Forni *et al. Gut Pathogens* (2025) 17:61
https://doi.org/10.1186/s13099-025-00740-1

Gut Pathogens

RESEARCH

Open Access

Positive selection at core genes may underlie niche adaptation in *Fusobacterium animalis*

Diego Forni[1*], Audun Sivertsen[2], Rachele Cagliani[1], Alessandra Mozzi[1], Cristian Molteni[1], Øyvind Kommedal[2] and Manuela Sironi[3]

Abstract

Background *Fusobacterium animalis* (*Fa*) was identified as the most enriched *Fusobacterium* species in colorectal cancer (CRC). Recently, a group of *Fa* core genes were found to be highly expressed intratumorally and to favor intracellular survival. We hypothesized that, because they promote bacterial fitness in the intracellular niche, these genes might be targets of positive selection, a process that often underlies adaptation to variable environments.

Results We performed an evolutionary analysis to identify selective events that occurred over different time frames, namely during the divergence of the *Fusobacterium* species and in the more recent separation of the *Fa* lineage from *F. paranimalis*. Results indicated that the coding sequences of these genes have been targeted by intense purifying selection, possibly as the result of their often-essential functions. However, localized signatures of positive selection were also detectable. During the divergence of *Fusobacterium* species, the major target of positive selection was represented by elongation factor-Tu, a finding that may be related to its moonlighting functions in adhesion and biofilm development. Additional targets were RpoC and the septum-determining protein MinD. We suggest that variations in the latter contribute to the observed differences in cell length and width between *F. watanabei* and *Fa*. We also searched for and detected beneficial changes that occurred specifically in the *Fa* lineage, suggesting that such variants promote intracellular growth or adaptation to the tumor microenvironment. The strongest target of selection was DnaK, which was shown to promote malignant transformation in other bacterial systems. Analysis of the selected sites in DnaK indicated that most of them are located in the C-terminal unstructured region and that they determine the appearance of eukaryotic linear motifs (ELMs). Specifically, one ELM is a casein kinase 2 phosphorylation site, whereas two additional ELMs are involved in SUMOylation and USP7-mediated deubiquitination. USP is a central modulator of the p53-MDM2 pathway and we propose that SUMOylation facilitates the nuclear import of *Fa* DnaK where USP7 promotes its stability.

Conclusion We identified specific proteins and amino acid changes that are expected to underlie phenotypic diversity in *Fusobacteria*. These data are relevant to inform future analyses of *Fa* oncogenic potential.

Keywords Fusobacterium, Positive selection, Adaptation, DnaK

*Correspondence:
Diego Forni
diego.forni@lanostrafamiglia.it
[1] Scientific Institute IRCCS E. MEDEA, Computational Biology Unit, 23842 Bosisio Parini, Italy
[2] Department of Microbiology, Haukeland University Hospital, Bergen, Norway
[3] School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

Background

Fusobacteria are Gram-negative, non-spore-forming anaerobes with a wide distribution [1]. Several members of the genus *Fusobacterium* are found in the human oral cavity, where they participate in the formation of polymicrobial biofilms and cause periodontal disease [1]. These bacteria also have the ability to spread to extraoral sites to cause invasive infections,

including Lemierre's syndrome, abscesses, pleural infections and osteomyelitis and are also associated with appendicitis, pericarditis, adverse pregnancy outcomes, inflammatory bowel disease, and cancer [1, 2]. In particular, several reports have indicated that *Fusobacterium* species are associated with the development of colorectal cancer (CRC), with most studies focusing on *Fusobacterium nucleatum* [1, 3–5].

Once considered a heterogeneous species consisting of four subspecies *(animalis, nucleatum, polymorphum,* and *vincentii),* recent evidence has indicated that these subspecies are phylogenetically divergent to the point that they should be considered distinct species [6–10]. Consistently, the distinct species were shown to occupy different niches, even within the oral cavity. For instance, *F. polymorphum* dominates the dental plaque, whereas *F. animalis* (*Fa*) is abundant in odontogenic abscesses [7, 11]. Importantly, *Fusobacterium* species were also reported to differ in their disease-causing potential, especially in relation to CRC. In particular, *Fusobacterium animalis* (*Fa*) was identified as the most enriched species in *Fusobacterium*-positive CRC [12–15]. It was thus suggested that *Fa* may thrive within sites of active inflammation and may even benefit from the inflammatory environment [7].

Phylogenetically, *Fa* is most closely related to *F. paranimalis* and *F. watanabei* [10]. We recently sequenced the genomes of both species and, through a wider comparison with other *Fusobacterium* species, we found no obvious differences in the repertoire of putative virulence factors that may explain the pathogenic potential of *Fa* [10]. Recently, Younginger and coworkers identified a group of *Fa* core genes that are highly expressed intratumorally [14]. Several of these genes are involved in the stress response and the authors suggested that they may favor intracellular survival. Indeed, experiments in *F. nucleatum* sensu stricto indicated that the protein product of one of these genes, the DNA-binding protein from starved cells (Dps), increases intracellular survival in tumor cells [16]. We hypothesized that, if they promote bacterial fitness in the intracellular niche, these genes might be targets of positive selection, a process that often underlies adaptation to different environments. We thus performed an evolutionary analysis to identify selective events that occurred over different time frames, namely during the divergence of the *Fusobacterium* species and in the more recent separation of the *Fa* lineage from *F. paranimalis*. We detected several targets of selection, some of which have been identified as cancer-associated in other bacterial species.

Methods

Genome sequences, PPanGGolin analysis, and phylogenetic tree generation

Bacterial genome sequences were retrieved from the National Center for Biotechnology Information (NCBI) database. The list of genomes analyzed in this study and their accession ID is available as Table S1. Genome sequences and their annotations were used as input to run the PPanGGOLiN tool with default settings [17]. PPanGGOLiN generates pangenomes by using a graphical model and clusters proteins in gene families. We then identified which gene families included 32 *Fa* core genes that are highly expressed in CRC, as defined by [14], and we checked the presence/absence of these genes in the genome of the 161 Fusobacteria strains (Table S2).

The phylogenetic tree was constructed using 120 core genes identified by the Genome Taxonomy Database [18] as previously described [9], using IQ-TREE v1.6.12 [19]. The substitution model (GTR + F + I + G4) was selected by the ModelFinder tool implemented in IQ-TREE.

Analysis of selective patterns across the fusobacterial phylogeny

Gene alignments were generated using the GUIDANCE2 suite [20], setting sequence type as codons and using MAFFT [21] as an aligner. GUIDANCE2 also allows the filtering of unreliably aligned positions; we removed codons with a score lower than 0.90 [22].

To account for recombination, we applied the 3SEQ software (v.1.7) [23], which tests all sequence triplets in a given alignment, scanning for mosaic recombination signals. The result is the identification of mosaic regions in which one of the three sequences is the recombinant (child) of the other two (parental). A total of 12 genes were found to be recombinant (Table S3) and the mosaic regions were masked before running the positive selection analyses.

The single-likelihood ancestor counting (SLAC) method was applied to calculate the average nonsynonymous substitution (dN)/synonymous (dS) substitution rate ratio [24].

Sites that show statistical evidence of purifying selection were identified using FUBAR (Fast Unconstrained Bayesian AppRoximation) [25].

To detect the action of positive selection, a *codeml* analysis, from the PAML (Phylogenetic Analysis by Maximum Likelihood) package, was run. Specifically, site model M8 (positive selection model) that allows a class of sites to evolve with $dN/dS > 1$ was compared to a model (M7, neutral model) that does not allow $dN/dS > 1$. To assess statistical significance, twice the difference of the likelihood (ΔlnL) for the models was compared to a $\chi 2$

distribution (2 degrees of freedom) [26–28]. Genes that were significant after FDR correction for all genes analyzed (p value < 0.05) were also tested by comparing the M8 model with another more conservative neutral model (M8a), which allows dN/dS ≤ 1. In this case, twice the difference of the likelihood for the models was compared to a $\chi 2$ distribution with 1 degree of freedom (p value < 0.05) [26, 27]. In order to identify specific sites subject to positive selection, we applied four different methods: (1) the Bayes Empirical Bayes (BEB) analysis (with a posterior probability cutoff of 0.90), which calculates the posterior probability that each codon is from the positive selection site class (under M8 model) [29, 30]; (2) FUBAR (with a posterior probability cutoff of 0.90), an approximate hierarchical Bayesian method that generates an unconstrained distribution of selection parameters to estimate the posterior probability of positive diversifying selection at each site in a given alignment [25]; (3) Mixed Effects Model of Evolution (MEME) (with a p value cutoff < 0.1), which allows the distribution of dN/dS to vary from site to site and from branch to branch at a site [31]; (4) Fixed Effects Likelihood (FEL) (with a p value cutoff < 0.1), a maximum-likelihood (ML) approach to infer dN/dS on a per-site basis, assuming that the selection pressure for each site is constant along the entire phylogeny [24]. Again, to be conservative and to limit false positives, only sites detected using at least two methods were considered as positive selection targets.

