

ORIGINAL RESEARCH ARTICLE

Investigating ethnic variation in the immune response to *Fusobacterium nucleatum*Sakshi Sunil Dhavale¹, Soumya Basu¹, and Tungadri Bose^{2*}¹Cancer and Translational Research Centre, Dr. D. Y. Patil Biotechnology and Bioinformatics Institute, Dr. D. Y. Patil Vidyapeeth, Pimpri, Pune (Deemed to be University), Pune, Maharashtra, India²Tata Consultancy Service Research, Tata Consultancy Services Limited, Pune, Maharashtra, India

Abstract

Fusobacterium nucleatum (FN) is an opportunistic pathogen associated with several diseases, including periodontitis, cardiovascular disease, and colorectal cancer. Owing to genetic variation, FN strains may display distinct epitope profiles to the host immune system. Likewise, the human leukocyte antigen allele repertoire is highly polymorphic and varies across ethnic groups. This study examines how these variations influence immune recognition of FN among different ethnicities and evaluates the potential for autoimmune responses arising from cross-reactivity with the FN proteome. A set of 140 experimentally verified epitopes and their variants was analyzed for binding affinities to major histocompatibility complex Class-I and Class-II molecules. Using the prevalence of these molecules across ethnicities and their affinity for FN epitopes, a new metric—pathogen detection capacity (PDC)—was formulated. PDC enabled comparison of FN recognition across populations. Additionally, the FN proteome was assessed for similarity to human epitopes implicated in autoimmune disorders. Results showed that European, South Asian, and Amerindian populations had a greater potential to detect FN, whereas Oceanian groups exhibited comparatively lower PDC values. Within super-populations, ethnicities that migrated from their ancestral homelands displayed altered PDC profiles relative to non-migrating groups. The analysis also identified four previously unreported FN peptide fragments with potential antigenic properties. Overall, this study provides a foundational framework for understanding FN immunogenicity and its variation across ethnicities. The findings highlight that disease pathophysiology can differ meaningfully among populations and that ethnicity-specific medical strategies may contribute to improved therapeutic outcomes.

***Corresponding author:**
Tungadri Bose
(tungadri.bose@tcs.com)

Citation: Dhavale SS, Basu S, Bose T. Investigating ethnic variation in the immune response to *Fusobacterium nucleatum*. *Microbes & Immunity*. 2026;3(3):025490129.
doi: 10.36922/MI025490129

Received: December 6, 2025

Revised: January 16, 2026

Accepted: February 5, 2026

Published online: April 15, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: *Fusobacterium nucleatum*; Human leukocyte antigen system; Major histocompatibility complex alleles; Epitopes and immune response; Ethnic variations

1. Introduction

The human leukocyte antigen (HLA) system, also referred to as the major histocompatibility complex (MHC) gene repertoire, is integral to immune recognition and modulates the host immune response to infections.^{1,2} It encodes cell-surface molecules that present antigenic peptides to T-cells.¹ HLA molecules are highly polymorphic, with various alleles influencing the adaptive immune system's ability to differentiate

between self and foreign antigens.² While MHC Class-I molecules, encoded by genes at loci *HLA-A*, *HLA-B*, and *HLA-C*, present peptides from intracellular pathogens to CD8-positive cytotoxic T-cells, MHC Class-II molecules, encoded by genes in the *HLA-DP*, *HLA-DQ*, or *HLA-DR* region, present peptides from extracellular pathogens to CD4-positive helper T-cells.¹ The polymorphism in *HLA* alleles contributes to genetic variability across populations and influences susceptibility to pathogens.^{2,3} Variations in *HLA* alleles can alter the range of peptides they bind, affecting immune responses to specific pathogens and the infections they cause.^{3,4}

Fusobacterium nucleatum (FN), a Gram-negative bacterium historically regarded as an oral commensal, has emerged as an opportunistic pathogen linked to multiple human diseases, including periodontal disorders, cardiovascular disease, colorectal cancer (CRC), and adverse pregnancy outcomes.⁴⁻¹² FN contributes to periodontal disease through biofilm formation and inflammation,⁵ while its dysregulation in the microbiota can lead to systemic inflammation and increased risk of atherosclerosis and other diseases.^{8,11} In CRC, FN fosters a tumor-promoting environment by inducing local inflammation and inhibiting anti-tumor immune responses through Fap2-mediated interactions with receptors such as the T-cell immunoreceptor with IgG.^{4,5,11,13,14} These dynamics raise critical questions about the bacterium's ability to evade immune detection, an aspect that has profound implications for disease manifestation.

Given the known polymorphisms in *HLA* alleles, it is likely that individuals from different ethnicities would also vary in their ability to detect FN through their immune repertoires (antigenic potential). Additionally, certain FN strains may possess genetic adaptations¹⁵ that allow them to evade host immunity, leading to increased risk of infection or disease. Therefore, our study aims to explore the hypothesis that the antigenic profiles of diverse FN strains differ across ethnicities and whether FN-derived epitopes may resemble human epitopes, potentially triggering autoimmune responses.

Results from our analysis indicate that the antigenic profiles of different FN strains indeed vary, with potential implications in immunogenic responses across ethnic groups. Notably, populations from Europe, South Asia, and the Americas showed higher pathogen detection capacity (PDC) against FN compared to Oceanian ethnicities, who demonstrated lower PDC. Furthermore, migrant South Asian (SAS) and African populations (in the USA) exhibited distinct PDC profiles, suggesting that ancestral homogeneity influences immune recognition. In our investigation, we also identified four novel peptide

fragments from the FN proteome that are homologous to human epitopes, underscoring the potential for autoimmune reactions.

2. Materials and methods

2.1. Data acquisition

The protein sequences of the 12 fully sequenced FN strains were retrieved from the National Center for Biotechnology Information Genome Database on September 25, 2023 (Table 1). Furthermore, the information on the experimentally verified epitopes in FN was obtained from the Immune Epitope Database (IEDB).¹⁶ This dataset consisted of 140 unique epitope sequences with lengths varying between 9 and 18 amino acids. Human epitopes possibly associated with autoimmune disorders were also obtained from IEDB. Information on the distribution of *HLA* alleles among 82 ethnic populations was obtained from earlier publications.³ This comprised 342 *HLA* allelic gene products, with 258 and 84 alleles corresponding to MHC molecules Class-I and Class-II, respectively, which have been observed to occur at frequencies ≥ 0.01 across various ethnicities.

2.2. Mapping the epitopes (and their variants) to the *Fusobacterium nucleatum* strains

The 140 epitope sequences obtained from IEDB were mapped on the protein sequences of the studied FN strains using the Basic Local Alignment Search Tool (BLAST) program (Version 2.13, NCBI, USA).¹⁷ In this step, epitope sequences were queried for exact matches (100% identity and 100% overlap) against a database comprising the protein sequences of all 12 studied FN strains. Results were parsed using in-house Python scripts (Version 3.10, Python Software Foundation, USA). It may be noted that none of the 140 reported epitopes mapped to the combined proteome of the 12 FN strains. Thus, to accommodate mutations within these 140 epitopes, further analysis of the BLAST output was performed. In this step, the FN protein sequence segments matching the epitope sequences with at most one mismatch—insertion, deletion, or substitution—were also noted. Since the antigenic potential of these sequences was unknown, they were considered potential epitopes.

