

# Evolution of Invasion in a Diverse Set of *Fusobacterium* Species

Abigail Manson McGuire,<sup>a</sup> Kyla Cochrane,<sup>b</sup> Allison D. Griggs,<sup>a</sup> Brian J. Haas,<sup>a</sup> Thomas Abeel,<sup>a,d</sup> Qiangdong Zeng,<sup>a</sup> Justin B. Nice,<sup>e</sup> Hanlon MacDonald,<sup>e</sup> Bruce W. Birren,<sup>a</sup> Bryan W. Berger,<sup>c,e</sup> Emma Allen-Vercoe,<sup>b</sup> Ashlee M. Earl<sup>a</sup>

Broad Institute, Cambridge, Massachusetts, USA<sup>a</sup>; Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada<sup>b</sup>; Department of Chemical Engineering<sup>c</sup> and Program in Bioengineering,<sup>e</sup> Lehigh University, Bethlehem, Pennsylvania, USA; VIB Department of Plant Systems Biology, Ghent University, Ghent, Belgium<sup>d</sup>

**ABSTRACT** The diverse *Fusobacterium* genus contains species implicated in multiple clinical pathologies, including periodontal disease, preterm birth, and colorectal cancer. The lack of genetic tools for manipulating these organisms leaves us with little understanding of the genes responsible for adherence to and invasion of host cells. Actively invading *Fusobacterium* species can enter host cells independently, whereas passively invading species need additional factors, such as compromise of mucosal integrity or coinfection with other microbes. We applied whole-genome sequencing and comparative analysis to study the evolution of active and passive invasion strategies and to infer factors associated with active forms of host cell invasion. The evolution of active invasion appears to have followed an adaptive radiation in which two of the three fusobacterial lineages acquired new genes and underwent expansions of ancestral genes that enable active forms of host cell invasion. Compared to passive invaders, active invaders have much larger genomes, encode FadA-related adhesins, and possess twice as many genes encoding membrane-related proteins, including a large expansion of surface-associated proteins containing the MORN2 domain of unknown function. We predict a role for proteins containing MORN2 domains in adhesion and active invasion. In the largest and most comprehensive comparison of sequenced *Fusobacterium* species to date, we have generated a testable model for the molecular pathogenesis of *Fusobacterium* infection and illuminate new therapeutic or diagnostic strategies.

**IMPORTANCE** *Fusobacterium* species have recently been implicated in a broad spectrum of human pathologies, including Crohn's disease, ulcerative colitis, preterm birth, and colorectal cancer. Largely due to the genetic intractability of member species, the mechanisms by which *Fusobacterium* causes these pathologies are not well understood, although adherence to and active invasion of host cells appear important. We examined whole-genome sequence data from a diverse set of *Fusobacterium* species to identify genetic determinants of active forms of host cell invasion. Our analyses revealed that actively invading *Fusobacterium* species have larger genomes than passively invading species and possess a specific complement of genes—including a class of genes of unknown function that we predict evolved to enable host cell adherence and invasion. This study provides an important framework for future studies on the role of *Fusobacterium* in pathologies such as colorectal cancer.

Received 26 August 2014 Accepted 13 October 2014 Published 4 November 2014

**Citation** Manson McGuire A, Cochrane K, Griggs AD, Haas BJ, Abeel T, Zeng Q, Nice JB, MacDonald H, Birren BW, Berger BW, Allen-Vercoe E, Earl AM. 2014. Evolution of invasion in a diverse set of *Fusobacterium* species. *mBio* 5(6):e01864-14. doi:10.1128/mBio.01864-14.

**Editor** Michael S. Gilmore, Harvard Medical School

**Copyright** © 2014 Manson McGuire et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Ashlee M. Earl, [aearyl@broadinstitute.org](mailto:aearyl@broadinstitute.org).

The bacterial genus *Fusobacterium* is comprised of at least 13 species that are primarily anaerobic, nonmotile, non-spore-forming, Gram-negative rods and members of the normal human microbiota (1). 16S rRNA gene-based sequencing projects have resolved the *Fusobacterium* genus into groups of species that can be loosely characterized by their interactions with the human host and potential to cause disease (1–3). Some *Fusobacterium* species are capable of “actively” invading host cells without the aid of other factors, whereas other species require compromise of mucosal integrity or coinfection with a virus for host cell invasion (4). The active invader species *F. nucleatum* and *F. periodonticum* are able to independently invade host cells (5, 6), in part using extracellular adhesin and invasion molecules such as FadA (7, 8). This invasion subverts host cell function in ways that are not well understood (9, 10). *F. nucleatum* and *F. periodonticum* are known to be highly adhesive species, displaying selective aggregative ten-

dencies both between strains of the same species as well as with certain unrelated microbial species (11–13). That these species—*F. nucleatum*, in particular—are rapidly gaining notoriety as pathogens contributing to a wide range of human pathologies, including adverse pregnancy outcomes, appendicitis, inflammatory bowel disease, and, most recently, colorectal cancer, makes understanding key steps in the pathogenesis of infection, such as cellular invasion, of great importance (5, 14–17).

