A newly developed PCR-based method revealed distinct *Fusobacterium nucleatum* subspecies infection patterns in colorectal cancer

Dexi Bi,† Yin Zhu,‡† Yaohui Gao,† Hao Li,‡ Xingchen Zhu,† Rong Wei,† Ruting Xie,† Qing Wei†,* and Huanlong Qin‡**

†Department of Pathology, Shanghai Tenth People’s Hospital, Tongji University School of Medicine, Shanghai, 200072, China.
‡Department of Gastrointestinal Surgery, Shanghai Tenth People’s Hospital, Tongji University School of Medicine, Shanghai, 200072, China.

Summary

*Fusobacterium nucleatum*, which has four subspecies (*nucleatum, animalis, vincentii* and *polymorphum*), plays an important role in promoting colorectal cancer (CRC). However, as there is no efficient method of differentiating these subspecies in the context of a rich gut microbiota, the compositions in CRC remain largely unknown. In this study, a PCR-based differentiation method enabling profiling of *F. nucleatum* infection in CRC at the subspecies level was developed. Based on the analysis of 53 *F. nucleatum* genomes, we identified genetic markers specific to each subspecies and designed primers for the conserved sequences of those markers. The PCR performance of the primers was tested with *F. nucleatum* and non-*nucleatum* *Fusobacterium* strains, and complete consistency with taxonomy was achieved. Additionally, no non-specific amplification occurred when using human DNA. The method was then applied to faecal (*n = 58*) and fresh-frozen tumour tissue (*n = 100*) samples from CRC patients, and wide heterogeneity in *F. nucleatum* subspecies compositions in the gut microbiota among CRC patients was observed. Single-subspecies colonization was common, whereas coexistence of four subspecies was rare. Subspecies *animalis* was most prevalent, while *nucleatum* was not frequently detected. The results of this study contribute to our understanding of the pathogenicity of *F. nucleatum* at the subspecies level and the method developed has potential for clinical and epidemiological use.

Introduction

*Fusobacterium nucleatum* has long been recognized as an opportunistic oral pathogen. However, knowledge about its pathogenicity is being updated, as emerging evidence has uncovered that it is also involved in a wide range of diseases, especially colorectal cancer (CRC), in humans (Brennan and Garrett, 2019, Lopez, et al., 2020). *F. nucleatum* is a commensal of the oral cavity and a key causative agent of oral diseases such as periodontitis (Roques, et al., 2000). Furthermore, it acts as a bridging microbe that can coaggregate other pathogens, such as Porphyromonas gingivalis and Aggregatibacter actinomycetemcomitans (Kolenbrander, et al., 2010). With high invasiveness and a tendency towards translocation, *F. nucleatum* can also induce extraoral infections, such as those involved in preterm birth and pregnancy complications (Han, 2015). More importantly, recent studies have demonstrated that *F. nucleatum* infection is not only prevalent in but also has a close association with CRC (Brennan and Garrett, 2019), emphasizing the urgent need for continued research into its exact roles and clinical relevance in this malignant disease.

*F. nucleatum* has well-recognized genotypic and phenotypic heterogeneity (Thurnheer, et al., 1999; Strauss, et al., 2008). Sequencing-facilitated phylogenetic studies have greatly advanced the taxonomy of *F. nucleatum*, which currently includes four subspecies: *nucleatum, animalis, vincentii* and *polymorphum* (Mira, et al., 2004; Strauss, et al., 2008; Kim, et al., 2010; Kook, et al., 2017, Brennan and Garrett, 2019). It should be noted that *vincentii* and the formerly proposed *fusiforme* are now considered as one subspecies (Kim, et al., 2010, Kook, et al., 2013). These subspecies display relatively high genetic divergence (Kook, et al., 2017), with possible variable phenotypic patterns in growth (Chew, et al., 2012), habitat preference (Thurnheer, et al., 2019) and pathogenicity (Kurgan, et al., 2017). Subspecies...
Fusobacterium nucleatum subspecies differentiation

Fusobacterium nucleatum is mostly found in periodontal diseases, and animalis and polymorphum are often associated with extraoral infectious diseases; in contrast, vincentii is usually considered as a commensal of the healthy oral flora (Han and Wang, 2013). The different subspecies also display distinct biofilm behaviours (Thurnheer, et al., 2019) and can differentially modulate neutrophil function (Kurgan, et al., 2017). Moreover, a recent study suggested that the subspecies animalis might be prevalent in CRC tissues and influence proinflammatory cytokine expression and monocyte activation (Ye, et al., 2017). However, the pathogenic effects and oncogenic mechanisms of particular subspecies in CRC are mostly unknown, and it remains unclear whether these subspecies have different levels of pathogenicity and/or play different roles in CRC. Hence, accurate classification of F. nucleatum subspecies could be of great importance, as it will help to further reveal pathogenic F. nucleatum members, establish precise etiological relationships with CRC and other associated diseases and ultimately benefit clinical practice and epidemiological investigation.