2.3. Predicting the binding affinity of epitopes

While a list of experimentally verified epitope sequences in FN was available from IEDB, information on the affinity of these epitopes for the MHC molecules was unavailable. Thus, the MHC I Binding and MHC II Binding prediction tools available on IEDB were used for this purpose.¹⁶ The NetMHCpan (Version 4.1, DTU Health Tech, Denmark)

Table 1. List of the 12 *Fusobacterium nucleatum* (FN) strains analyzed in this study

Strain/biosample ID	Isolation		Mapped epitopes			% of variant epitopes
	Geography	Site	Total	Original	Variant	
FNP/SAMN33169818	China: Shandong	Oral/dental	130	75	55	42.31
FNA/SAMN33170517	China: Shandong	Oral/dental	129	77	52	40.31
FNU/SAMN33170998	China: Shandong	Oral/dental	124	73	51	41.13
FNV/SAMN33170584	China: Shandong	Oral/dental	130	69	61	46.92
MGYG-HGUT-01347/ SAMEA5850850	UK	Gut	128	70	58	45.31
ChDC F317/SAMN03263083	South Korea: Gwangju	Oral/dental	134	103	31	23.13
KCOM 1322/SAMN03263583	South Korea: Gwangju	Oral/dental	134	103	31	23.13
2_1_50A/SAMN33244396	USA	Pure cell culture	126	73	53	42.06
25586/SAMN08706662	USA: Virginia	Oral/dental	133	107	26	19.55
JD-Fn1/SAMN32721952	China: Shanghai	Oral/dental	126	75	51	40.48
ATCC 25586/SAMN02603417	USA: Illinois	Oral/dental	134	108	26	19.40
ATCC 23726/SAMN08501025	USA: Virginia	Urogenital tract	133	102	31	23.31

Note: In each strain, the number of epitopes that are mapped to the list of FN epitopes available in the Immune Epitope Database with (variant) and without (original) mismatches is indicated.

binding affinity method (as recommended by IEDB) was employed for binding prediction, with the MHC source species set to human. The list of 342 *HLA* alleles (258 MHC Class-I and 84 MHC Class-II), mentioned in the data acquisition step, was used to assess epitope-binding affinity. For the MHC I Binding prediction tool, amino acid segments of length 9 to 14 were considered. The corresponding length range for the MHC II Binding prediction tool was 11–18 amino acid stretches. In addition to the 140 epitopes downloaded from IEDB, the exercise was also performed for the potential epitopes obtained in an earlier step. Epitope binding to MHC molecules with

half maximal inhibitory concentration (IC_{50}) values < 50 nM was considered as a strong binding, wherein the lower IC_{50} values were indicative of the stronger binding affinity of an MHC molecule for an epitope.

2.4. Determining the propensity of each ethnicity to detect the *Fusobacterium nucleatum* strains

The epitopes present in each of the 12 FN strains varied. Therefore, the propensity of an ethnic group to detect each FN through their repertoire of the *HLA* alleles prevalent in the population would also vary. Furthermore, given the polymorphic nature of *HLA* genes, this potential to detect

the pathogen would vary across ethnic populations. To capture this, we devised a metric, PDC. It was determined as in **Equation 1**:

$$PDC_{strain}^{ethnicity} = \sum_{e=1}^n \sum_{hla=1}^m \left(\frac{1}{IC50_e^{hla}} \times frequency_{hla}^{ethnicity} \right) \quad (1)$$

where $PDC_{strain}^{ethnicity}$ is the PDC of an ethnicity to detect a given FN strain; e (ranging from 1 to n) represents the number of epitopes encoded by the FN strain; $IC50_e^{hla}$ represents the $IC50$ value for an epitope e for the corresponding *HLA* gene product; hla is the total number of *HLA* gene alleles considered ($m = 342$ in this case); and $frequency_{hla}^{ethnicity}$ is the frequency of occurrence of a *HLA* gene allele in the ethnicity.

2.5. Aligning human epitopes to *Fusobacterium nucleatum* proteome

Similarities between the experimentally verified human and the FN epitopes were analyzed using BLAST. The results were further parsed using in-house Python scripts to identify human epitopes with at least 75% sequence identity and 75% sequence overlap with any FN epitope. In additional analysis, human epitopes potentially associated with autoimmune disorders were queried for exact matches (100% identity over the entire length of the epitope) against the entire repertoire of protein sequences of the 12 studied FN strains. Only linear human epitopes were considered for this analysis.

2.6. Statistical analysis and data visualization

Statistical analyses, including pairwise *t*-tests to assess differences in epitope recognition potential among super-populations comprising all the ethnicities, were performed using matplotlib.pyplot and pandas packages in Python and visualized using Excel (Microsoft Corporation, USA). Benjamini–Hochberg corrected *p*-values (*q*-value) of 0.05 were used to report the significant differences. The clustering of the FN strains was generated using Web-Igloo (TCS Research, India).¹⁸

3. Results

3.1. *Fusobacterium nucleatum* isolated from China encoded a higher proportion of variant epitopes

The 140 experimentally verified epitopes (Table S1A), as obtained from IEDB¹⁶ and their variants, were mapped against the protein sequences of FN strains using BLAST.¹⁷ Twelve fully sequenced FN strains were used for the analysis. The site of isolation and the distribution of the epitopes among these 12 FN strains are provided in **Table 1**. While the number of epitopes among the studied strains varied, the strains isolated from China encoded a higher

proportion of variant epitope sequences with respect to those reported in IEDB (Table S1B). One strain each from the UK and the USA also followed this trend. Notably, the strains isolated from South Korea and strain 25586 (isolated from the USA) did not harbor any of the variant epitope sequences. No major variations between the strains based on the site of infection/isolation were found. While 111 epitopes were mapped to at least one of the FN strains, 40 epitopes mapped to all the studied strains (Table S1C). After allowing for mismatch, variants of 77 out of these 111 epitopes were further mapped to one or more FN strains. Of these, 11 variants mapped to all the FN strains (Table S1B). A total of 45 epitopes (including variants) could be mapped to only one strain. These were predominantly found in the FN strains isolated from China and a strain isolated from the USA (2_1_50A). Among the 29 epitopes that could not be mapped to any of the FN strains, five epitopes were found to be overlapping with FN protein sequences after allowing for mismatches (i.e., the epitope variants). The overall clustering of the studied FN strains based on the presence of epitopes is shown in **Figure A1**.