In contrast, other *Fusobacterium* species are “passive” invaders, including the well-known veterinary pathogen *F. necrophorum*, which is also the causative agent of human disorders, including Lemierre's syndrome (18). *F. necrophorum* causes damage to host tissues by promoting necrosis (19). The gut resident *F. gonidiaformans* (20), which is primarily nonpathogenic but occasionally causes disease, is closely related to *F. necrophorum*.

A third, less-studied group of fusobacteria consists of *F. mor-*

*tiferum* and *F. varium*, which are frequent residents of the human gut (1, 21), and *F. ulcerans*, which is thought to contribute to the development of tropical skin ulcers (22). While there are some experimental data to suggest that *F. varium* is able to invade host epithelial cells in an active manner, the mechanism for this invasion is unknown (21).

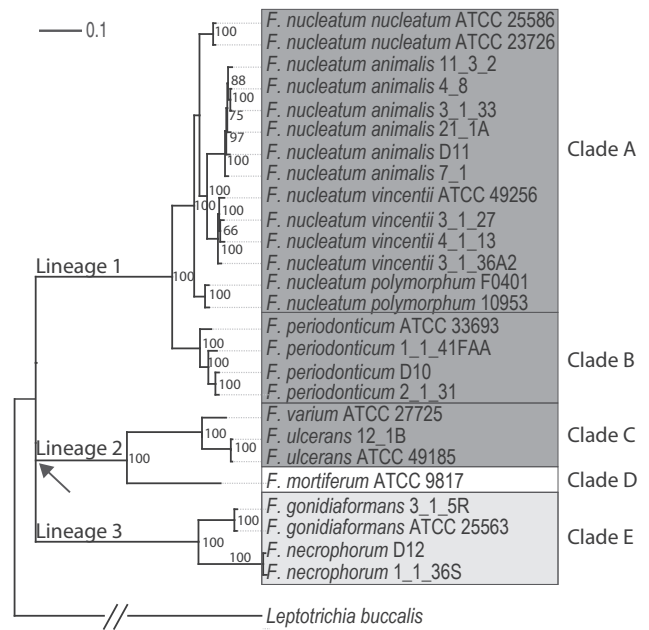
Characterization of *Fusobacterium* biology has been slowed by the fact that members of this genus are largely genetically intractable. They have no known transducing phage or mechanisms for conjugation or natural transformation. Sonoporation has been used to genetically manipulate one species (7, 23), but methods for chemical and electrical competence induction have yet to be developed. As such, it is difficult to engineer mutations and genetically characterize important traits (e.g., active invasion). However, comparative genomics provides a tool to make quantitative associations between traits inferred from gene sequences and known phenotypes of *Fusobacterium* species.

Here, we report whole-genome comparisons of 26 strains representing 7 species belonging to the genus *Fusobacterium*. Our analyses indicate that *Fusobacterium* experienced an adaptive radiation, where three lineages diverged from a common ancestor around the same time. Of these three lineages, two have the ability to actively invade host cells. Features enriched in actively invading strains included a massive expansion of genes encoding membrane-associated proteins, including the known virulence adhesins FadA and RadD, and a set of short, repeated, membrane-associated protein domains designated MORN2 (for *m*embrane *o*ccupation and *r*ecognition *n*exus). MORN2 domain-containing proteins were encoded within sets of genes with no known function and clustered in the same genomic neighborhoods as other adhesins associated with invasion, including FadA and RadD. MORN2 domains were rarer in passively invading species as well as in most other sequenced bacterial species, except for *Helicobacter bilis*, another bacterial species implicated in promoting cancer and preterm birth in animals (24–26). We propose a model in which proteins containing MORN2 domains function to enhance adhesive, aggregative, and invasive traits within select *Fusobacterium* species and may serve an important role in specific disease manifestations like colorectal cancer.