To date, various efforts to differentiate F. nucleatum subspecies have been made. Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) has been employed for subspecies identification, and among 34 F. nucleatum clinical isolates, it performed accurately for 28 (82%) but misclassified or failed to classify the remaining isolates (Nie, et al., 2015). An arbitrarily primed PCR (AP-PCR) method has also been developed to subtype F. nucleatum, but it was designed for isolate comparison rather than subspecies typing (George, et al., 1997). In addition, both of the above methods require successful isolation prior to use and cannot be applied directly to microbiota-containing samples such as faeces and tumour tissues from CRC patients. Although 16S rRNA gene amplicon sequencing and metagenomic sequencing are frequently used for microbiota profiling, neither has achieved subspecies-level resolution for F. nucleatum (He, et al., 2021). Yeoh, et al., 2020). A 16S rRNA gene-targeted sequencing method determining the V2-V4 region was also used to identify specific F. nucleatum subspecies in CRC tumour specimens, but with the requirement of multistep PCR, sequencing and complex analysis (Ye, et al., 2017). Additionally, F. nucleatum shows extensive 16S rRNA genes conservation (Mira, et al., 2004; Kim, et al., 2010), and a short partial region would hardly provide a clear resolution (He, et al., 2021). Accordingly, there is still a lack of genotyping methods to accurately and intuitively differentiate F. nucleatum subspecies.

By taking advantage of large-scale comparative genomics, this study has identified subspecies-specific genetic markers and subsequently developed a PCR-based typing assay that allows easy differentiation of F. nucleatum subspecies in CRC.

Results

Large-scale comparative genomic analysis identified subspecies-specific markers in F. nucleatum

The genomes of 54 unique F. nucleatum strains from the NCBI genome database were used in this study (Table S1). Whole-genome phylogenetic analysis was performed to ensure correct subspecies assignment. The strains formed five clades in the phylogenetic tree: four corresponded well to the subspecies nucleatum, animalis, vincentii and polymorphum, but one showed a distant relationship to these four clades and constituted only strain W1481 (Fig 1). It has been suggested that strain W1481 does not belong to F. nucleatum (Ang, et al., 2016), and it was thus excluded from subsequent analysis and reassigned as Fusobacterium sp. W1481. The 53 F. nucleatum strains, including 6 subsp. nucleatum, 17 subsp. animalis, 10 subsp. vincentii and 20 subsp. polymorphum, were evaluated by comparative analysis to search for specific regions (SRs) that can act as subspecies markers. Genes or non-coding sequences occurring in all genomes of one subspecies but completely absent in other subspecies and in non-F. nucleatum organisms were selected as SRs. Six (designated SR-n1 to SR-n6) and three (designated SR-a1 to SR-a3) genes were found to be specific to the subspecies nucleatum and animalis respectively (Fig 2). We did not find any subsp. vincentii-specific genes, but instead identified two intergenic regions (IGRs) specific to this subspecies (designated SR-v1 and SR-v2) (Fig 2). With regard to subspecies polymorphum, we were unable to obtain any specific gene or IGR that could independently cover all its members. However, we detected three genes (designated SR-p1 to SR-p3) specific to different subsets of subsp. polymorphum strains that could be used in combination (SR-p1 + SR-p2 or SR-1 + SR-p3) to achieve full-member coverage (Fig 2). Importantly, inter- and intra-subspecies comparisons showed that the identified markers (or marker combinations) exhibited strict subspecies specificity, with highly conserved sequences (with 87.6%-100% identities to reference strains; median: 99.2%) (Fig 2). Homologs of these markers were not found in other organisms, as examined in the NCBI nucleotide database. There was only one exception: SR-v2 exhibits 100% identity with Fusobacterium sp. oral taxon 203. Note that for intersubspecies comparison and wider comparison involving other organisms, somewhat homologous hits were inevitable when non-strict parameters were used, but they showed only limited identity and/or coverage.