3.2. Recognition potential of *Fusobacterium nucleatum* epitopes by human HLAs varied widely

The MHC I and MHC II binding prediction tools hosted by IEDB¹⁶ were used to predict the binding affinities between human MHC molecules and the experimentally verified FN epitopes and their variants (Section 2). More than half of the original 140 FN epitopes ($n = 81$) were predicted to have strong binding affinity (IC_{50} values < 50 nM) for at least one of the 342 human HLAs (Table S2A). This set of 342 human HLAs is known to be abundant among the studied ethnic groups. Among the 258 MHC Class-I associated *HLA* alleles, each of *HLA-A*02:22*, *HLA-A*02:11*, *HLA-A*02:50*, *HLA-A*02:02*, and *HLA-A*02:03* was found to recognize more than 10 FN epitopes (**Figure 1A**). Among the 84 MHC Class-II associated *HLA* alleles, *HLA-DRB1*01:01*, *HLA-DRB1*04:01*, *HLA-DRB1*10:01*, and *HLA-DRB1*13:02* presented the highest number of epitopes (**Figure 1B**). Furthermore, several epitopes were recognized by multiple *HLA* gene products. The most widely recognized epitopes were Epitope 136 (YPITPSSPM), Epitope 79 (LDNKIILIPSIGIDK), Epitope 28 (FTTD TAAAL), and Epitope 72 (KNTLIIMTSNIGSHF), which were recognized by 39, 26, 19, and 15 alleles, respectively (**Figure 1C**).

Among the variants, 53 epitopes exhibited strong binding affinity to at least one human *HLA* molecule (Table S2B). These were variants of 35 experimentally verified epitopes obtained from IEDB (Section 2). Of these, 16 epitope variants were predicted to have better binding affinity to a human MHC molecule compared with

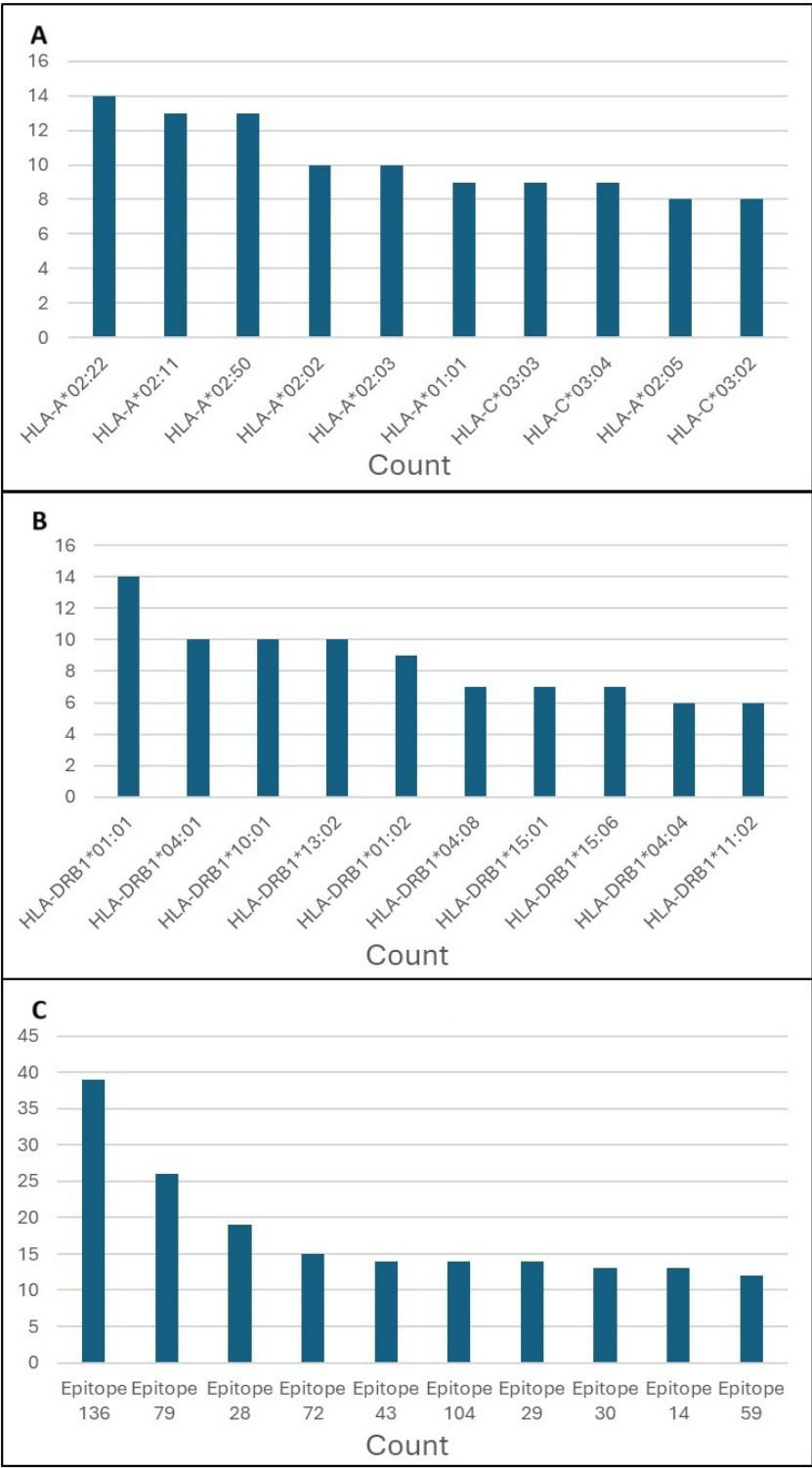


Figure 1. The top 10 human (A) major histocompatibility complex (MHC) Class-I and (B) MHC Class-II alleles capable of recognizing the highest number of *Fusobacterium nucleatum* (FN) epitopes. (C) The top 10 FN epitopes that are recognized by the highest number of human MHC alleles.

the original epitope, i.e., the epitope variant reported in IEDB (Table S2C). While four variants of Epitope 63 were predicted to have better binding affinities (IC_{50} value) than the variant available in the IEDB database, the greatest improvement in binding affinity with respect to the original epitope was seen in the case of Epitope 27. Notably, 159 MHC Class-I and 44 MHC Class-II *HLA* alleles were predicted to lack binding affinity for any of the FN epitopes reported in IEDB or their variants (Table S2A).