## RESULTS

**Phylogeny and genome content for a highly diverse set of *Fusobacterium* strains.** We generated high-quality draft genome sequences for 21 *Fusobacterium* strains, representing 7 species. Strains were isolated from a variety of human habitats that were either healthy or inflamed at the time of sample collection (see Table S1 in the supplemental material) (5, 22, 27–29). Additionally, we generated finished genomes for five *F. nucleatum* strains (see Table S1) that, together with the two previously finished genomes (30, 31), resulted in the inclusion in our study of seven finished *F. nucleatum* genomes, representing four subspecies.

At the time of sequencing, only partial 16S sequence information was available for 14 isolates, so definitive taxonomic classification was not possible. Using our newly sequenced genomes and 5 previously sequenced *Fusobacterium* genomes (30–32; <http://www.hgsc.bcm.tmc.edu/>; <http://genome.wustl.edu>) (see Table S1 in the supplemental material), we constructed a phylogenetic tree from alignment of 498 orthologous genes, or orthogroups, conserved across all strains (Fig. 1; see Materials and Methods and Table S2 in the supplemental material) and taxo-



**FIG 1** Phylogenetic tree based on nucleotide sequences of 498 core orthogroups, or orthogroups containing exactly one copy from each of the 26 *Fusobacterium* strains plus the outgroup, *Leptotrichia buccalis*. Bootstrap values are indicated for each node. The node indicated with an arrow illustrates a 3-way trifurcation (based on bootstrap values and individual orthogroup trees [see Text S1 in the supplemental material]), representing an adaptive radiation. The five clades are outlined with boxes. The clades containing species believed to actively or passively invade host cells are indicated with dark and light gray shading, respectively. The mechanism by which *F. mortiferum* (in white) can invade cells is unknown.

nomically classified previously unnamed clinical isolates (see Fig. S1 in the supplemental material). The resolved taxonomy, showing that species fall into three main lineages (Fig. 1), was in general agreement with the taxonomy based on 16S sequences from the Living Tree Project (2). However, our results allowed for a more highly resolved view of evolution within this genus and revealed that there was an adaptive radiation, where the last common ancestor of all *Fusobacterium* species diversified into three major lineages (see Text S1 in the supplemental material) at an early point in its evolution from *Leptotrichia*.

We observed substantial variation in genome sizes and gene content among the sequenced species (see Table S1 and Fig. S2 in the supplemental material). Genome sizes differed by as much as 2 Mb (1.7 to 3.7 Mb). Only 556 orthogroups (24%) were shared among all strains. Members of the same species and subspecies exhibited remarkable gene content plasticity: *F. nucleatum* strains shared only 59% of their orthogroups, while members of the closely related *F. nucleatum* subsp. *animalis* subspecies shared only 70%, despite having >99% nucleotide identity among shared genes. In addition, few orthogroups uniquely defined each species and subspecies. (For example, 1%, or 36 orthogroups, were found exclusively in all 14 *F. nucleatum* genomes, and 1%, or 32 orthogroups, were found exclusively in all 6 *F. nucleatum* subsp. *animalis* genomes.) Average nucleotide identity (ANI) plots based on whole-genome data (33, 34) (see Fig. S1 in the supplemental material) suggested that each *F. nucleatum* subspecies could be considered a separate species (89 to 93% ANI), and *F. periodonticum* could also be subdivided into separate species, with some strain

comparisons having only 92 to 94% ANI (see Text S1 in the supplemental material), likely contributing to the low numbers of species-defining genes. Species with fewer genomes had more defining orthogroups (e.g., 6%, or 220, were found exclusively in *F. ulcerans*), likely due to their being underrepresented in our data set (see Table S3 in the supplemental material) and their higher ANI (>95%) (see Text S1). Species-specific orthogroups included genes encoding extracellular features, such as adhesins, membrane-associated transporters, receptors, and extracellular solute-binding proteins (see Text S1). That some adhesins and membrane-associated proteins were species specific has been previously documented in *Fusobacterium* (35) and points to surface attachment as a driver of the evolution of this genus.

Since the majority of species- and subspecies-defining orthogroups encoded “hypothetical proteins” of unknown function, we compared the sequences of genes against those in functional databases, including the KEGG (36), Pfam (37), and Gene Ontology (GO) (38) databases, to increase our power to assign function to hypothetical proteins. We then searched for broader functional categories expanded in individual species and subspecies that were not obvious by gene annotation and ortholog clustering (see Table S4 and Text S1 in the supplemental material) but that might help to explain what drove speciation and adaptation of *Fusobacterium* species to different environments. Again, the results indicated that the majority of species-specific orthogroups encoded extracellular or membrane-related proteins, including 46 of the 72 species-specific orthogroups present in either *F. nucleatum* or *F. periodonticum* (64%). Expansions were also observed among genes related to amino acid metabolism and cofactor biosynthesis (*F. nucleatum* subsp. *polymorphum* and *F. periodonticum*), gene regulation and signaling (*F. ulcerans*), YadA-related autotransporter adhesins (*F. necrophorum*), and species-specific hemagglutinin-related genes (*F. nucleatum*).