© 2021 The Authors. Microbial Biotechnology published by Society for Applied Microbiology and John Wiley & Sons Ltd., Microbial Biotechnology, 14, 2176–2186
Development of a PCR-based assay for robust F. nucleatum subspecies differentiation

Universal primers were designed for the identified markers within their consensus or near-consensus regions (Table 1). For practical use, all primers had similar Tms, and the theoretical amplicon sizes were 100 bp–400 bp (mostly <300 bp). The performance of the designed primers was assessed using F. nucleatum strains spanning the four subspecies and strains of other Fusobacterium species, including F. hwasookii, F. mortiferum, F. periodonticum and F. varium. Remarkably, the amplification results were exactly in accord with the taxonomy of these strains (Fig 3). For subspecies nucleatum, animalis and vincentii markers, each primer set resulted in amplification for its target subspecies of a clear single band at the expected position. For markers of subspecies polymorphum, we obtained positive amplification for SR-p1 and SR-p3, but not SR-p2 with the strains used, which was not unexpected, as we have shown that these markers could not independently cover the entire subspecies. The results for combinations SR-p1 + SR-p2 (expected as being positive for either SR-p1 or SR-p2 or both) and SR-p1 + SR-p3 (positive for either SR-p1 or SR-p3 or both) matched well with this subspecies. None of the designed primers yielded amplification with strains of other subspecies or other Fusobacterium species or human DNA. In addition, we examined the primer set (Fn-F/R) (Castellarin, et al., 2012; Mima, et al., 2015) widely used for F. nucleatum detection in these strains, and it also produced bright bands when using F. hwasookii and F. periodonticum. Indeed, sequence analysis revealed that its target sequence is not strictly specific to F. nucleatum and is found in F. hwasookii, F. periodonticum and Fusobacterium pseudoperiodonticum (Fig S1).

F. nucleatum subspecies detection in the gut microbiota of colorectal cancer (CRC) patients

With the developed method, we profiled the subspecies composition of F. nucleatum in the gut microbiota of CRC patients. One representative primer set or set combination was used for each subspecies according to the performance in bacterial strains (Fig 3). Both faeces (n = 58) and fresh-frozen tumour tissues (n = 100) were evaluated (Figs 4 and 5). All positive amplifications yielded a single band at the expected position, and a non-specific band was rarely observed. Interestingly, we found diverse F. nucleatum subspecies composition
patterns across the samples (Figs 4A and 5A), varying in types and numbers of colonized subspecies. *F. nucleatum* was detected in 31 (53.5%) faecal and 50 (50.0%) tumour samples. Among them, colonization by one subspecies was mostly observed (14 faeces and 31 tumours); colonization by two subspecies was also common (12 faeces and 14 tumours), whereas colonization by three or four subspecies was detected only in a limited number of samples (5 faeces and 5 tumours). Moreover, the four subspecies showed distinct detection rates, and their distribution patterns were similar in the two types of samples (Figs 4B and 5B). Detected in 37.9% of faeces and 30.0% of tumours, *animalis* was consistently the most prevalent subspecies, followed by *polymorphum* and *vincentii*; *nucleatum* had the lowest detection rates (6.9% and 7.0% in faeces and tumours respectively). In addition, we again found that the primer set Fn-F/R lacked sufficient specificity, resulting in amplification when using samples negative for subspecies-specific primers (Figs 4 and 5).

**Discussion**

In this study, we developed a PCR-based method to discriminate *F. nucleatum* subspecies by probing subspecies-specific genetic markers, and with this method, we uncovered that the gut microbiota of CRC patients shows heterogeneous *F. nucleatum* subspecies patterns.

We obtained fourteen subspecies-specific markers for *F. nucleatum*, including twelve genes and two IGRs. Among the genes, four are predicted to encode a transcriptional regulator and enzymes likely participating in metabolism; eight have unknown functions (Table 2). These genes might be involved in biological behaviours that are exclusive to the corresponding subspecies. Nevertheless, the intersubspecies difference cannot be explained only by these markers, as genome-wide nucleotide variations are predominant (Kook, et al., 2017). Additionally, other undiscovered subspecies-specific genetic markers may exist, as our analysis included unfinished genomes. We next designed primers targeting the identified markers and tested them on *Fusobacterium* strains. All primer sets exhibited satisfactory performance. The results indicated that for subspecies *nucleatum, animalis* and *vincentii* markers, each primer set can be independently used to probe its targeted subspecies; conversely, combined sets should be used for markers of subsp. *Polymorphum*. Given that we have not yet obtained positive amplification with P2-F/R, the combination P1-F/R + P3-F/R (for SR-p1 + SR-p3) is recommended.