It was therefore evident that the recognition potential (binding affinity) of the human *HLA* allelic repertoire for the individual FN epitopes varied widely. Thus, the distribution of FN epitopes, which were recognized by varying numbers of *HLA* alleles, was studied across strains. The epitopes were categorized as (i) most frequently recognized epitopes, those recognized by five or more *HLA* alleles, (ii) less frequently recognized epitopes, those recognized by less than five *HLA* alleles, and (iii) unrecognized epitopes, those not recognized by any of the *HLA* alleles used in this study. The distribution of the number of epitopes most and least frequently identified as well as unidentified by the human *HLA* repertoire was found to be similar for most of the studied strains (Table S2D), suggesting that the probability of differential recognition of an FN strain by the human *HLA* system is expected to be low. However, this potential may also be affected by the variability in the distribution of *HLA* genes among ethnicities.³

3.3. Ethnic groups differed in their potential to detect pathogens

It is evident from the results presented in Sections 3.1 and 3.2 that the binding affinity/recognition potential of the *HLA* alleles for each of the FN epitopes varies widely. Furthermore, earlier studies have shown that the distribution of *HLA* alleles across ethnic groups differs substantially.³ Given this, we hypothesized that the combined potential of the repertoire of *HLA* genes encoded by an ethnicity to determine the presence of FN would also vary. We designed a metric termed PDC to test the above hypothesis. Using this metric, we compared the FN epitope detection potential of 82 ethnicities constituting seven super-populations (Table S3A) for each of the 12 FN strains studied. Tables S3B–D list the computed PDC values of ethnicities for each of the 12 studied FN strains. The highest PDC values were observed in the Caucasoid European (EUR) and Amerindian (AMR) populations, with the Finnish population in Finland (EUR_FIN) and Netherlands UMCU (EUR_NEU) showing the highest values (Table S3B). In contrast, the Oceanian (OCN) ethnicities of Australia, Cape York Peninsula Aborigine (OCN_AUS), and Papua New Guinea, Madang (OCN_

PNG) reported the lowest PDC values (Table S3B). While the PDC via the MHC Class-II repertoire was highest among EUR ethnicities, a few AMR ethnicities from Mexico, such as Oaxaca Mixe (MXX) and Chiapas Lacandon Mayans (MXM), showed the highest PDC via the MHC Class-I repertoire (Table S3C, D). Interestingly, many of these Mexican ethnicities had extremely low PDC through MHC Class-II repertoire (Table S3D). Overall, EUR, SAS, and AMR demonstrated significantly higher PDC (q -value < 0.05) compared to most other super-populations (Figure 2 and Table S3E). In general, OCN demonstrated the lowest recognition potential. Significant differences in strain-specific detection capacities were also noted for a few ethnicities. The detection capacities were lower in three strains, namely, FNU, JD-Fn, and 2_1_50A (Table S3B). In this context, the relatively higher ability of MXX to detect JD-Fn compared to the other FN strains was worth noting (Table S3D).

Within the super-populations originating from South Asia and Africa, ethnicities migrating to Western countries demonstrated a lower PDC via the MHC Class-II repertoire when compared to those who did not migrate and reside in their native environment (Figure 3A–F). Furthermore, in contrast to native AMRs, migrants of EUR (Caucasoid), SAS, and African origins had lower PDC values for MHC Class-I alleles but higher PDC values due to MHC Class-II alleles (Figure 3G and 3H). Among them, the EUR ethnicities residing in the USA demonstrated the highest PDC score via the MHC Class-II repertoire. Implications of this observation have been discussed in Section 4.

3.4. *Fusobacterium nucleatum* epitopes might induce autoimmunity by mimicking human epitopes

The list of human epitopes associated with autoimmune disorders was obtained from IEDB.¹⁶ The FN epitopes and their variants were compared to the human epitopes to detect FN epitopes that might induce autoimmunity in the host. None of the FN epitopes studied matched the human epitopes with 100% identity over the entire peptide length. The best match was noted for a variant of Epitope 41 (Epitope 41.2), which aligned with a human epitope (IEDB id: 562550) with an e -value of 0.08 (identity 100%; coverage 80%). This variant of Epitope 41 was omnipresent in all studied FN strains. We further investigated the entire FN proteome to identify FN peptide fragments that might mimic a human epitope. Based on sequence similarity search, four peptides (potential epitopes) were found to be identical to human epitopes (Table 2). Three of these peptides were encoded by the genomes of all the FN strains studied. Notably, a human epitope identical to two of these potential FN epitopes (viz., LLDVTPLSL and ILDSRGNPTV) demonstrated the highest binding affinity

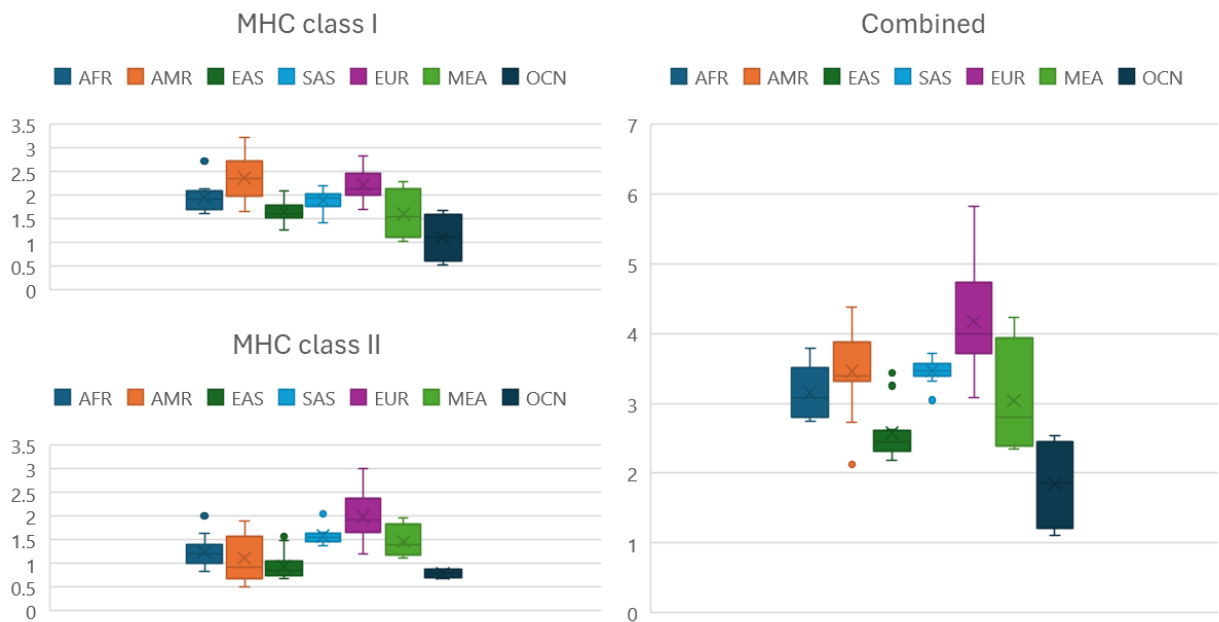


Figure 2. Comparison of *Fusobacterium nucleatum* epitope detection potential using the pathogen detection capacity (PDC) metric for 82 ethnicities constituting seven super-populations: Africans (AFRs), Amerindians (AMRs), East Asians (EASs), Europeans (EURs), Middle East and Africans (MEAs), Oceanians (OCNs), and South Asians (SASs)
Abbreviation: MHC: Major histocompatibility complex.