FUBAR, MEME, FEL, and SLAC analyses were run locally through the HyPhy suite V2.5.29 [32, 33].

Analysis of selective events on the *Fa* lineage

Selective events that accompanied the evolution of *Fa* were investigated with gammaMap, which uses intra-species variation and inter-species diversity to estimate the distribution of selection coefficient (γ) [34]. The sequences of the 32 genes from fifty-five available *Fa* genomes, as well as their corresponding reconstructed outgroup sequences, were analyzed. In order to reconstruct the nucleotide sequences of the last common ancestor (LCA), we applied the GRASP (Graphical representation of ancestral sequence predictions) tool, which reconstructs the ancestral sequences at internal nodes of a phylogeny [35]. We used all *Fa* sequences plus the sequences of *F. watanabei, F. paranimalis* and *F. vincentii*. We retrieved the reconstructed sequence from the internal node that separates *F. paranimalis* from all the *Fa* sequences.

gammaMap requires configuring prior distributions for some of the parameters. Thus, we selected weakly informative distributions, meaning improper log-uniform distributions for the transition/transversion ratio (k), branch length (T), and for the neutral mutation rate per site (θ) parameter. A uniform distribution was chosen for the probability of adjacent codons to share the same selection coefficient (p). The frequency distribution of non-stop codons was calculated by merging all the 32 genes. To check the effect of prior selection, a second run was performed by using log normal distributions for k, T, and θ. Runs of 100,000 iterations were performed with a thinning interval of ten iterations and a burn-in of 10,000.

gammaMap categorizes selection coefficient into twelve predefined classes: strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (−1), moderately deleterious (−5, −10), strongly deleterious (−50, −100), and inviable (−500) [34]. Specifically, the program assigns to each codon a posterior probability for each selection coefficient. Because it is often difficult to infer the relative frequency of similar selection coefficients, individual codons are rarely assigned to one selection class with high reliability (i.e. with high posterior probability). This issue can be overcome by grouping coefficients into larger classes. For instance, for the analysis of *Fa* sequences, we were interested in identifying changes at codons that conferred a selective advantage. We thus called sites as positively selected if they showed a posterior probability > 0.75 of γ ≥ 1.

Protein models, disorder prediction and ELM analysis

We used AlphaFold3 [36], through the AlphaFold Server (https://alphafoldserver.com/), to model the structure of human USP7 in complex to the linear motif USP7_MACH of *Fa* DnaK, and *Fa* Prp in complex with the N-terminal extension of *Fa* L27 (LFNIQLFAHKK). ipTM + pTM scores ≥ 0.9 were used as a confidence metric, as previously suggested [37]. The model of USP7 in complex with the DnaK ELM was superposed to the solved structure of human USP7 in complex with the same motif found in human p53 (PDB IDs: 2foj) and MDM2 (PDB ID: 2fop). For Prp, the model was superposed to that of *S. aureus* in complex with a product peptide (PDB ID: 7kld).

Homology modeling was performed through the SWISS-MODEL server [38]. Template files were derived from the Protein Data Bank (PDB) (7vmx.1.B, 7vmx.1.A, 6riq.1.C) and used for the modelling of Ef-Tu, Ef-Ts, and MinD, respectively. For *Fa* ClpX we used the Alphafold DB model of CLPX_FUSNN (gene: *clpX*, organism: *Fusobacterium nucleatum subsp nucleatum*) as template, in order to also model the N-terminus. GMQE and QMEANDisCo global scores were used for quality estimation.

3D structures were rendered using PyMOL (The PyMOL Molecular Graphics System, Version 1.8.4.0; Schrödinger, LLC).

Positively selected sites in *Fa* Prp were analyzed for their effects on binding affinity to L27 using the MutaBind2 tool [39], by comparing amino acid states in *Fa* Prp with the ones in *F. paranimalis*. The effect on binding was evaluated in terms of $\Delta\Delta G_{bind}$ values.

Structural disorder was inferred using the Metapredict tool [40, 41], which applies a deep-learning algorithm based on a consensus score calculated from eight different disorder predictors. Metapredict V2 was run using default parameters.

ELMs were identified using the scan tool of the Eukaryotic Linear Motif resource [42], restricting the search to *Homo sapiens* and using the default motif probability cutoff of 100.

Results and discussion

Positively selected genes encode potential mediators of *Fusobacterium* niche adaptation

The study gene set comprised 32 genes that were found to be highly expressed in *Fa* positive tumors (as defined in a previous study with the exclusion of three genes encoding hypothetical proteins) [14] (Table S3).

We first sought to define the distribution of these genes across *Fa* strains and in Fusobacteria that are closely related to *Fa*. To this aim, we used PPanGGOLiN [17] to query the genomes of 161 Fusobacteria (Fig. 1A, Table S1 and S2). We found that all genes are present in most genomes. The most notable exception was *mliC* that seems to be absent in the *F. periodonticum/pseudoperiodonticum* lineage.

The widespread distribution of these genes across Fusobacteria allows the investigation of their evolutionary patterns. We thus retrieved sequence information for the reference genomes of the *Fusobacteria* species most closely related to *Fa* (as reported in Fig. 1A). Only in a minority of cases some orthologous genes were not present in all reference genomes (see Table S3). Using alignments of the genes and after accounting for recombination (see Methods), we first calculated the average non-synonymous substitution/synonymous substitution rate ratio (dN/dS) using the single-likelihood ancestor counting (SLAC) method [24]. The dN/dS metric is commonly used to measure the selective pressure acting on coding sequences: $dN/dS < 1$ indicates purifying selection, dN/dS around 1 is indicative of neutrality, and $dN/dS > 1$ reflects positive diversifying selection [43]. For all the studied genes, dN/dS was much lower than 1, indicating a major role for purifying selection in shaping their genetic diversity (Fig. 1B). To gain more detailed insight into the strength of selection, we used FUBAR (Fast Unconstrained Bayesian AppRoximation) to calculate the fraction of sites that show statistical evidence of purifying selection [25]. We obtained variable fractions among genes, ranging from ~ 10% to more than 80% (Fig. 1B). As expected, this measure showed an inverse relationship with dN/dS and confirmed that most of these genes are evolutionarily constrained, possibly as a result of their often-essential functions.

A major effect of purifying selection is not incompatible with positive selection acting on specific sites or domains [43]. To test for evidence of positive selection, we applied the likelihood ratio tests (LRTs) implemented in the *codeml* program [27]. Specifically, we compared models that allow (NSsite model M8, positive selection model) or disallow (NSsite models M7, null model) a class of codons to evolve with $dN/dS > 1$. After false discovery rate (FDR) correction for multiple tests, nine genes showed a significant LRT (p value < 0.05), with a better fit of model M8 vs M7 (Table S3). These genes were further tested with a more conservative M8 vs M8a model comparison. Five of them retained a significant LRT (p value < 0.05) and were considered targets of positive selection (Table 1, genes marked in red in Fig. 1B). *MliC*, which was the gene with the highest dN/dS, was not among them. Together with the comparatively lower fraction of sites targeted by purifying selection, this suggests that this gene experiences a weaker constraint compared to the others we analyzed (Fig. 1B).

(See figure on next page.)