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Microbial Biotechnology*, 14, 2176–2186
TABLE 1. Primers designed to differentiate *Fusobacterium nucleatum* subspecies.

| Subspecies | Marker | Primers | Sequences (5’ to 3’) | Amplicon size (bp) |
|------------|--------|---------|----------------------|-------------------|
| *nucleatum* | SR-n1 | N1-F | GAGATGGAAAGTACAAACAAATTACAGTAG | 196 |
|           |       | N1-R | TCTTGGACATCACAATCATAAAACAAGA | 285 |
|           | SR-n2 | N2-F | GAGTTTACACCTGCTATATTGGACACCA | 216 |
|           |       | N2-R | CAAAACATTGTAATTTCTCTGCAAT | 216 |
|           | SR-n3 | N3-F | CAAGCAACTGAAAATGCTTTAAAG | 183 |
|           |       | N3-R | TCCAGGTAAGAAAATTACACTACTG | 124 |
|           | SR-n4 | N4-F | GGGGCTACATTITTTCTACTCTC | 241 |
|           |       | N4-R | TCTATCTACACCTGACGACCTTT | 102 |
|           | SR-n5 | N5-F | AAACTTCTGCTATCCTATGCTTTAA | 189 |
|           |       | N5-R | CTTTACCAAGTTAGGTTCTATCCATAAT | 224 |
|           | SR-n6 | N6-F | GATAGACAACAGAAGAAAGAGACT | 196 |
|           |       | N6-R | GTATCCCTATTCCTACAAAGAAGTC | 107 |
| *animals*  | SR-a1 | A1-F | TTTGTTTATGTTGACATGCTCTT | 124 |
|           |       | A1-R | CATCTGCTGTTATGATCCCTG | 224 |
|           | SR-a2 | A2-F | AACTAATTATTATGATGATGAAAGA | 189 |
|           |       | A2-R | GCATCTGAAAGGATGAAATGCTG | 125 |
|           | SR-a3 | A3-F | ATTTTTATCCTGTGTTATGATCAAT | 241 |
|           |       | A3-R | ACAAATTCTAYGGATATTGACAGATT | 118 |
| *vincentii* | SR-v1 | V1-F | GAGGCTATGCAAAATTAACGCTG | 189 |
|           |       | V1-R | GCTTACCTATTATAAATATGAC | 118 |
|           | SR-v2 | V2-F | GAGCAARYSAGAAGTAGGGCA | 183 |
|           |       | V2-R | GACCTTCACATTCTGAAARCAACCA | 183 |
| *polymorphum* | SR-p1 | P1-F | TTAGGAAATCTTTTAGACCA | 107 |
|           |       | P1-R | TCTACTGTAAATTTGACCTG | 107 |
|           | SR-p2 | P2-F | TGTGCTCTACCTCCTTGTATTG | 375 |
|           |       | P2-R | TACCGAYCTTTTTTGGCTCTTGT | 375 |
|           | SR-p3 | P3-F | CTTATGYYTTTGGACCTATTG | 375 |
|           |       | P3-R | CCAAGTAAATTTAGGCTCTTGAAC | 375 |

In samples such as faeces and tumour tissues from CRC patients, isolation of *F. nucleatum* is tedious due to the complex and rich commensal microbiota context and lack of highly selective medium (Gur, et al., 2015; Komiya, et al., 2019; Abed, et al., 2020). Therefore, the implementation of isolate-based classification approaches for such samples can hardly be achieved in clinical practice. Moreover, the sequencing techniques currently applied to microbiota-containing samples have not yet achieved sufficient resolution to distinguish *F. nucleatum* subspecies (He, et al., 2021, Yeoh, et al., 2012; Mima, et al., 2015). The widely used Fn-F/R primers (Castellarin, et al., 2012; Mima, et al., 2015) are only able to detect *F. nucleatum* at the species level, and we also found that they are not strictly specific to this species. In this study, we adopted a stringent strategy to ensure that the markers identified and corresponding primer sets have strict specificity as they were cross-compared against not only the *F. nucleatum* genomes but also the sequences of other organisms in the public database, making it possible to implement our method for both *F. nucleatum* isolates and microbiota-containing samples. Indeed, the primers resulted in no amplification in human DNA and did not produce non-specific bands when using faecal or tissue samples.