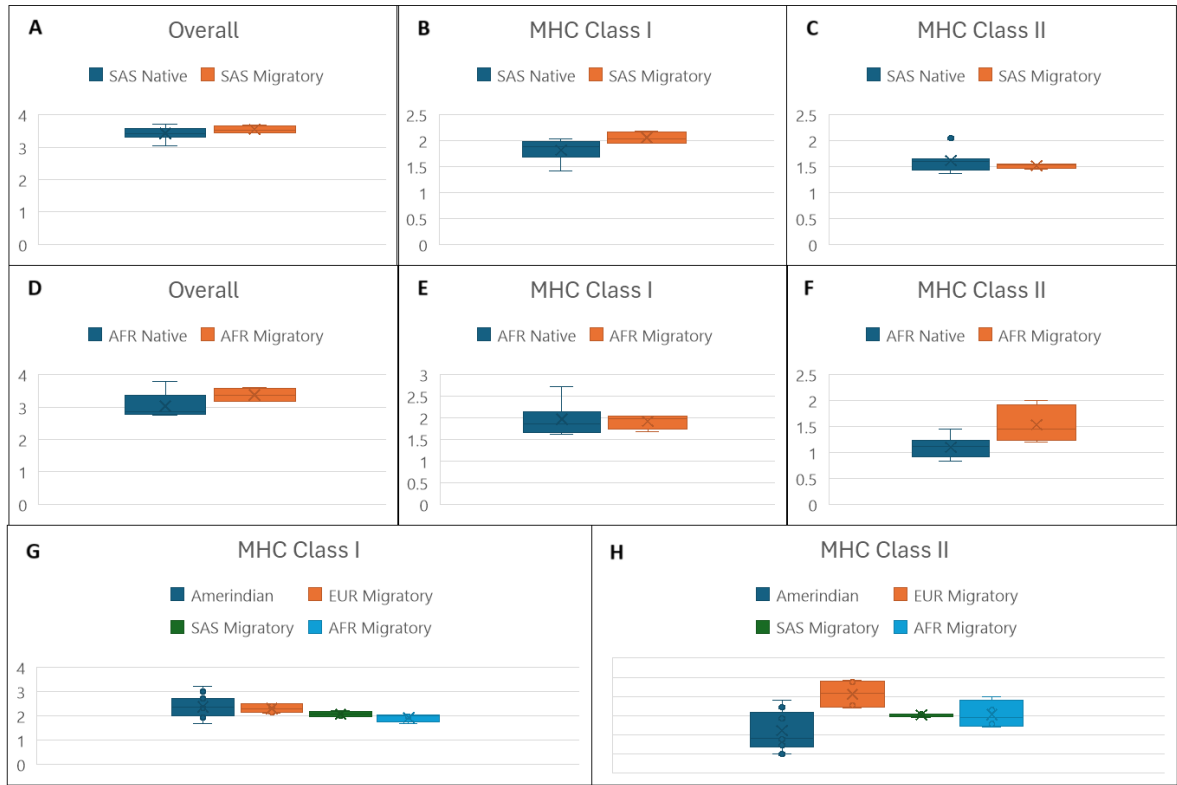


Figure 3. Comparison of pathogen detection capacity (PDC) values for different ethnic groups. (A–C) Native vs. migrant ethnicities from South Asia (SAS). (D–F) Native vs. migrant ethnicities from Africa (AFR). (G–H) Native Amerindians, European (EUR) migrants to the American continent, and migrants from SAS and AFR.

with *HLA-A*02:11*, which is missing among individuals of Middle Eastern and African (MEA) ethnicity. The fourth potential epitope (viz., ILYGPPGTGK) was encoded by the genome of only a few of the strains isolated in China. Table S4 lists recently published studies that have examined the possible role of human epitopes (Table 2) in immune-associated disorders. This aspect has been further explored in the next section.

4. Discussion

F. nucleatum is an opportunistic colonizer, capable of co-aggregating with various microorganisms in low-oxygen environments.⁶ FN contributes to diseases such as periodontitis, oral mucositis, peri-implantitis, dental cavities, atherosclerosis, coronary artery disease, and CRC.^{4,6,8,11,12,15} Several studies have reported that variations

Table 2. The list of *Fusobacterium nucleatum* peptides that mapped to the human epitopes possibly associated with autoimmune responses

Peptide sequence/ human epitope	Human epitope ID	Bacterial protein	Human protein	HLA-distribution (IC ₅₀ value)	FN-epitope
LLDVTPLSL	Epitope-128046	Molecular chaperone DnaK	Heat shock cognate 71 kDa protein (UniProt: P11142); stress-70 protein, mitochondrial (UniProt: P38646)	<i>HLA-A*02:11</i> (15.35), <i>HLA-C*05:01</i> (29.58), <i>HLA-C*08:02</i> (42.87)	Present in all studied strains
ILDSRGNPTV	Epitope-193945	Phosphopyruvate hydratase	Gamma-enolase (UniProt: P09104); beta-enolase (UniProt: P13929)	<i>HLA-A*02:11</i> (17.67)	Present in all studied strains
ILYGPPGTGK	Epitope-212333	AAA family ATPase; McrB family protein	26S proteasome regulatory subunit 4 (UniProt: P62191)	<i>HLA-A*03:01</i> (9.38), <i>HLA-A*03:02</i> (40.91)	Present in three strains (FNA, FNU, JD-Fn1)
KEIFLRELI	Epitope-212997	Hsp90 protein; molecular chaperone HtpG	Heat shock protein HSP 90-alpha (UniProt: P07900); heat shock protein HSP 90-alpha A2 (UniProt: Q14568); heat shock protein HSP 90-beta (UniProt: P08238); endoplasmic (UniProt: P14625)	<i>HLA-B*40:04</i> (15.84), <i>HLA-B*40:02</i> (18.4), <i>HLA-B*41:02</i> (21.47), <i>HLA-B*41:03</i> (22.37), <i>HLA-B*40:09</i> (35.73), <i>HLA-B*40:01</i> (39.64)	Present in all studied strains

Abbreviations: FN: *Fusobacterium nucleatum*; HLA: Human leukocyte antigen.

in immune responses to FN invasion may be linked to CRC incidence and disease outcome.^{12,13,19} Given this, it is important to understand the host factors that are involved in the recognition and/or neutralization of the pathogen. In humans, immune recognition and subsequent host responses to infections are orchestrated via the HLA system. Being highly polymorphic, variations in *HLA* allelic makeup can modulate the epitopes (peptides) that can bind to (and get recognized by) the *HLA* gene products, i.e., the MHC Class-I and MHC Class-II molecules. This often manifests as alternative responses to the same pathogen among individuals of varying ethnicity who may harbor different genetic makeup.³

Furthermore, strain-specific variations in the bacterial genome may also lead to differential immune responses and disease outcomes. For example, recent reports indicate that a distinct FN clade (namely, FN subspecies *animalis* C2) dominates the CRC niche.²⁰ While FN proteins like Fad2 and Fap2 have been identified as potential factors related to FN's pathogenic processes,^{11,13,19} the knowledge in this area remains incomplete. FN's association with various diseases highlights its pathogenic potential and underscores the importance of understanding its mechanisms of infection and interaction with the host's immune system. This understanding is crucial for developing targeted therapeutic interventions. Therefore, we aim to explore how variations within the FN genome influence its recognition by the human HLA repertoire. Furthermore, the impact of variation in the human *HLA* gene repertoire across ethnicities has been examined in this context. By using state-of-the-art bioinformatics tools and computational analyses, we sought to understand the interplay among FN recognition potential, immune responses, and genetic variation across diverse populations.