Fig. 1 Gene representation and evolutionary patterns in the Fusobacteria phylogeny. **A** A phylogenetic tree of fusobacterial core genomes is shown, color-coded by species. A gene presence/absence matrix for the 32 studied genes, as calculated by PpanGGOLiN, is also reported. **B** dN/dS values (as per SLAC) analysis are plotted against the fraction of sites showing significant evidence of purifying selection. Genes targeted by positive selection are in red. **C** Ribbon representation of the 3D model of the Ef-Tu/Ef-Ts complex based on the crystal structure of the *M. tuberculosis* dimer (PDB: 7VMX) [93]. Sites targeted by positive selection across the Fusobacterial phylogeny are in yellow. Beneficial changes detected by gammaMap are in red. The coiled-coil region is circled **D** Alignments of representative fusobacterial MinD sequences (left) and ribbon representation of the 3D model of minD based on the structure of the MinCD copolymeric filament from *Pseudomonas aeruginosa* (6riq.1.C) [94] (right). The positively selected sites are marked in yellow. Regions involved in homodimerization and in interaction with both MinE and MinC proteins are indicated. **E** Ribbon representation of the 3D model of the *Fa* ClpX monomer (light teal) imposed onto the structure of the hexameric ClpX unfoldase ring from *Neisseria meningitidis* (PDB:6VFS, grey). The ClpP protease contact region is depicted. Serine 62, which was found as positively selected by both methods is marked in orange. In all panels, amino acid numbering refers to the *Fa* reference strain (KCOM 1325) genome (NZ_CP012715.1)

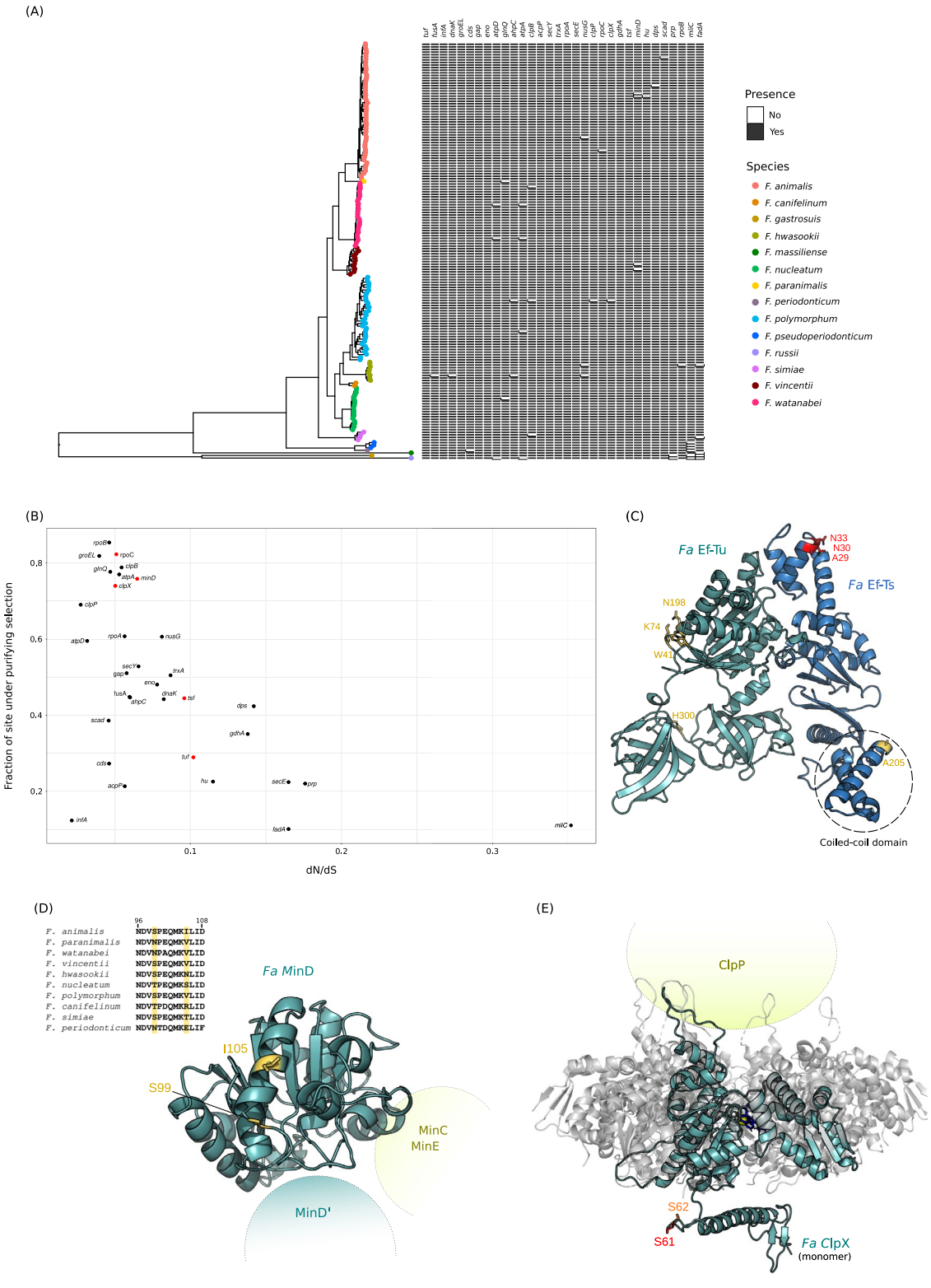

Fig. 1 (See legend on previous page.)

Table 1 Likelihood ratio test statistics for models of variable selective pressure among sites

Protein product	Gene	M8 vs M8a		Positively selected sites[c]
		−2ΔlnL[a]	p value[b]	
Translation elongation factor Tu	*tuf*	7.71	5.49×10^{-3}	41, 74, 198, 300
DNA-directed RNApolymerase subunit beta'	*rpoC*	5.90	1.15×10^{-2}	507
ATP-dependent Clp protease ATP-binding subunit clpX	*clpX*	6.42	1.13×10^{-2}	62
Translation elongation factor Ts	*tsf*	5.77	1.63×10^{-2}	205
Septum site-determining protein minD	*minD*	4.92	2.48×10^{-3}	99, 105

[a] Twice the difference of likelihood for the two models compared

[b] *p* value of rejecting the neutral models (M8a) in favor of the positive selection model (M8)

[c] positively selected sites detected by at least two methods among BEB, FEL, FUBAR, and MEME. Amino acid numbering refers to the *Fa* reference strain (KCOM 1325) genome

The positively selected genes have different functions. Inspection of the Virulence factor database (VFDB, https://www.mgc.ac.cn/VFs) [44] indicated that the protein products of two of these genes (elongation factor Tu, Ef-Tu, encoded by *tuf,* and septum site-determining protein MinD) are known virulence factors, whereas amino acid replacements in Ef-Tu and the RNA polymerase subunit beta'(encoded by *rpoC*) are associated with antibiotic resistance according to the Comprehensive Antibiotic Resistance Database (CARD Comprehensive Antibiotic Resistance Database, https://card.mcmaster.ca/) [45].

To identify specific codons subject to positive selection, we applied four methods: the Bayes Empirical Bayes (BEB) analysis from model M8, FUBAR, FEL (fixed effects likelihood) and MEME (Mixed Effects Model of Evolution) [24, 25, 30, 31]. To be conservative, only sites detected using at least two methods were considered targets of positive selection (Table S3). We detected few selected sites, indicating that positive selection is highly localized and indeed it occurs in a background of very strong constraint for some of these genes (e.g., *rpoC*, *clpX*, and *minD*) (Table 1; Fig. 1B).

We next used structural modelling to map the positively selected sites onto the 3D protein structures, and searched the literature for possible functional data. Ef-Tu and Ef-Ts function as translation elongation factors catalyzing the binding of aminoacyl-tRNA to the A-site of the ribosome. However, in many bacterial species, Ef-Tu has evolved moonlighting functions and the protein can either be exposed at the cell surface or secreted. It contributes to adhesion and invasion of host cells, modulation of immune responses, and binding of host secreted molecules [46]. For instance, in streptococci, secreted Ef-Tu promotes cell-adhesion, biofilm development, and periodontitis onset, whereas the surface exposed Ef-Tu of *Lactobacillus johnsonii* mediates attachment to intestinal epithelial cells and can induce a proinflammatory responses [47, 48]. Mapping of the positively selected sites on the 3D structure of *Fa* Ef-Tu indicated that all of them are exposed at the surface and may thus mediate attachment or interaction with host factors. Indeed, one of them (H300) is located in the barrel-like adhesion domain (Fig. 1C) [48]. Conversely, Ef-Ts has no known moonlighting functions. In this protein, the single positively selected site is in the coiled-coil domain (Fig. 1C). Interestingly, mutations in this region have been shown to confer resistance against contact-dependent antibacterial tRNase toxins produced by enterohemorrhagic *Escherichia coli* [49]. Thus, the positively selected site might represent a signature of inter-bacterial conflict.