*F. nucleatum* is known to play an important role in promoting CRC, but studies at the subspecies level are scarce (Ye, et al., 2017; Richardson, et al., 2020, partly due to the lack of an efficient approach to clearly identify *F. nucleatum* subspecies in the gut microbiota. Remarkably, our method provided a well-resolved view of the composition of *F. nucleatum* subspecies in CRC. Given the heterogeneity within *F. nucleatum*, it is worth exploring whether composition diversity has an impact on the pathogenesis of CRC and whether it can be linked to tumour heterogeneity. Notably, subspecies *animals* was most prevalent in the gut of CRC patients, whereas *nucleatum* ATCC 25586 and ATCC 23726 strains, it is worth exploring whether composition diversity has an impact on the pathogenesis of CRC and whether it can be linked to tumour heterogeneity. Notably, subspecies *animals* was most prevalent in the gut of CRC patients, whereas *nucleatum*, which has been extensively used as a model organism (such as its ATCC 25586 and ATCC 23726 strains) in CRC studies, actually had a much lower prevalence than the other subspecies, suggesting the necessity of comparing the pathogenicity of different subspecies and choosing more representative model organisms in future CRC studies. Furthermore, *F. nucleatum* subspecies determination might contribute to strain-level studies, as it may help to narrow down the taxonomic range of pathogenic strains. In addition, there were significant variations in the detection rate of *F. nucleatum* among CRC patients across studies, owing to not only the regional and ethnic differences but also the different primers used targeting either *Fusobacterium* spp. or specifically *F. nucleatum* (Lee, et al., 2019). Therefore, the degree to which *F. nucleatum* potentiates tumorigenesis still requires in-depth investigation.
method can be implemented using various types of specimens to further elucidate the distribution and clinical relevance of *F. nucleatum* at the subspecies level in large clinical and epidemiological studies with increased credibility of detection.

To our knowledge, a genotyping method for *F. nucleatum* subspecies classification, such as ours, has not been developed to date. Compared with MALDI-TOF MS, this method can be easily implemented because it does not require special equipment (Singhal, *et al*., 2015). However, due to the limited resources of isolates, further assays using more *F. nucleatum* strains are needed to validate the performance of this method. In addition, the current procedure requires separate PCRs for each primer set, leading to increased consumption of the DNA template and repetitive labour. Further studies should explore the feasibility of multiplex PCR based on these markers.

**Experimental procedures**

**Ethics**

This study was approved by the Ethics Committee of Shanghai Tenth People’s Hospital (No. SHSY-IHC-4.1/
Written consent was obtained from each participant.

**Genomic data**

A total of 68 *F. nucleatum* genomes were retrieved from the NCBI database (ftp://ftp.ncbi.nih.gov/genomes/). Duplicated genomes, unannotated genomes or those not meeting the RefSeq criteria were excluded. After selection, 54 unique genomes, including 14 complete and 40 draft genomes, were included for analysis (Table S1). The phylogeny of those genomes was analysed with kSNP3 (Gardner, *et al.*, 2015).

**Bacterial strains**

*F. nucleatum* subsp. *nucleatum* ATCC 25586 and ATCC 23726 were purchased commercially. Nineteen *Fusobacterium* strains previously isolated from freshly resected tumour tissues of CRC cancer patients were also used, including four *F. nucleatum* subsp. *animalis* (THCT5A4, THCT6B3, THCT7A2 and THCT5A5), two *F. nucleatum* subsp. *polymorphum* (THCT7E2 and THCT15E1), two *F. nucleatum* subsp. *vincentii* (THCT14A3 and THCT14B3), one *Fusobacterium hwasookii* (THCT14E2), one *Fusobacterium periodonticum* (THCT6B2), one *Fusobacterium periodonticum* (THCT18E1) and eight *F. varium* strains (THCT1E1, THCT1E2, THCT4E2, THCT4E4, THCT13E1, THCT23E1, THCT23B1 and THCT23E3). The bacteria were cultivated on Columbia agar supplemented with 5% sheep blood (Comagal Microbial).

**Clinical specimens**

Faecal (n = 58) or fresh-frozen tumour tissue (n = 100) samples were collected from 158 newly diagnosed primary CRC patients (aged >18 years) at Shanghai Tenth People’s Hospital between March 2017 and January 2020. The samples were retrieved from the...
Biobank of Shanghai Tenth People’s Hospital. The faecal samples had been collected upon admission. Patients who had received antibiotics or probiotics within one month prior to sample collection were excluded.