In this study, experimentally verified 140 FN epitopes from the IEDB and their variants were analyzed. In general, strains isolated from China were found to harbor a higher proportion of variant epitopes (Table 1). However, the number of epitopes in each of the studied strains was comparable (between 124 and 134), irrespective of the strain's geographical origin or the site of isolation/infection. Among the epitopes reported in IEDB, 81 were predicted to bind strongly (IC_{50} value < 50 nM) with at least one of the human *HLA* gene products (Table S2A). For this purpose, a list of 342 human HLAs previously reported as most prevalent across diverse ethnic groups was used.³ Among the 258 MHC Class-I associated *HLA* alleles, *HLA-A*02:22*, *HLA-A*02:11*, *HLA-A*02:50*, *HLA-A*02:02*, and *HLA-A*02:03* were found to be the most potent in the antigenic recognition by cytotoxic T-lymphocyte cells (i.e., CD8+ T-cells) of FN epitopes

(Figure 1A). Among the 84 MHC Class-II associated *HLA* alleles, which pertained to CD4+ T-cells immunity, *HLA-DRB1*01:01*, *HLA-DRB1*04:01*, *HLA-DRB1*10:01*, and *HLA-DRB1*13:02* presented the highest number of epitopes (Figure 1B). Epitope 136 (YPITPSSPM), Epitope 79 (LDNKIILPSIGIDK), Epitope 28 (FTTDTAAAL), and Epitope 72 (KNTLIIMTSNIGSHF) were noted to be the most widely recognized FN epitopes (Figure 1C). Furthermore, 16 variant epitopes (variants of the epitopes reported in IEDB) were predicted to have better binding affinities/immune recognition potentials for human *HLA* molecules as compared to the experimentally identified/originally reported epitope (Table S2C). The remaining 37 variant epitopes were predicted to elicit a muted immune response when compared to the originally reported epitope. This indicates the pathogen's ability to bypass the host immune response. In this context, we noted that a large proportion of the 342 human *HLA* alleles (159 MHC Class-I and 44 MHC Class-II) analyzed in this study were not involved in the recognition of any of the experimentally validated FN epitope or their variants. While differences in the antigenic profiles of the 12 studied strains could be observed, it was not evident if the differences in the profiles would translate to an altered immunogenic response, as the total number of epitopes present in each of the strains and those recognized by the human *HLA* system were comparable (Table 1 and Table S2D).

We further investigated if there were any ethnicity-specific variations in the recognition of FN epitopes across seven super-populations: Africans, AMR, EUR, East Asians, SAS, MEA, and OCN. We designed a metric called PDC for this purpose. The analysis focused on both MHC Class-I and MHC Class-II recognition potentials and revealed certain interesting patterns. EUR, SAS, and AMR ethnicities were found to have higher PDC values when compared to most other super-populations (Figure 2 and Table S3). Furthermore, within the super-population, PDC values differed between ethnicities settled in the native habitat and those who migrated to other parts of the world (Figure 3). Intriguingly, a few Mexican ethnicities had high PDC through MHC Class-I repertoire, but their capability to recognize FN via the MHC Class-II repertoire was low (Table S3C, D). Notably, FN is known to induce local inflammation, increase cytokine production, and promote tumor growth in CRC.^{4,5} Given that the immune recognition via the *HLA* Class-II molecules is primarily associated with this immunostimulation,^{3,21} it may be speculated that the Mexican ethnicities, which are characterized by lower PDC scores for the MHC Class-II molecules, would face lesser activation of pro-inflammatory cytokine storms within the CRC tumors. Several indications to support the notion could be obtained from existing scientific literature,

which states that Mexicans following a traditional lifestyle are less prone to cancer and/or systemic inflammation.^{22,23} Moreover, contrary to most reports available globally, in a study involving CRC patients in Mexico, Cuellar-Gómez *et al.*²⁴ observed that interleukin-23 was significantly reduced in the tumor tissue infected with FN when compared with the normal samples. However, more comprehensive data would be required to draw any definite conclusions.

In contrast to Mexican ethnicities, the Caucasoid EUR ethnicities had higher PDC values via the MHC Class-II alleles (Table S3). We noted that EUR Caucasoids who migrated to the USA had higher PDC values compared to native AMRs and migratory populations from South Asia and Africa (**Figure 3**). Furthermore, these EUR Caucasoid ethnicities also had a higher PDC value for MHC Class-I when compared to the migratory populations from South Asia and Africa. It may therefore be hypothesized that owing to the presence of a greater repertoire of both MHC Class-I and MHC Class-II molecules, the Caucasoids would be better equipped to recognize and neutralize FN. This may have further ramifications, as a deficiency in the recruitment of immune cells and the ability to mount appropriate systemic secretion of cytokines could lead to poor survival rates, as has been observed in African Americans suffering from CRC.²⁵ In this context, observations made by Lo *et al.*²⁶ is noteworthy. Although they reported higher seroprevalence of FN antibodies among non-White participants than among White participants in the USA, they concluded that this observation could be due to antibody responses resulting from cross-reactivity with other closely related bacterial species.

Several recent reports have also indicated that FN may be associated with the manifestation of autoimmune disorders like rheumatoid arthritis.^{7,27} The FadA protein, which is present in the outer membrane vesicles of FN, has been implicated in this case.²⁷ Given the association of FN with a wide array of systemic diseases, we investigated whether any of the epitopes analyzed in this study mimic the human epitopes that are known to be associated with autoimmune diseases. Among the FN epitopes (either reported in IEDB or its variants identified in this study), only Epitope 41.2 demonstrated similarity to the epitopes associated with autoimmunity in humans. It had a 100% identity with the human epitope (IEDB id:562550) over 80% of its length. This epitope is part of the mitochondrial translation factor *GUF1* gene product and is known to be associated with a form of epileptic encephalopathy (Q8N442). Fascinatingly, a recent study in China reported a higher abundance of *Fusobacterium* in the gut as a biomarker for the diagnosis of epilepsy in adults.²⁸ Thus, further studies deliberating

on this aspect might be helpful towards understanding the crosstalk between members of the *Fusobacterium* genus and neuronal health. Apart from the epitopes reported in IEDB or their variants, four stretches of FN peptides were identified that matched the human epitopes (**Table 2**). These human epitopes may have possible roles in autoimmune responses and have previously been studied in the context of various diseases and metabolic disorders (Table S4). While numerous studies reported the role of these epitopes in the manifestation of viral and bacterial pathogenic mechanisms,^{29,30} others found their involvement in various forms of cancer and autoimmune diseases like IBD.^{31,32} Thus, the list of FN epitopes available in IEDB may not be exhaustive, and new epitope candidates are likely to emerge as new high-throughput experiments are being designed.