MinD, with its partners MinC and MinE, are required for the correct placement of the division site and act to spatially regulate cell division [50]. In *Neisseria gonorrhoeae* and *E. coli*, *min* mutants have altered cell shape and size, as well as decreased virulence [51–54]. The positively selected sites in *minD* are not predicted to be located at the homodimer interface, nor in the MinC/MinE binding region (Fig. 1D). As in the case of Ef-Tu the sites are surface-exposed and may mediate interaction with other proteins or components. In this respect, it is interesting to note that *F. watanabei* and *Fa*, which are closely related but carry different residues at the positively selected MinD sites (Fig. 1D), were shown to differ in cell length and width [10, 15]. This is relevant because bacterial cell shape and size have been shown to modulate colonization of distinct host niches, the susceptibility to host defenses, and, for many pathogens, disease progression [55]. Experimental studies will however be necessary to determine whether the positively selected sites have a role in modulating *Fa* cell shape and pathogenicity.

We also detected one positively selected site in RpoC. This variant was not associated with resistance as determined by consulting CARD [45]. However, multiple evidence suggests that changes within housekeeping bacterial genes can have antibiotic-independent adaptive effects [56]. In *E. coli*, mutations in *rpoC* have been

demonstrated to arise in response to various selective pressures, including prolonged resource exhaustion, growth in minimal media or at high temperatures or under acidic conditions, as well as changes in nutrient sources [57]. Such mutations are often antagonistically pleiotropic—for instance they are adaptive under resource exhaustion but reduce exponential growth rates in full media [58]. Thus, an interesting possibility is that the positively selected site in *rpoC* contributed to the adaptation of Fusobacteria to different niches within the host [9]. Whether it also evolved to facilitate intracellular survival remains to be evaluated.

Finally, we detected a positively selected site in ClpX, which is located on a flexible loop that joins the zinc coordinating and ATP binding domains of the protease (Fig. 1E) (see below).

The chaperone DnaK and the ribosomal-processing protease Prp are major targets of selection on the *Fa* lineage

We next aimed to assess whether adaptive changes emerged specifically on the *Fa* lineage. To this aim, we retrieved 55 complete or almost-complete *Fa* genomes and extracted sequence information for the 32 genes under study (Tables S1 and S2). As the outgroup, we reconstructed the sequences of the ancestor of *Fa* and *F. paranimalis* (see methods). We next used the gammaMap program [34], which jointly uses intra-species variation and inter-species diversity, to estimate the distribution of fitness effects (i.e., selection coefficients, γ). In practical terms, γ values can be considered a measure of the fitness consequences of new non-synonymous mutations. The method categorizes selection coefficients into twelve predefined classes ranging from −500 (inviable) to 100 (strongly beneficial). As expected, and in line with the results above, we observed a preponderance of codons evolving under strong purifying selection ($-500 \leq \gamma \leq -10$). The median γ value for most of the genes was lower than − 50, with only a few exceptions, and the highest median ($\gamma = -1$) was observed for *prp* (Fig. 2A). Thus, these genes also experience intense constraint in the circulating *Fa* population.

To determine whether any amino acid changes conferred a fitness advantage to *Fa*, we estimated codon-wise posterior probabilities for each selection coefficient. We called a codon as positively selected if its cumulative posterior probability of $\gamma \geq 1$ was > 0.75. A total of thirty-three sites in seven genes were found to be positively selected (Fig. 2B). For most such sites, the selected derived allele is fixed or at high frequency in the *Fa* population (Fig. 2B). The largest numbers of positively selected sites were detected in Prp and DnaK (Fig. 2B). As above, we integrated structural information and literature searches, but also motif analysis, to mine possible effects of the positively selected sites.

Analysis of beneficial changes in ClpX indicated that one of the sites corresponds to the one that was targeted by positive selection in the extended Fusobacterial phylogeny and the other is flanking (Fig. 1E). Thus, the same loop in ClpX has been a target of selection over different time-frames—i.e. during the emergence of the Fusobacterial species and more recently, as the *Fa* population diverged from the common ancestor with closely related *Fusobacteria*. ClpX is the ATP-dependent component of the Clp protease, which determines substrate specificity [59]. In many bacteria, the protease controls virulence, stress tolerance, and biofilm formation [59, 60]. More recently, ClpX was also reported to have a role in phage biology, as it directly interacts with the phage repressor CI and performs an essential role in prophage induction by abolishing the ability of CI to repress lytic phage genes [61]. We therefore speculate that the positively selected sites in ClpX modulate the specificity of the protease towards endogenous or phage proteins. Unlike the observations for ClpX, the beneficial changes in Ef-Ts do not map to the coiled-coil region, nor to the surface involved in the interaction with Ef-Tu (Fig. 1C).

Next, we moved to the analysis of selected sites in the two major targets, Prp and DnaK. Prp is essential for the assembly and maturation of the ribosome, as it removes an N-terminal extension from a precursor of ribosomal protein L27. The protein acts as a dimer. We thus modeled the structure of *Fa* Prp in complex with the N-terminal extension of *Fa* L27 using AlphaFold3 [36] (Fig. 2C). The generated complex was then superimposed to the crystal structure of the orthologous protein from *Staphylococcus aureus,* which was crystallized in complex with a peptide substrate (Fig. 2C) [62], showing a perfect overlap between complexes. Mapping of the positively selected sites indicated that three of them project into the binding pocket with a distance of less than 3 Å from the L27 fragment. An additional selected site is at the homodimer interface (Fig. 2C). To evaluate the effect of the positive sites on binding efficiency, we mutated in silico these sites by recapitulating the amino acid states found in *F. paranimalis.* We then estimated changes in binding affinity (Table S4): as expected the *F. paranimalis* mutations were not predicted to be deleterious, but they diminished binding affinity ($\Delta\Delta G_{bind} > 0$) (Table S4). Overall, this analysis suggests that *Fa* Prp may bind the L27 N-terminal tail more efficiently than the protein encoded by *F. paranimalis*. Whether these changes also modulate the cleavage efficiency of L27 remains to be evaluated. Furthermore, an interesting possibility is that the positively selected sites modulate the specificity of the protease for other substrates. Recent evidence in