Identification of subspecies-specific markers and primer design

*Fusobacterium nucleatum* genomes were subject to pangenome analysis with PGAP to sort homologous genes into

© 2021 The Authors. *Microbial Biotechnology* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Microbial Biotechnology*, 14, 2176–2186
clusters with the GeneFamily method (Zhao, et al., 2012). Gene clusters occurring in all genomes of one subspecies but completely absent in other subspecies were selected. The genes in the selected clusters were then aligned with all the genomes to verify whether they were selected. The genes in the selected clusters were also compared against the NCBI nucleotide database to exclude those with homologous sequences in non-F. nucleatum organisms. Clusters having genes with truncation or disruption were also excluded. The clusters ultimately included were defined as subspecies-specific markers. For subspecies with no gene cluster meeting the above criteria, markers were similarly searched for in intergenic regions. If both strategies failed, gene clusters specific to a subspecies but not occurring in all its genomes were selected as markers to use in combination to cover all subspecies members. For each defined marker, multiple sequence alignment of its allelic sequences was performed with MUSCLE (Edgar, 2004). Primers were designed to have a similar melting temperature (Tm) within a consensus or near-consensus region. The primer sequences were compared against the NCBI nucleotide database to assess putative non-specific amplification.

**DNA preparation**

Genomic DNA of bacterial strains was prepared with the TiANamp Bacteria DNA Kit (TIANGEN Biotech) according to the manufacturer’s instructions. Total DNA from faeces and tumour tissues was prepared with the cetyltrimethylammonium bromide (CTAB)-based method.

**Polymerase chain reaction**

Polymerase chain reaction (PCR) was conducted with Premix Taq™ (Ex Taq™ Version 2.0 plus dye, TAKARA) using a Mastercycler nexus gradient thermal cycler (Eppendorf) according to the manufacturer’s instructions; 10 ng DNA was used for each reaction. The PCR conditions were 30 cycles (or 35 cycles for faecal and tissue DNA) of 98°C for 30 seconds, 55°C for 30 seconds and 72°C for 20 seconds with the Cool Start procedure unless otherwise specified. The products were electrophoresed on 0.8% or 2% agarose gels as applicable, stained with SYBR Green I (Yeasen Biotech) and visualised with a Tanon 3500 UV gel image system (Tanon Science & Technology). The subspecies-specific primers designed in this study are listed in Table 1. For bacterial strains, 16S rRNA genes were amplified with the universal primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-CGTTACCTTGGTACCTT-3') as a positive control (with a 90 second extension process in each PCR cycle). For faecal and tissue samples, the V6 regions of the 16S rRNA genes were amplified with a set of five primers (5'-CAGCCGGAACCTTANC-3', 5'-ATACGCGARGAACCACTTTACC-3', 5'-CTAAACCAGANACCTTACC-3', 5'-CAGCCGARAAACTTTACC-3', 5'-CGACRRCACTGCANCACCT-3') (Geller, et al., 2017) as a positive control. The primer set widely used to detect F. nucleatum (Fn-F: 5'-CAACCTACTTAACTCTACACGTTGTTACGACTT-3' and Fn-R: 5'-GTTGACTTTACAGAAGGAGATTGTAAATAG-3') (Castellarin, et al., 2012; Mima, et al., 2015) was also included for comparison. Human blood DNA from healthy volunteers was used as a negative control.

**Heatmap**

A heatmap was generated with the ComplexHeatmap (Gu, et al., 2016) package in R. The accompanying dendrogram was drawn with default parameters.

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China under grant 82072236 (to D.B.), 82072634 (to Q.W.) and 81972221 (to H.Q.), Clinical Research Plan of SHDC under grant SHDC 2020CR2069B (to Q.W.) and the Pandeng Research Foundation of Shanghai Tenth People’s Hospital under grant 2018SYPDRC015 (to D.B.).