Overall, this study introduces a systematic approach to understanding the immunogenic properties of FN and their variation across ethnic groups. In addition, the work presents a view of variation in the epitope landscape across different FN strains. While such studies have been conducted to understand viral pathogens,³ our study is possibly the first attempt to delineate a bacterial pathogen. Given that the FN genome is over 2 Mb in length, an entirely *ab initio* approach to detect the bacterial epitopes could not be adopted due to computational limitations. Hence, only experimentally verified FN epitopes and their variants across strains have been analyzed in this study. Further, the current study has examined only linear epitopes associated with T-cell immunity and has not assessed non-linear/structural B-cell epitopes due to similar logistical challenges. The host's microbiota is yet another factor that could impact the immunogenic properties of FN^{11,12,33} but was outside the scope of the current study. Despite these limitations, our study provides a foundational framework for exploring and understanding FN's interactions with host immune mechanisms and how they might vary across large populations. The perspectives presented herein are promising and invite further validation of how the pathophysiology of diseases often varies across ethnicities and, therefore, might warrant ethnicity-specific medical interventions using advanced *in vitro* and *in vivo* methods. Future studies should aim to expand the repertoire of FN epitopes and incorporate broader experimental data to enhance the understanding of FN's role in disease pathogenesis and its potential for developing personalized/ethnicity-specific medical interventions.

5. Conclusion

In summary, this study offers a detailed computational view of the immune-related features of FN and illustrates how both bacterial genomic variation and human HLA diversity influence FN recognition across populations. By

examining experimentally validated FN epitopes and their strain-specific variants, we show that although the overall epitope landscape is largely conserved, even small sequence changes can alter HLA binding strength and potentially shift the host immune response. The PDC metric discussed here allowed us to compare these recognition patterns across ethnic groups, revealing population-level differences that may help explain variations in disease susceptibility, clinical progression, and therapeutic responses. Our observation that some FN epitopes resemble human autoimmune-associated peptides also raises the possibility that molecular mimicry contributes to disease mechanisms, an aspect that will require a deeper investigation. While our analyses are limited by the current availability of characterized FN epitopes and do not capture structural B-cell determinants, the framework presented here provides an important starting point for understanding FN–host immune interactions on a global scale. Future studies incorporating broader epitope datasets, microbiome-level interactions, and experimental validation will be essential for refining these insights and for guiding the development of more targeted, potentially ethnicity-specific approaches to diagnosing and managing FN-associated diseases.

Acknowledgments

None.

Funding

None.

Conflict of interest

At the time of submission, Tungadri Bose was employed by Tata Consultancy Services Limited. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions

Conceptualization: Soumya Basu, Tungadri Bose

Formal analysis: Sakshi Sunil Dhavale

Investigation: Tungadri Bose, Sakshi Sunil Dhavale

Methodology: Soumya Basu, Tungadri Bose

Writing–original draft: Sakshi Sunil Dhavale, Tungadri Bose

Writing–review & editing: Sakshi Sunil Dhavale, Tungadri Bose

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

No datasets were generated or analyzed during the current study.

References

1. Crux NB, Elahi S. Human Leukocyte Antigen (HLA) and Immune Regulation: How Do Classical and Non-Classical HLA Alleles Modulate Immune Response to Human Immunodeficiency Virus and Hepatitis C Virus Infections? *Front Immunol.* 2017;8:832.
doi: 10.3389/fimmu.2017.00832
2. Medhasi S, Chantratita N. Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections. *J Immunol Res.* 2022;2022:9710376.
doi: 10.1155/2022/9710376
3. Bose T, Pant N, Pinna NK, Bhar S, Dutta A, Mande SS. Does immune recognition of SARS-CoV2 epitopes vary between different ethnic groups? *Virus Res.* 2021;305:198579.
doi: 10.1016/j.virusres.2021.198579
4. Yu MR, Kim HJ, Park HR. *Fusobacterium nucleatum* Accelerates the Progression of Colitis-Associated Colorectal Cancer by Promoting EMT. *Cancers.* 2020;12(10):2728.
doi: 10.3390/cancers12102728
5. Brennan CA, Garrett WS. *Fusobacterium nucleatum* — symbiont, opportunist and oncobacterium. *Nat Rev Microbiol.* 2019;17(3):156–166.
doi: 10.1038/s41579-018-0129-6
6. Mark Welch JL, Ramírez-Puebla ST, Borisy GG. Oral Microbiome Geography: Micron-Scale Habitat and Niche. *Cell Host Microbe.* 2020;28(2):160–168.
doi: 10.1016/j.chom.2020.07.009
7. Fan Z, Tang P, Li C, *et al.* *Fusobacterium nucleatum* and its associated systemic diseases: epidemiologic studies and possible mechanisms. *J Oral Microbiol.* 2023;15(1):2145729.
doi: 10.1080/20002297.2022.2145729
8. Ghosh A, Jaaback K, Boulton A, *et al.* *Fusobacterium nucleatum*: An Overview of Evidence, Demi-Decadal Trends, and Its Role in Adverse Pregnancy Outcomes and Various Gynecological Diseases, including Cancers. *Cells.* 2024;13(8):717.
doi: 10.3390/cells13080717
9. Han YW. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Curr Opin Microbiol.* 2015;23:141–147.
doi: 10.1016/j.mib.2014.11.013