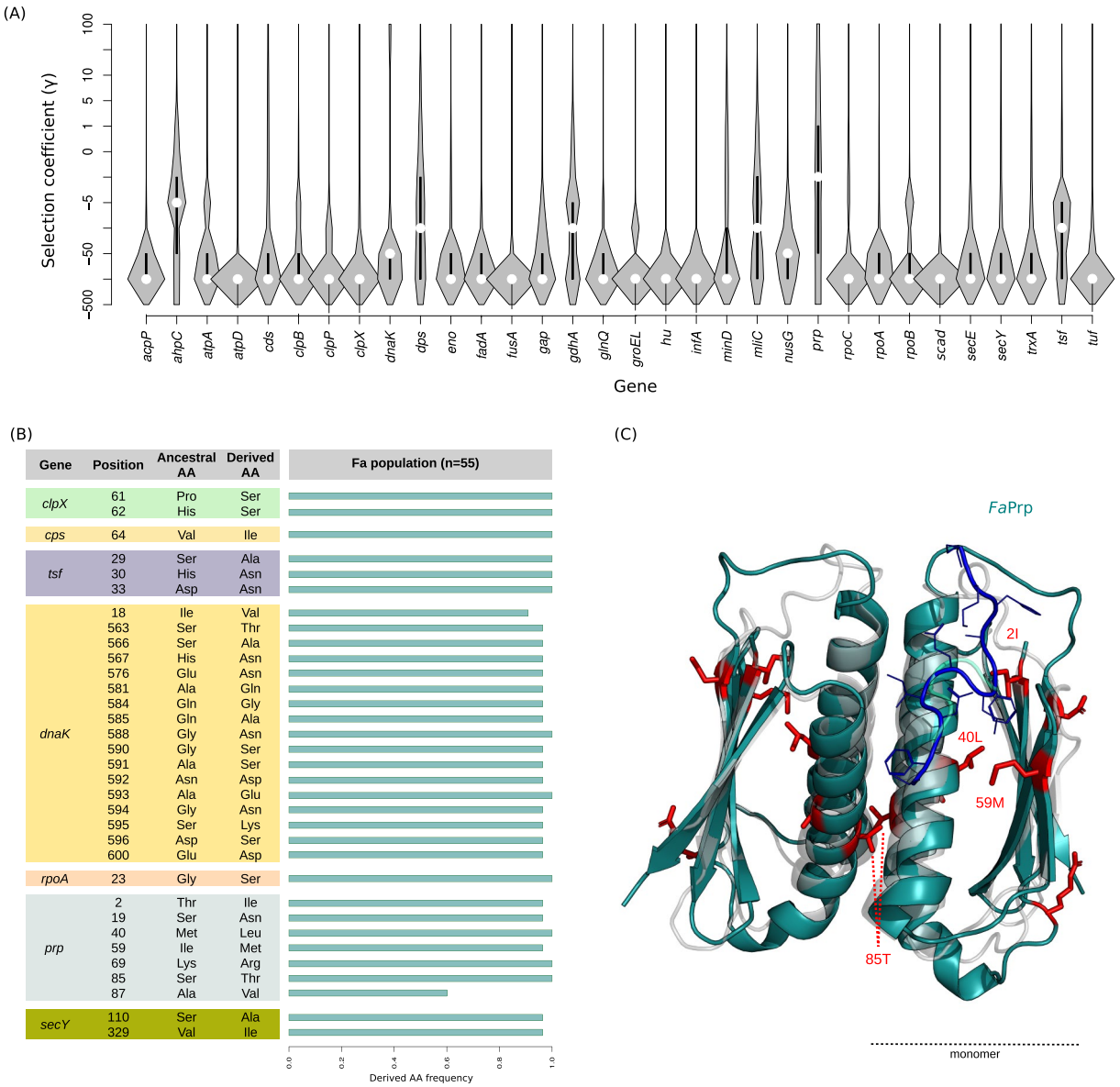

Fig. 2 Analysis of selective events on *Fa* lineage. **A** Violin plot of selection coefficients for the 32 genes (median, white dot; interquartile range, black bar). Selection coefficients (γ) are classified as strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (− 1), moderately deleterious (− 5, − 10), strongly deleterious (− 50, − 100), and inviable (− 500). **B** Beneficial changes that occurred on the *Fa* lineage. The plot shows all the identified positively selected sites (posterior probability > 0.75 of γ ≥ 1), their position (relative to the *Fa* reference strain KCOM 1325), and their frequency in the *Fa* population. The ancestral amino acid refers to the amino acid coded by the reconstructed LCA (see Materials and Methods section for details). **C** Ribbon representation of the 3D model of *Fa* Prp (light teal) in complex with the N-terminal segment of *Fa* L27 (blue) and superposed to *S. aureus* Prp (gray) with its product peptide (cyan) (PDB: 7KLD) [62]. Positively selected sites are in red. The three sites less than 3 Å apart from the substrate peptide and one at the homodimer interface are labelled

Streptococcus mutans indicates that a Prp homolog can cleave cellular enzymes, eventually modulating carbohydrate metabolism, biofilm formation, and interaction with other microorganisms (e.g., *Candida albicans*) [63].

The most interesting target of selection identified in this study was DnaK, not only because it accumulated the largest number of beneficial changes, but also for its role in tumor biology. Indeed, the DnaK chaperone encoded by *Mycoplasma fermentans* has been demonstrated to promote malignant transformation in vivo and in vitro [64–66], and both the *M. fermentans* and *F. nucleatum* DnaK proteins reduce the effectiveness of anticancer drugs such as cisplatin and 5-Fluorouracil [67]. Experiments with *M. fermentans* showed that DnaK can be

internalized by uninfected human cells and that the protein localizes to several cellular compartments, including the cytoplasm, perinuclear membrane, and nucleus [66]. The DnaK protein of *F. nucleatum* was found to be secreted [68], suggesting that, irrespective of whether intracellular infection occurs, the protein may gain access to human cells. In the *M. fermentans* model, DnaK was shown to bind several host proteins, including human USP10 (ubiquitin-specific protease 10), thus reducing p53 stability and anti-cancer functions [66, 69].

We identified 17 beneficial changes in *Fa* DnaK. With the exclusion of 18 V, all of them are located in the C-terminal region (Fig. 3A). Because the terminal tail of some DnaK proteins was shown to be intrinsically disordered (i.e., to lack a fixed 3D structure) [70, 71], we used Metapredict V2, a deep-learning-based approach that combines different predictors to generate consensus disorder scores, to test whether this was also the case for the fusobacterial proteins. Results indicated that the ~ 30 C-terminal residues of DnaK from *Fa*, *F. paranimalis, F. watanabei, F. vincentii* and the LCA are unstructured (Fig. 3B). In eukaryotes, intrinsically disordered protein regions can accommodate eukaryotic linear motifs (ELMs), short peptide modules that are important for protein regulation, often mediating protein–protein interactions or acting as sites for posttranslational modifications [72–74]. ELMs are often mimicked by viral and bacterial pathogens to hijack host cell regulatory networks [75, 76]. Indeed, these elements of host molecular mimicry have been studied on a large scale in *F. nucleatum* using a computational strategy. This enabled the identification of a network of protein interactions between the putative *F. nucleatum* secretome and human proteins [77]. We thus used an ELM prediction tool to investigate the presence of linear motifs in the disordered C-terminal tails of the fusobacterial DnaK proteins [42]. No ELM was found in the proteins encoded by *F. paranimalis, F. watanabei*, and *F. vincentii*, nor in

