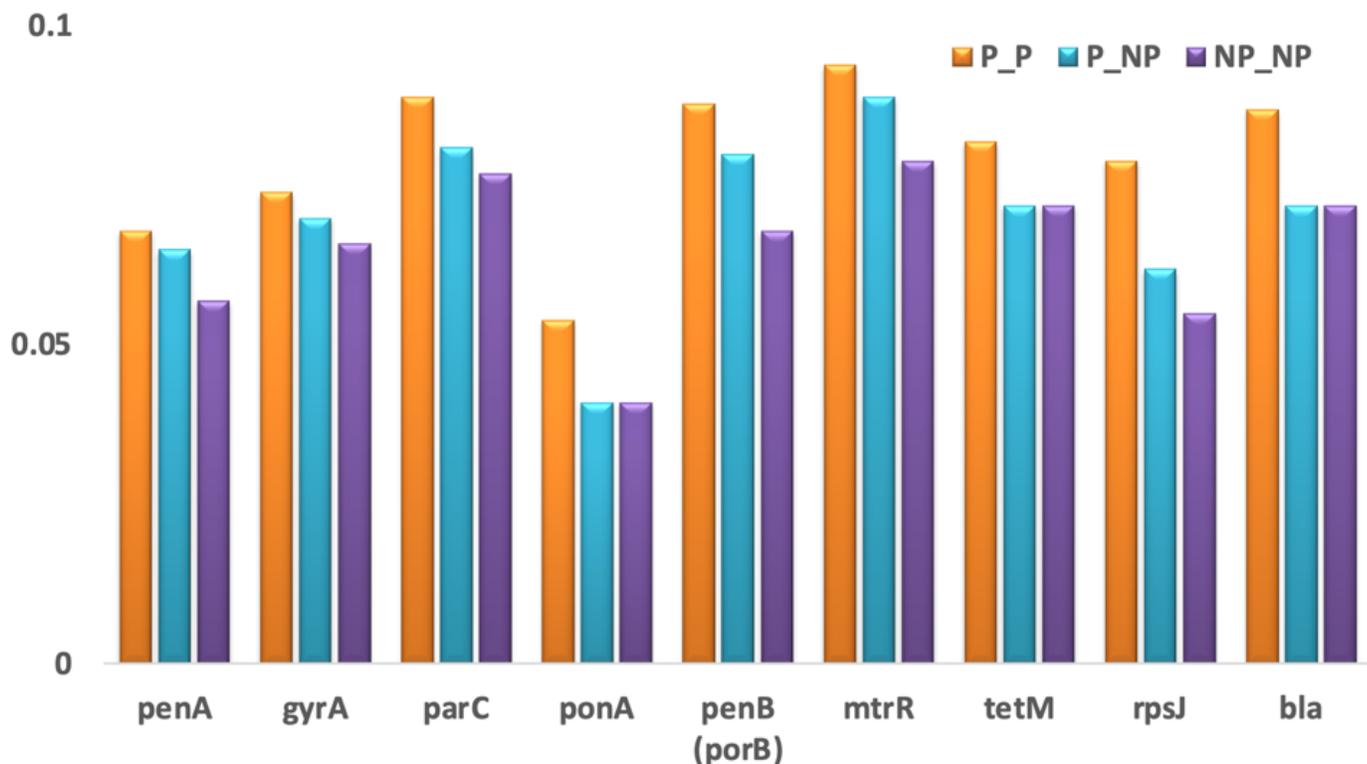


Figure 4

Evolutionary analysis of resistome among nine protein coding genes associated with AMR. The evolutionary rate of these genes is highest among pathogens and lowest in non-pathogens.

Figure 5

(a) Network analysis among *N. gonorrhoea* signal peptides. P-value > 0.8 was only considered. The clustering was done based on COG analysis. The COG based clustering of both the networks showed “cellular processing and signaling” category (red circle) contained most connected proteins. Those proteins were also connected with others associated to “information storage and processing” (yellow circles) and “Metabolism” (blue circles) categories. Few proteins (crimson circles) were proteins with uncharacterized COG category and their connectedness was less than other proteins. (b) Network analysis among *N. meningitidis* signal peptides. P-value > 0.8 was only considered. The clustering was done based on COG analysis. The pink circles were protein for “cellular processing and signaling”, green circles were for “information storage and processing”, yellow circles were for “Metabolism” and red circles were unknown categories. (c) PPI analysis among Gonorrhoea related human protein and *N. gonorrhoea* pathogenic proteins. The pink circles are human proteins and green circles are microbial virulent protein. (d) PPI analysis among Meningitis related human protein and *N. meningitidis* pathogenic proteins. The yellow circles are human proteins and green circles are microbial virulent protein.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1.xlsx](#)
- [Supplementaryfile2.xlsx](#)

- [Supplementaryfile3.xlsx](#)