---

**Table 2. Annotation of the subspecies marker genes.**

| Marker | Representative gene | Product |
|--------|---------------------|---------|
| SR-n1  | C7Y58_07315         | Hypothetical protein |
| SR-n2  | C7Y58_06070         | LysR family transcriptional regulator |
| SR-n3  | C7Y58_06085         | RdA family protein |
| SR-n4  | C7Y58_06080         | Pyridoxal phosphate-dependent aminotransferase |
| SR-n5  | C7Y58_05110         | Hypothetical protein |
| SR-n6  | C7Y58_08045         | Hypothetical protein |
| SR-a1  | FSDG_01049          | Hypothetical protein |
| SR-a2  | FSDG_01599          | M20/M25/M40 family metallo-hydrolase |
| SR-a3  | FSDG_00096          | Hypothetical protein |
| SR-p1  | EII28_03630         | Hypothetical protein |
| SR-p2  | EII28_07115         | Hypothetical protein |
| SR-p3  | EII28_10930         | Hypothetical protein |

Note: SR-v1 and SR-v2 are not listed because they are in intergenic regions.
Conflict of interest

D. B., Y. Z., Q.W. and H.Q are the inventors on a patent application (No. 202110250360.7) filed by the Shanghai Tenth People’s Hospital relating to the F. nucleatum sub-species differentiation method described herein.

REFERENCES

Abed, J., Maalouf, N., Manson, A.L., Earl, A.M., Parhi, L., Ermgard, J.E.M., et al. (2020) Colon cancer-associated Fusobacterium nucleatum may originate from the oral cavity and reach colon tumors via the circulatory system. Front Cell Infect Microbiol 10: 400.

Ang, M.Y., Dutta, A., Wee, W.Y., Dymock, D., Paterson, I.C., and Choo, S.W. (2016) Comparative genome analysis of Fusobacterium nucleatum. Genome Biol Evol 8: 2928–2938.

Brennan, C.A., and Garrett, W.S. (2019) Fusobacterium nucleatum-symbiont, opportunist and oncobacterium. Nat Rev Microbiol 17: 156–166.

Brenneman, D., Warren, R.L., Freeman, J.D., Drellinger, L., Krzywinski, M., Strauss, J., et al. (2012) Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res 22: 299–306.

Colin, H., Zilm, P.S., Fuss, J.M., and Guly, N.J. (2012) A proteomic investigation of Fusobacterium nucleatum alkaline-induced biofilms. BMC Microbiol 12: 189.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

Gardner, S.N., Slezak, T., and Hall, B.G. (2015) kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31: 2877–2878.

Geller, L.T., Barzily-Rokni, M., Danino, T., Jonas, O.H., Shental, N., Nejman, D., et al. (2017) Potential role of intratumor bacteria in mediating tumor resistance to the chemotherapeutic drug gemcitabine. Science 357: 1156–1160.

George, K.S., Reynolds, M.A., and Falkler, W.A. Jr (1997) Arbitrarily primed polymerase chain reaction fingerprinting and clonal analysis of oral Fusobacterium nucleatum isolates. Oral Microbiol Immunol 12: 219–226.

Gu, Z., Eils, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32: 2847–2849.

Gur, C., Ibrahim, Y., Isaacson, B., Yamin, R., Abed, J., Gamliel, M., et al. (2015) Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity 42: 344–355.

Han, Y.W. (2015) Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol 23: 141–147.

Han, Y.W., and Wang, X. (2013) Mobile microbiome: oral bacteria in extra-oral infections and inflammation. J Dent Res 92: 485–491.

He, Y., Mujagond, P., Tang, W., Wu, W., Zheng, H., Chen, X., et al. (2021) Non-nucleatum Fusobacterium species are dominant in the southern Chinese population with distinctive correlations to host diseases compared with F. nucleatum, Gut 70(4): 810–812. http://doi.org/10.1136/gutjnl-2020-322090

Kim, H.S., Lee, D.S., Chang, Y.H., Kim, M.J., Koh, S., Kim, J., et al. (2010) Application of rpoB and zinc protease gene for use in molecular discrimination of Fusobacterium nucleatum sub-species. J Clin Microbiol 48: 545–553.

Kolenbrander, P.E., Palmer, R.J. Jr, Periasamy, S., and Jakubovics, N.S. (2010) Oral multispecies biofilm development and the key role of cell-cell distance. Nat Rev Microbiol 8: 471–480.

Komiya, Y., Shimomura, Y., Higurashi, T., Sugi, Y., Arimoto, J., Umezawa, S., et al. (2019) Patients with colorectal cancer have identical strains of Fusobacterium nucleatum in their colorectal cancer and oral cavity. Gut 68: 1335–1337.