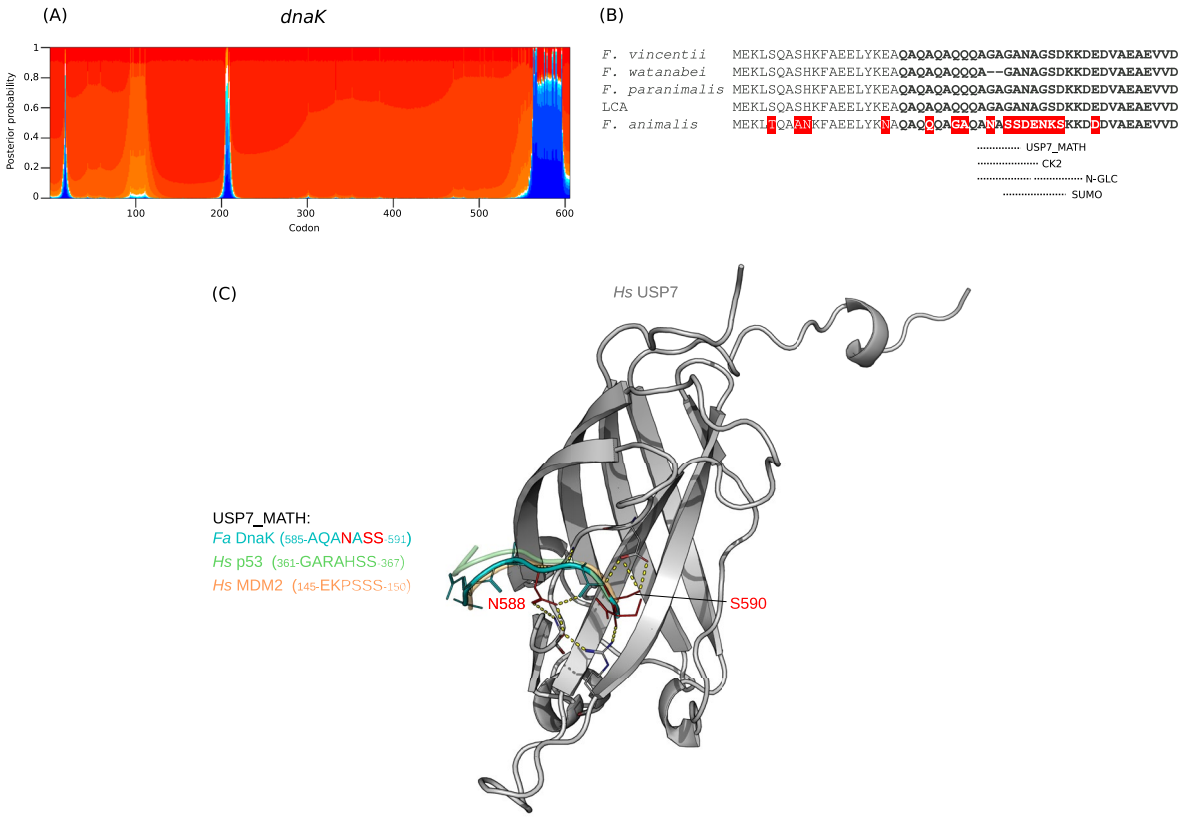

Fig. 3 Positive selection in DnaK. **A** The posterior probability of selection coefficients for non-synonymous mutations along the *dnaK* coding region (Log uniform model). The height of the colored bars represents the posterior probability of the corresponding selection coefficient. Colors closer to red represent increasingly deleterious variants, white indicates neutral variants, and colors closer to blue represent increasingly beneficial variants. **B** Alignment of the C-terminal portion of DnaK, where most positively selected sites (red) are located. The regions that are predicted to be intrinsically disordered are in bold. The location of the detected ELMs is shown. **C** Cartoon representation of the 3D model of human USP7 (gray) in complex with USP7_MATH motifs of *Fa* DnaK (cyan). The corresponding linear motifs of human p53 and MDM2 are also shown by superimposition of the solved complexes from PDB (IDs: 2FOJ and 2FOP, respectively). *Fa* DnaK positively selected sites are in red, residues interacting with USP7 are labeled, hydrogens bonds are yellow dotted

Table 2 Details of the ELMs identified in the *Fa* DnaK disordered region

ELM Name	Instances	Positions	ELM Description	Probability
DOC_USP7_MATH_1	ANASS	587–591	The USP7 MATH domain binding motif variant based on the MDM2 and p53 interactions	0.01239
MOD_CK2_1	ANASSDE	587–593	Casein kinase 2 (CK2) phosphorylation site	0.01457
MOD_N-GLC_1	ANASSD	587–592	Generic motif for N-glycosylation	0.005018
	ENKSKK	593–598		
MOD_SUMO_rev_2	SSDENKS	590–596	Inverted version of SUMOylation motif recognized for modification by SUMO-1	0.0128
	SDENKS	591–596		
	DENKS	592–596		

the LCA sequence. Conversely, five ELMs were detected in the *Fa* DnaK sequence (Fig. 3B; Table 2). Two of them are N-glycosylation sites, which are unlikely to be functional because, in eukaryotic cells, the N-glycan precursor is biosynthesized in the endoplasmic reticulum. Another one is a casein kinase 2 (CK2) phosphorylation motif (Fig. 3B; Table 2). CK2 is a constitutively active kinase frequently over-expressed in cancer cells, where it has multiple roles, including proliferation, cell cycle control, and protection from apoptosis [78]. Notably, in CRC, CK2 was suggested to regulate the epithelial-mesenchymal transition (EMT) and *Fa* is particularly associated with tumors of the mesenchymal subtype [14, 79]. Although these observations do not necessarily imply causation, they indicate that the high expression of CK2 in malignant cells may indeed result in the phosphorylation of the disordered tail of DnaK, a modification that is known to modulate the functional properties of unstructured regions [80].

The two final ELMs we detected are involved in post-translational modifications, namely ubiquitination and SUMOylation (Fig. 3B; Table 2). The latter is a dynamic, reversible process involving the covalent modification of specific lysine residues on target proteins with SUMO (small ubiquitin-related modifier). SUMO conjugation is an important regulatory mechanism for protein stability, subcellular localization, and protein–protein interactions [81]. Interestingly, SUMOylation (and ubiquitination) have been shown to regulate the nuclear import of specific proteins [82]. For instance, SUMO modification promotes the nuclear import of polo-like kinase and suppresses its ubiquitin-mediated proteasomal degradation [83]. Likewise, caspase-8 SUMOylation is associated with nuclear localization [84]. It is thus possible that SUMOylation of DnaK allows its access to the nuclear compartment. It is also worth noting that ubiquitin-specific protease 7 (USP7) is predicted to target a motif flanking the SUMOylation site. USP7 is a key regulator of the p53-MDM2 pathway and, as in the case of CK2, it is over-expressed in most cancers [85]. Thus, the pharmacological inhibition of USP7 is regarded as a promising therapeutic strategy [86]. Notably, USP7 functions as a SUMO deubiquitinase: by acting on SUMOylated proteins, USP7 counteracts their ubiquitination [87]. It is thus tempting to speculate that SUMOylation facilitates the nuclear import of *Fa* DnaK where USP7 promotes its stability. Given its potential biological relevance, we used AlphaFold3 to model the interaction between human USP7 and the cognate motif in DnaK. This confirmed that the *Fa* DnaK ELM peptide binds USP7 in the same region as the p53 and MDM2 motifs. We also observed that two sites found as positively selected (N588 and S590) contribute to this interaction (Fig. 3C).

Clearly, these analyses will require experimental validation. We acknowledge that the lack of in vitro/in vivo data to support our hypotheses represents a limitation of this study. Another limitation is the relatively small number of genes we analyzed. Such genes were selected due to their high expression in *Fa* positive tumors and their possible role in promoting intracellular survival [14]. It is however possible that additional genes modulate the intratumor survival of *Fa* and its oncogenic potential.

Conclusions

In summary, our data show that the genes we analyzed, which are highly expressed in *Fa* positive tumors, are present in most Fusobacteria genomes and evolved under purifying selection across different time frames. This clearly underscores the functional relevance and the housekeeping functions of many of them. However, localized signatures of positive selection were detectable. Within the host, distinct *Fusobacterium* species are adapted to different niches that differ in many respects, including oxygen tension, nutrient availability, exposure to the host defense system, and competition with other microorganisms. Previous analyses in other bacteria have indicated that, whereas horizontal gene transfer certainly has strong adaptive potential, changes in core genes also contribute to the colonization and

exploitation of novel niches or environments [57, 58, 88–91]. Thus, the selection signatures we identified in the fusobacterial phylogeny may contribute to some aspects of niche adaptation. We also searched for and detected beneficial changes that occurred specifically in the *Fa* lineage, suggesting that such variants promote intracellular growth or, more generally, adaptation to the tumor microenvironment. Whether niche adaptation in itself is necessarily important for cancer development is presently impossible to determine. One interesting possibility is that niche adaptation and cancer development are linked processes, as the bacterium benefits from stimulating the expansion of its preferred habitat (the tumor). In this respect, the identification of DnaK as the protein targeted by the strongest selection is extremely interesting and, based on data from other bacteria [64–67, 69, 92], suggests that tumor formation and progression confer a fitness advantage to *Fa*. Clearly these hypotheses will need experimental validation. In this respect, our identification of specific proteins and amino acid changes that are e xpected to underlie phenotypic diversity is relevant to inform future analyses of *Fa* oncogenic potential.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13099-025-00740-1.

Additional file 1

Additional file 2

Additional file 3

Additional file 4

Author contributions

MS and DF conceived the study and designed the experiment; DF, RC, CM, and AM performed the experiments and generated the data, with assistance from MS; DF and AS conducted the data analysis with support from MS and ØK; AS and ØK provided data; MS and DF drafted the manuscript with input from all authors; MS and DF organized the data and finalized the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Italian Ministry of Health ("Ricerca Corrente").