**Identification of new determinants associated with active invasion.** Only some *Fusobacterium* species can invade host cells independently (active invaders), while others require either helper organisms, coinfection with viruses, or compromised mucosal integrity (passive invaders) (4–6). The FadA adhesin was previously identified as an important factor in active invasion (7, 8) and has been identified in several actively invading *Fusobacterium* species (39). To identify other genetic factors associated with active invasion, we first stratified species into active and passive invader clades using a Bayesian approach (see Materials and Methods) and then correlated this structure with information from the literature regarding each species’ invasion potential (4–6, 21). Five distinct clades were identified among *Fusobacterium* species (Fig. 1). Three clades—clades A (*F. nucleatum*) and B (*F. periodonticum*) belonging to lineage 1 and clade C (*F. ulcerans*/*F. varium*) belonging to lineage 2—consisted entirely of species known to actively invade host cells (active invaders). Clade D, which included the remaining species in lineage 2, *F. mortiferum*, was of unknown invasion potential. Clade E constitutes lineage 3 and contained the known passive invader species *F. necrophorum* and *F. gonidiaformans*.

Active invaders had genomes that were, on average, 560 kb larger than those of passive invaders and contained 257 more genes (see Table S1 in the supplemental material). A large fraction of the additional genes encoded membrane-related proteins (as defined by GO criteria), doubling the membrane-related protein coding capacity in active invaders. Active invaders also had 1.6-fold more genes with predicted signal peptides (365 versus 233 per

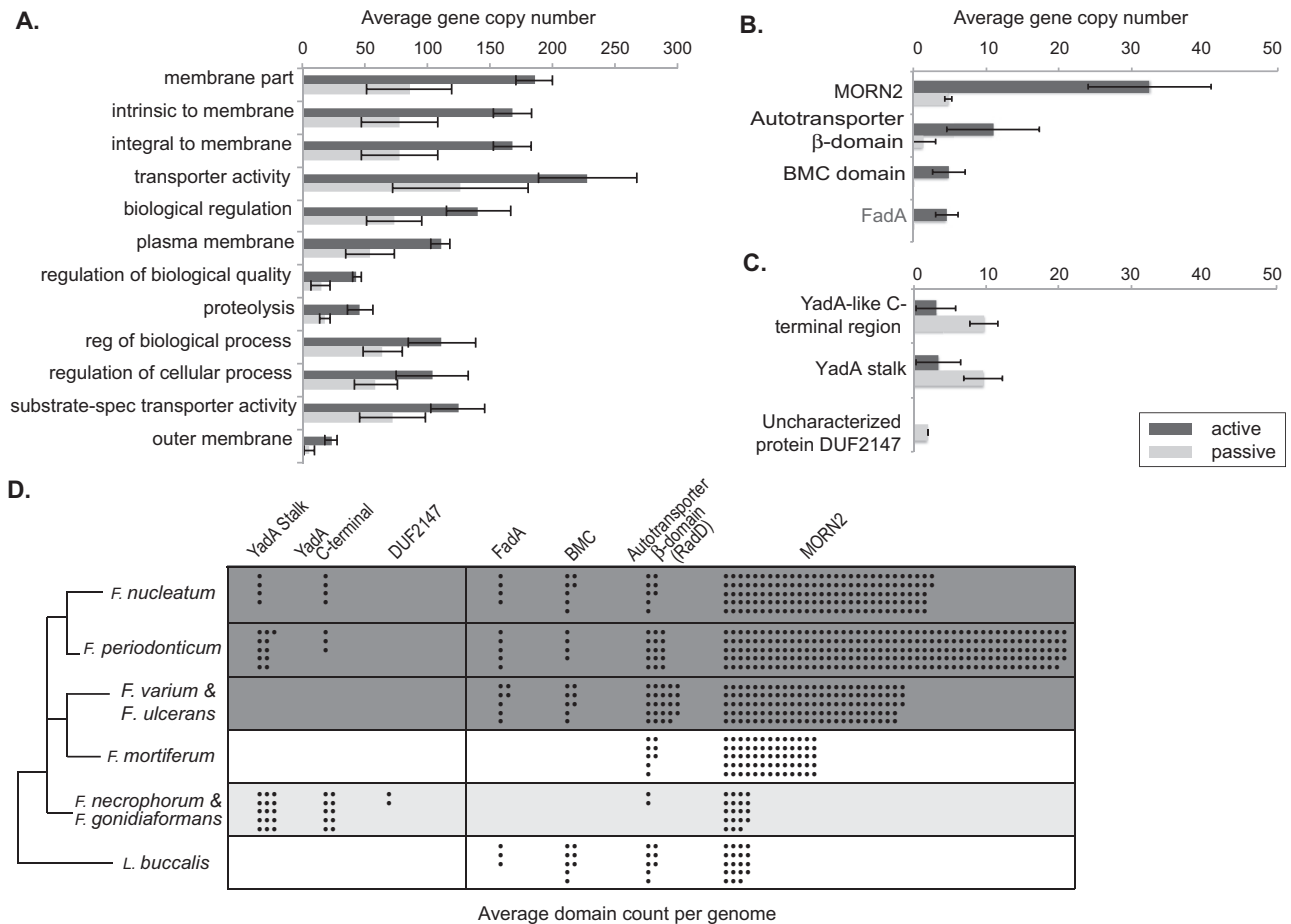
genome), suggesting an extracellular role for expanded gene families in active invasion.