10. Yang C, Chowdhury D, Zhang Z, *et al.* A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data. *Comput Struct Biotechnol J.* 2021;19:6301-6314.
doi: 10.1016/j.csbj.2021.11.028
11. Tang W, Li F, Zheng H, *et al.* Unveiling hidden players: the role of intratumoral microbiota in gastrointestinal cancer dynamics. *J Cancer Res Clin Oncol.* 2025;152(1):15.
doi: 10.1007/s00432-025-06401-9
12. Mehmood MS, Iqbal I, Hajj F. Intratumoral microbiota as prognostic biomarkers in gastrointestinal cancers. *Ann Med Surg.* 2025;87(12):9137-9138.
doi: 10.1097/MS9.0000000000004144
13. Ganesan K, Guo S, Fayyaz S, Zhang G, Xu B. Targeting Programmed Fusobacterium nucleatum Fap2 for Colorectal Cancer Therapy. *Cancers.* 2019;11(10):1592.
doi: 10.3390/cancers11101592
14. Mesa F, Mesa-López MJ, Egea-Valenzuela J, *et al.* A New Comorbidity in Periodontitis: Fusobacterium nucleatum and Colorectal Cancer. *Medicina (Mex).* 2022;58(4):546.
doi: 10.3390/medicina58040546
15. Ma X, Sun T, Zhou J, *et al.* Pangenomic Study of Fusobacterium nucleatum Reveals the Distribution of Pathogenic Genes and Functional Clusters at the Subspecies and Strain Levels. *Microbiol Spectr.* 2023;11(3):e0518422.
doi: 10.1128/spectrum.05184-22
16. Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2018;47(D1):D339-D343. Available from: <https://academic.oup.com/nar/article/47/D1/D339/5144151?login=false> [Last accessed on 2025 Aug 21].
17. Altschul SE, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403-410.
doi: 10.1016/S0022-2836(05)80360-2
18. Kuntal BK, Mande SS. Web-igloo: a web based platform for multivariate data visualization. *Bioinformatics.* 2017;33(4):615-617.
doi: 10.1093/bioinformatics/btw669
19. Zeddou M. Class I HLA Allele Predicted Restricted Antigenic Coverages for Fap2 Protein of Fusobacterium Nucleatum Are Associated with Colorectal Cancer Incidence. *Asian Pac J Cancer Prev.* 2023;24(10):3629-3636.
doi: 10.31557/APJCP.2023.24.10.3629
20. Zepeda-Rivera M, Minot SS, Bouzek H, *et al.* A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature.* 2024;628(8007):424-432.
doi: 10.1038/s41586-024-07182-w
21. Sconocchia G, Eppenberger-Castori S, Zlobec I, *et al.* HLA class II antigen expression in colorectal carcinoma tumors as a favorable prognostic marker. *Neoplasia.* 2014;16(1):31-42.
doi: 10.1593/neo.131568
22. Loroña NC, Santiago-Torres M, Lopez-Pentecost M, *et al.* Traditional Mexican dietary pattern and cancer risk among women of Mexican descent. *Cancer Causes Control.* 2024;35(6):887-896.
doi: 10.1007/s10552-024-01849-5
23. Santiago-Torres M, Tinker LE, Allison MA, *et al.* Development and Use of a Traditional Mexican Diet Score in Relation to Systemic Inflammation and Insulin Resistance among Women of Mexican Descent. *J Nutr.* 2015;145(12):2732-2740.
doi: 10.3945/jn.115.213538
24. Cuellar-Gómez H, Ocharán-Hernández ME, Calzada-Mendoza CC, Comoto-Santacruz DA. Association of Fusobacterium nucleatum infection and colorectal cancer: A Mexican study. *Rev Gastroenterol Méx Engl Ed.* 2022;87(3):277-284.
doi: 10.1016/j.rgmexn.2021.07.001
25. Paredes J, Zabaleta J, Garai J, *et al.* Immune-Related Gene Expression and Cytokine Secretion Is Reduced Among African American Colon Cancer Patients. *Front Oncol.* 2020;10:1498.
doi: 10.3389/fonc.2020.01498
26. Lo CH, Blot WJ, Teras LR, *et al.* Prediagnostic Antibody Responses to Fusobacterium nucleatum Proteins Are Not Associated with Risk of Colorectal Cancer in a Large U.S. Consortium. *Cancer Epidemiol Biomarkers Prev.* 2021;30(6):1279-1282.
doi: 10.1158/1055-9965.EPI-20-1471
27. Hong M, Li Z, Liu H, *et al.* Fusobacterium nucleatum aggravates rheumatoid arthritis through FadA-containing outer membrane vesicles. *Cell Host Microbe.* 2023;31(5):798-810.e7.
doi: 10.1016/j.chom.2023.03.018
28. Dong L, Zheng Q, Cheng Y, *et al.* Gut Microbial Characteristics of Adult Patients With Epilepsy. *Front Neurosci.* 2022;16:803538.
doi: 10.3389/fnins.2022.803538
29. Lorente E, Redondo-Antón J, Martín-Esteban A, *et al.* Substantial Influence of ERAP2 on the HLA-B*40:02 Peptidome: Implications for HLA-B*27-Negative Ankylosing Spondylitis. *Mol Cell Proteomics.* 2019;18(11):2298-2309.
doi: 10.1074/mcp.RA119.001710
30. Weingarten-Gabbay S, Klaeger S, Sarkizova S, *et al.* Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell.* 2021;184(15):3962-3980.e17.
doi: 10.1016/j.cell.2021.05.046

31. Fujiwara K, Shao Y, Niu N, *et al.* Direct identification of HLA class I and class II-restricted T cell epitopes in pancreatic cancer tissues by mass spectrometry. *J Hematol Oncol.* 2022;15(1):154.
doi: 10.1186/s13045-022-01373-6
32. Marcu A, Bichmann L, Kuchenbecker L, *et al.* HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J Immunother Cancer.* 2021;9(4):e002071.
doi: 10.1136/jitc-2020-002071
33. Henne K, Schilling H, Stoneking M, Conrads G, Horz HP. Sex-specific differences in the occurrence of *Fusobacterium nucleatum* subspecies and *Fusobacterium periodonticum* in the oral cavity. *Oncotarget.* 2018;9(29):20631-20639.
doi: 10.18632/oncotarget.25042

Appendix

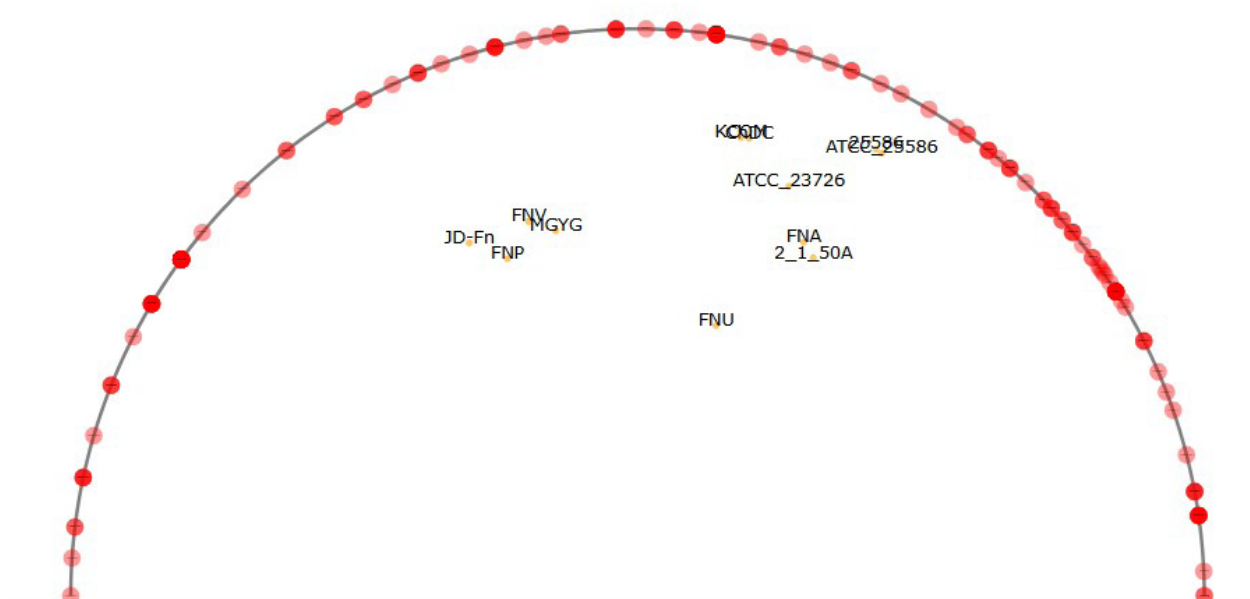


Figure A1. The clustering pattern of the studied *Fusobacterium nucleatum* strains based on the presence of epitopes. Red dots on the semi-circle represent the epitopes, and the strains are projected inside the semi-circle using a Hooke's law-based approach.