Availability of data and materials

All accession IDs are listed in Tables S1 and S2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 June 2025 Accepted: 8 August 2025
Published online: 11 August 2025

References

1. Brennan CA, Garrett WS. *Fusobacterium nucleatum*—symbiont, opportunist and oncobacterium. Nat Rev Microbiol. 2019;17:156–66.
2. Dyrhovden R, Eagan TM, Fløtten Ø, Siljan W, Leegaard TM, Bø B, et al. Pleural empyema caused by *Streptococcus intermedius* and *Fusobacterium nucleatum*: a distinct entity of pleural infections. Clin Infect Dis. 2023;77:1361–71.
3. Barot SV, Sangwan N, Nair KG, Schmit SL, Xiang S, Kamath S, et al. Distinct intratumoral microbiome of young-onset and average-onset colorectal cancer. EBioMedicine. 2024;100: 104980.
4. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of *Fusobacterium* persistence and antibiotic response in colorectal cancer. Science. 2017;358:1443–8.
5. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. Genome Res. 2012;22:292–8.
6. Kook J-K, Park S-N, Lim YK, Cho E, Jo E, Roh H, et al. Genome-based reclassification of *Fusobacterium nucleatum* subspecies at the species level. Curr Microbiol. 2017;74:1137–47.
7. Krieger M, AbdelRahman YM, Choi D, Palmer EA, Yoo A, McGuire S, et al. Stratification of *Fusobacterium nucleatum* by local health status in the oral cavity defines its subspecies disease association. Cell Host Microbe. 2024;32:479-488.e4.
8. Manson McGuire A, Cochrane K, Griggs AD, Haas BJ, Abeel T, Zeng Q, et al. Evolution of invasion in a diverse set of *Fusobacterium* species. MBio. 2014;5:e01864.
9. Molteni C, Forni D, Cagliani R, Sironi M. Comparative genomics reveal a novel phylotaxonomic order in the genus *Fusobacterium*. Commun Biol. 2024;7:1102.
10. Sivertsen A, Forni D, Molteni C, Bivand J, Dimmen G, Sironi M, et al. Reassessing taxonomy and virulence in the *Fusobacterium nucleatum* group—Rebuttal of *Fusobacterium animalis* clades "*Fna* C1" and "*Fna* C2", genome announcement for *Fusobacterium watanabei* and description of *Fusobacterium paranimalis* sp. nov. MBio. 2025. https://doi.org/10.1101/2025.03.20.644344.
11. Connolly JP, Kelly L. The physical biogeography of *Fusobacterium nucleatum* in health and disease. Graf J, editor. Bio. 2025;16:e02989–24.
12. Borozan I, Zaidi SH, Harrison TA, Phipps AI, Zheng J, Lee S, et al. Molecular and pathology features of colorectal tumors and patient outcomes are associated with *Fusobacterium nucleatum* and its subspecies *animalis*. Cancer Epidemiol Biomarkers Prev. 2022;31:210–20.
13. Ye X, Wang R, Bhattacharya R, Boulbes DR, Fan F, Xia L, et al. *Fusobacterium nucleatum* subspecies *animalis* influences proinflammatory cytokine expression and monocyte activation in human colorectal tumors. Cancer Prev Res. 2017;10:398–409.
14. Younginger BS, Mayba O, Reeder J, Nagarkar DR, Modrusan Z, Albert ML, et al. Enrichment of oral-derived bacteria in inflamed colorectal tumors and distinct associations of *Fusobacterium* in the mesenchymal subtype. Cell Rep Med. 2023;4: 100920.
15. Zepeda-Rivera M, Minot SS, Bouzek H, Wu H, Blanco-Míguez A, Manghi P, et al. A distinct *Fusobacterium nucleatum* clade dominates the colorectal cancer niche. Nature. 2024;628:424–32.
16. Wu Y, Guo S, Chen F, Li Y, Huang Y, Liu W, et al. Fn-dps, a novel virulence factor of *Fusobacterium nucleatum*, disrupts erythrocytes and promotes metastasis in colorectal cancer. PLoS Pathog. 2023;19: e1011096.
17. Gautreau G, Bazin A, Gachet M, Planel R, Burlot L, Dubois M, et al. PPanGGOLiN: Depicting microbial diversity via a partitioned pangenome graph. Ouzounis CA, editor. PLoS Comput Biol. 2020;16:e1007732.
18. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50:D785–94.
19. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

20. Sela I, Ashkenazy H, Katoh K, Pupko T. Guidance2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 2015;43: 7.
21. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
22. Privman E, Penn O, Pupko T. Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol Biol Evol. 2012;29:1–5.
23. Lam HM, Ratmann O, Boni MF. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. Mol Biol Evol. 2018;35:247–51.
24. Kosakovsky Pond SL, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22:1208–22.
25. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Pond SLK, et al. FUBAR: a fast, unconstrained bayesian approximation for inferring selection. Mol Biol Evol. 2013;30:1196–205.
26. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics. 1997;13:555–6.
27. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91.
28. Álvarez-Carretero S, Kapli P, Yang Z. Beginner's guide on the use of PAML to detect positive selection. Crandall K, editor. Mol Biol Evol. 2023;40: msad041.
29. Anisimova M, Bielawski JP, Yang Z. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Mol Biol Evol. 2002;19:950–8.
30. Yang Z, Wong WS, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–18.
31. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8: e1002764.
32. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21:676–9.
33. Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL, Murrell B, et al. HyPhy 2.5—a customizable platform for evolutionary hypothesis testing using phylogenies. Mol Biol Evol. 2020;37:295–9.
34. Wilson DJ, Hernandez RD, Andolfatto P, Przeworski M. A population genetics-phylogenetics approach to inferring natural selection in coding sequences. PLoS Genet. 2011;7: e1002395.
35. Foley G, Mora A, Ross CM, Bottoms S, Sützl L, Lamprecht ML, et al. Engineering indel and substitution variants of diverse and ancient enzymes using Graphical Representation of Ancestral Sequence Predictions (GRASP). Elofsson A, editor. PLoS Comput Biol. 2022;18:e1010633.
36. Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630:493–500.
37. Jeppesen M, André I. Accurate prediction of protein assembly structure by combining AlphaFold and symmetrical docking. Nat Commun. 2023;14:8283.
38. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22:195–201.
39. Zhang N, Chen Y, Lu H, Zhao F, Alvarez RV, Goncearenco A, et al. MutaBind2: predicting the impacts of single and multiple mutations on protein-protein interactions. Science. 2020;23:100939.
40. Emenecker RJ, Griffith D, Holehouse AS. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. Biophys J. 2021;120:4312–9.
41. Emenecker RJ, Griffith D, Holehouse AS. Metapredict V2: An update to metapredict, a fast, accurate, and easy-to-use predictor of consensus disorder and structure. https://doi.org/10.1101/2022.06.06.494887
42. Kumar M, Michael S, Alvarado-Valverde J, Zeke A, Lazar T, Glavina J, et al. ELM—the Eukaryotic Linear Motif resource—2024 update. Nucleic Acids Res. 2024;52:D442–55.
43. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host-pathogen interactions from mammalian sequence data. Nat Rev Genet. 2015;16:224–36.
44. Liu B, Zheng D, Zhou S, Chen L, Yang J. VFDB 2022: a general classification scheme for bacterial virulence factors. Nucleic Acids Res. 2022;50:D912–7.
45. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51:D690–9.
46. Harvey KL, Jarocki VM, Charles IG, Djordjevic SP. The diverse functional roles of elongation factor Tu (EF-Tu) in microbial pathogenesis. Front Microbiol. 2019;10:2351.
47. Granato D, Bergonzelli GE, Pridmore RD, Marvin L, Rouvet M, Corthésy-Theulaz IE. Cell surface-associated elongation factor Tu mediates the attachment of *Lactobacillus johnsonii* NCC533 (La1) to human intestinal cells and mucins. Infect Immun. 2004;72:2160–9.
48. Xiao L, Pu Y, Cui Y, Chen C, Xiao Q, Wang Y, et al. Elongation factor Tu promotes the onset of periodontitis through mediating bacteria adhesion. npj Biofilms Microbiomes. 2025;11:47.
49. Jones AM, Garza-Sánchez F, So J, Hayes CS, Low DA. Activation of contact-dependent antibacterial tRNase toxins by translation elongation factors. Proc Natl Acad Sci U S A. 2017;114(10):E1951–7.
50. De Boer PA, Crossley RE, Hand AR, Rothfield LI. The MinD protein is a membrane ATPase required for the correct placement of the *Escherichia coli* division site. EMBO J. 1991;10:4371–80.
51. De Boer PAJ, Crossley RE, Rothfield LI. A division inhibitor and a topological specificity factor coded for by the minicell locus determine proper placement of the division septum in *E. coli*. Cell. 1989;56(4):641–9.
52. Parti RP, Biswas D, Helgeson S, Michael FS, Cox A, Dillon J-AR. Attenuated virulence of min operon mutants of *Neisseria gonorrhoeae* and their interactions with human urethral epithelial cells. Microbes Infect. 2011;13:545–54.
53. Parti RP, Biswas D, Wang M, Liao M, Dillon J-AR. A MinD mutant of enterohemorrhagic *E. coli* O157:H7 has reduced adherence to human epithelial cells. Microb Pathog. 2011;51:378–83.
54. Szeto J, Ramirez-Arcos S, Raymond C, Hicks LD, Kay CM, Dillon J-AR. Gonococcal MinD affects cell division in *Neisseria gonorrhoeae* and *Escherichia coli* and exhibits a novel self-interaction. J Bacteriol. 2001;183:6253–64.
55. Yang DC, Blair KM, Salama NR. Staying in shape: the impact of cell shape on bacterial survival in diverse environments. Microbiol Mol Biol Rev. 2016;80:187–203.
56. Hershberg R. Antibiotic-independent adaptive effects of antibiotic resistance mutations. Trends Genet. 2017;33:521–8.
57. Cohen Y, Hershberg R. Rapid adaptation often occurs through mutations to the most highly conserved positions of the RNA polymerase core enzyme. Ogbunu B, editor. Genome Biol Evol. 2022;14:evac105.
58. Avrani S, Bolotin E, Katz S, Hershberg R. Rapid genetic adaptation during the first four months of survival under resource exhaustion. Mol Biol Evol. 2017;34:1758–69.
59. Frees D, Gerth U, Ingmer H. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of *Staphylococcus aureus*. Int J Med Microbiol. 2014;304:142–9.
60. Aljghami ME, Barghash MM, Majaesic E, Bhandari V, Houry WA. Cellular functions of the ClpP protease impacting bacterial virulence. Front Mol Biosci. 2022;9:1054408.
61. Thabet MA, Penadés JR, Haag AF. The ClpX protease is essential for inactivating the CI master repressor and completing prophage induction in *Staphylococcus aureus*. Nat Commun. 2023;14:6599.
62. Hotinger JA, Pendergrass HA, Peterson D, Wright HT, May AE. Phage-related ribosomal protease (Prp) of *Staphylococcus aureus*. In vitro Michaelis-Menten kinetics, screening for inhibitors, and crystal structure of a covalent inhibition product complex. Biochemistry. 2022;61(13):1323–36.
63. Treerat P, De Mattos C, Burnside M, Zhang H, Zhu Y, Zou Z, et al. Ribosomal-processing cysteine protease homolog modulates *Streptococcus mutans* glucan production and interkingdom interactions. El-Naggar MY, editor. J Bacteriol. 2024;206:e00104–24.
64. Benedetti F, Cocchi F, Latinovic OS, Curreli S, Krishnan S, Munawwar A, et al. Role of *Mycoplasma* chaperone DnaK in cellular transformation. IJMS. 2020;21:1311.
65. Benedetti F, Silvestri G, Denaro F, Finesso G, Contreras-Galindo R, Munawwar A, et al. *Mycoplasma* DnaK expression increases cancer development in vivo upon DNA damage. Proc Natl Acad Sci U S A. 2024;121: e2320859121.