Kook, J. K., Park, S. N., Lim, Y. K., Choi, M. H., Cho, E., Kong, S. W., et al. (2013) Fusobacterium nucleatum subsp. fusiforme Gharia and Shah 1992 is a later synonym of Fusobacterium nucleatum subsp. vincentii Dzink et al. 1990. Curr Microbiol 66: 414–417.

Kook, J. K., Park, S. N., Lim, Y.K., Cho, E., Jo, E., Roh, H., et al. (2017) Genome-based reclassification of Fusobacterium nucleatum sub-species at the species level. Curr Microbiol 74: 1137–1147.

Kurgan, S., Kansal, S., Nguyen, D., Stephens, D., Koroneos, Y., Hasturk, H., et al. (2017) Strain-specific impact of Fusobacterium nucleatum on neutrophil function. J Periodontol 88: 380–389.

Lee, S.A., Liu, F., Riordon, S.M., Lee, C.S., and Zhang, L. (2019) Global investigations of Fusobacterium nucleatum in human colorectal cancer. Front Oncol 9: 566.

Lopez, L.R., Kleel, R.M., and Arthur, J.C. (2020) Microbiota effects on carcinogenesis: initiation, promotion and progression. Annu Rev Med 72.

Mima, K., Sükawa, Y., Nishihara, R., Qian, Z.R., Yamauchi, M., Inamura, K., et al. (2015) Fusobacterium nucleatum and T cells in colorectal carcinoma. JAMA Oncol 1: 653–661.

Mira, A., Pushker, R., Legault, B.A., Moreira, D., and Rodriguez-Valera, F. (2004) Evolutionary relationships of Fusobacterium nucleatum based on phylogenetic analysis and comparative genomics. BMC Evol Biol 4: 50. https://doi.org/10.1186/1471-2148-4-50

Nie, S., Tian, B., Wang, X., Pincus, D.H., Welker, M., Gilhuus-Melhus, A., Pushker, R., Legault, B.A., Moreira, D., and Rodrigues-Valera, F. (2015) Fusobacterium nucleatum sub-species identification by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 53: 1399–1402.

Richardson, M., Ren, J., Rubininstein, M.R., Taylor, J.A., Friedman, R.A., Shen, B., and Han, Y.W. (2020) Analysis of 16S rRNA gene reveals reduced Fusobacterial community diversity when translocating from saliva to GI sites. Gut Microbes 12: 1–13.

Roques, C.G., El kaddouri, S., Barhet, P., Duffort, J. F., and Arellano, M. (2000) Fusobacterium nucleatum involvement in adult periodontitis and possible modification of strain classification. J Periodontol 71: 1144–1150.

Singhal, N., Kumar, M., Kanaujia, P.K., and Virdi, J.S. (2015) MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol 6: 791.

Strauss, J., White, A., Ambrose, C., McDonald, J., and Allen-Vercoe, E. (2008) Phenotypic and genotypic differences in Fusobacterium nucleatum sub-species differentiation. 2185
analyses of clinical *Fusobacterium nucleatum* and *Fusobacterium periodonticum* isolates from the human gut. *Anaerobe* 14: 301–309.

Thurnheer, T., Guggenheim, B., Gruica, B., and Gmür, R.J.A. (1999) Infinite serovar and ribotype heterogeneity among oral *Fusobacterium nucleatum* strains? *Anaerobe* 5: 79–92. https://doi.org/10.1006/anae.1999.0188

Thurnheer, T., Karygianni, L., Flury, M., and Belibasakis, G.N. (2019) *Fusobacterium* species and subspecies differentially affect the composition and architecture of supra- and subgingival biofilms models. *Front Microbiol* 10: 1716.

Ye, X., Wang, R., Bhattacharya, R., Boulbes, D.R., Fan, F., Xia, L., et al. (2017) *Fusobacterium nucleatum* subspecies *animalis* influences proinflammatory cytokine expression and monocyte activation in human colorectal tumors. *Cancer Prev Res (Phila)* 10: 398–409.

Yeoh, Y.K., Chen, Z., Wong, M.C.S., Hui, M., Yu, J., Ng, S.C., et al. (2020) Southern Chinese populations harbour non-nucleatum Fusobacteria possessing homologues of the colorectal cancer-associated FadA virulence factor. *Gut* 69: 1998–2007.

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012) PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28: 416–418.

**Supporting information**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1.** Genomes used in this study.

**Table S1.** Regions targeted by the widely used *Fusobacterium nucleatum* detection primer set have homologs in other *Fusobacterium* species.