66. Zella D, Curreli S, Benedetti F, Krishnan S, Cocchi F, Latinovic OS, et al. *Mycoplasma* promotes malignant transformation in vivo, and its DnaK, a bacterial chaperone protein, has broad oncogenic properties. Proc Natl Acad Sci U S A. 2018;115: E12005–E12014.
67. Benedetti F, Mongodin EF, Badger JH, Munawwar A, Cellini A, Yuan W, et al. Bacterial DnaK reduces the activity of anti-cancer drugs cisplatin and 5FU. J Transl Med. 2024;22:269.
68. Skår CK, Krüger PG, Bakken V. Characterisation and subcellular localisation of the GroEL-like and DnaK-like proteins isolated from *Fusobacterium nucleatum* ATCC 10953. Anaerobe. 2003;9:305–12.
69. Curreli S, Benedetti F, Yuan W, Munawwar A, Cocchi F, Gallo RC, et al. Characterization of the interactome profiling of *Mycoplasma fermentans* DnaK in cancer cells reveals interference with key cellular pathways. Front Microbiol. 2022;13:1022704.
70. Rosendahl S, Ainelo A, Hõrak R. The disordered C-terminus of the chaperone DnaK increases the competitive fitness of *Pseudomonas putida* and facilitates the toxicity of GraT. Microorganisms. 2021;9:375.
71. Smock RG, Blackburn ME, Gierasch LM. Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity. J Biol Chem. 2011;286:31821–9.
72. Davey NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, et al. Attributes of short linear motifs. Mol Biosyst. 2012;8:268–81.
73. Fuxreiter M, Tompa P, Simon I. Local structural disorder imparts plasticity on linear motifs. Bioinformatics. 2007;23:950–6.
74. Van Roey K, Uyar B, Weatheritt RJ, Dinkel H, Seiler M, Budd A, et al. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem Rev. 2014;114:6733–78.
75. Sámano-Sánchez H, Gibson TJ. Mimicry of short linear motifs by bacterial pathogens: a drugging opportunity. Trends Biochem Sci. 2020;45:526–44.
76. Via A, Uyar B, Brun C, Zanzoni A. How pathogens use linear motifs to perturb host cell networks. Trends Biochem Sci. 2015;40:36–48.
77. Zanzoni A, Spinelli L, Braham S, Brun C. Perturbed human sub-networks by *Fusobacterium nucleatum* candidate virulence proteins. Microbiome. 2017;5:89.
78. Trembley JH, Kren BT, Afzal M, Scaria GA, Klein MA, Ahmed K. Protein kinase CK2—diverse roles in cancer cell biology and therapeutic promise. Mol Cell Biochem. 2023;478:899–926.
79. Zou J, Luo H, Zeng Q, Dong Z, Wu D, Liu L. Protein kinase CK2α is overexpressed in colorectal cancer and modulates cell proliferation and invasion via regulating EMT-related genes. J Transl Med. 2011;9:97.
80. Iakoucheva LM. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004;32:1037–49.
81. Geiss-Friedlander R, Melchior F. Concepts in sumoylation: a decade on. Nat Rev Mol Cell Biol. 2007;8:947–56.
82. Rodríguez JA. Interplay between nuclear transport and ubiquitin/SUMO modifications in the regulation of cancer-related proteins. Semin Cancer Biol. 2014;27:11–9.
83. Wen D, Wu J, Wang L, Fu Z. Sumoylation promotes nuclear import and stabilization of Polo-like kinase 1 to support its mitotic function. Cell Rep. 2017;21:2147–59.
84. Besnault-Mascard L, Leprince C, Auffredou MT, Meunier B, Bourgeade MF, Camonis J, et al. Caspase-8 sumoylation is associated with nuclear localization. Oncogene. 2005;24:3268–73.
85. Saha G, Roy S, Basu M, Ghosh MK. USP7—a crucial regulator of cancer hallmarks. Biochimica et Biophysica Acta (BBA)—Rev Cancer. 2023;1878: 188903.
86. Carreira LD, Oliveira RI, Moreira VM, Salvador JAR. Ubiquitin-specific protease 7 (USP7): an emerging drug target for cancer treatment. Expert Opin Ther Targets. 2023;27:1043–58.
87. Lecona E, Rodriguez-Acebes S, Specks J, Lopez-Contreras AJ, Ruppen I, Murga M, et al. USP7 is a SUMO deubiquitinase essential for DNA replication. Nat Struct Mol Biol. 2016;23:270–7.
88. Anisimova M, Bielawski J, Dunn K, Yang Z. Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. BMC Evol Biol. 2007;7:154.
89. Gross J, Katz S, Hershberg R. *Pseudomonas putida* dynamics of adaptation under prolonged resource exhaustion. Eyre-Walker A, editor. Genome Biol Evol 2024;16:evae117.
90. Lefébure T, Stanhope MJ. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. Genome Res. 2009;19:1224–32.
91. Zion S, Katz S, Hershberg R. Escherichia coli adaptation under prolonged resource exhaustion is characterized by extreme parallelism and frequent historical contingency. Behringer M, editor. PLoS Genet. 2024;20:e1011333.
92. Benedetti F, Silvestri G, Saadat S, Denaro F, Latinovic OS, Davis H, et al. Mycoplasma DnaK increases DNA copy number variants in vivo. Proc Natl Acad Sci USA. 2023;120: e2219897120.
93. Zhan B, Gao Y, Gao W, Li Y, Li Z, Qi Q, et al. Structural insights of the elongation factor EF-Tu complexes in protein translation of *Mycobacterium tuberculosis*. Commun Biol. 2022;5:1052.
94. Szewczak-Harris A, Wagstaff J, Löwe J. Cryo-EM structure of the MinCD copolymeric filament from *Pseudomonas aeruginosa* at 3.1 Å resolution. FEBS Lett. 2019;593:1915–26.

Publisher's Note