When we compared orthogroups between active and passive invaders, 44 were present in all active invaders and absent in all passive invaders (see Table S2b and Table S5 in the supplemental material). Of the genes exclusive to active invaders, 32 were annotated as “hypothetical proteins.” However, we were able to gain clues to the function of many by examining Pfam or GO functional annotations. This highlighted several genes with known links to virulence and pathogenesis, including those encoding branched-chain amino acid transport (40, 41), components of type IV pili (42) and the related bacterial type II secretion system (43–46), a patatin-like phospholipase (47), and META domain-containing (48) and chorismate mutase domain-containing (49, 50) proteins. In addition, 43% (19 of 44) of the active invader-specific genes were predicted to contain either a signal peptide or transmembrane domain, highlighting the importance of evolved surface features on the invasive strains.

To ensure that the maximum number of genes were accounted for in the comparative analysis of these high-quality draft genomes, we used the GO, Pfam, and KEGG functional annotations to identify additional differences in actively versus passively invading species. Using this approach, additional GO-determined functional families were found that had significantly larger numbers of genes in the active invaders, with most being associated with the membrane ( $Q < 5e-10$ ) (Fig. 2A; see Table S5 in the supplemental material), consistent with our initial analysis identifying a 2-fold enrichment in membrane proteins among this group. Of the seven significantly enriched Pfam domains (Fig. 2B and C), five were associated with adherence, including the FadA adhesin domain ( $Q < 4.5e-21$ ). By this analysis, FadA family-encoding genes were found to be exclusive to active invader genomes, including both strains of *F. ulcerans*. It had been suggested previously (39) that FadA was not present in *F. ulcerans* ATCC 49185; however, we found seven related FadA family genes in that strain and eight in *F. ulcerans* 12-1B. These were not clustered within the same orthogroup as canonical FadA (FN0264 in *F. nucleatum*) because they shared only 30 to 38% amino acid identity. The passive invaders, *F. necrophorum* and *F. gonidiaformans*, completely lacked FadA family genes, as did *F. mortiferum*.

Interestingly, active invaders also contained 6 times as many genes encoding the adhesin-related autotransporter  $\beta$ -domain ( $Q < 4.4e-7$  [11.0 versus 1.6 genes]) (51), a component of the RadD family of outer membrane protein adhesins known to be involved in pathogenesis of *F. nucleatum* through cell death in human lymphocytes (51, 52). Bacterial microcompartment (BMC) domains were also found exclusively in the active invaders ( $Q < 9.9e-5$  [4.9 genes]). BMC-containing proteins are associated with polyhedral microcompartment organelles, such as carboxysomes, which give bacteria the ability to adapt to new niches via metabolic innovation and have been previously implicated in pathogenesis (53) (Fig. 2B). While BMC domains have not been directly implicated in adherence, these genes in *F. nucleatum* subsp. *nucleatum* ATCC 25586 were recently shown to be highly induced under conditions that promote aggregation (11), suggesting that these proteins are also involved in aggregation and adhesion.

Of perhaps greatest importance, MORN2 domains (Pfam identifier PF07661) were found to be the most extensively enriched domain among active invaders (Fig. 2B) ( $Q < 2.5e-21$  [32.4 versus 4.8 genes]). None of the 697 genes encoding MORN2



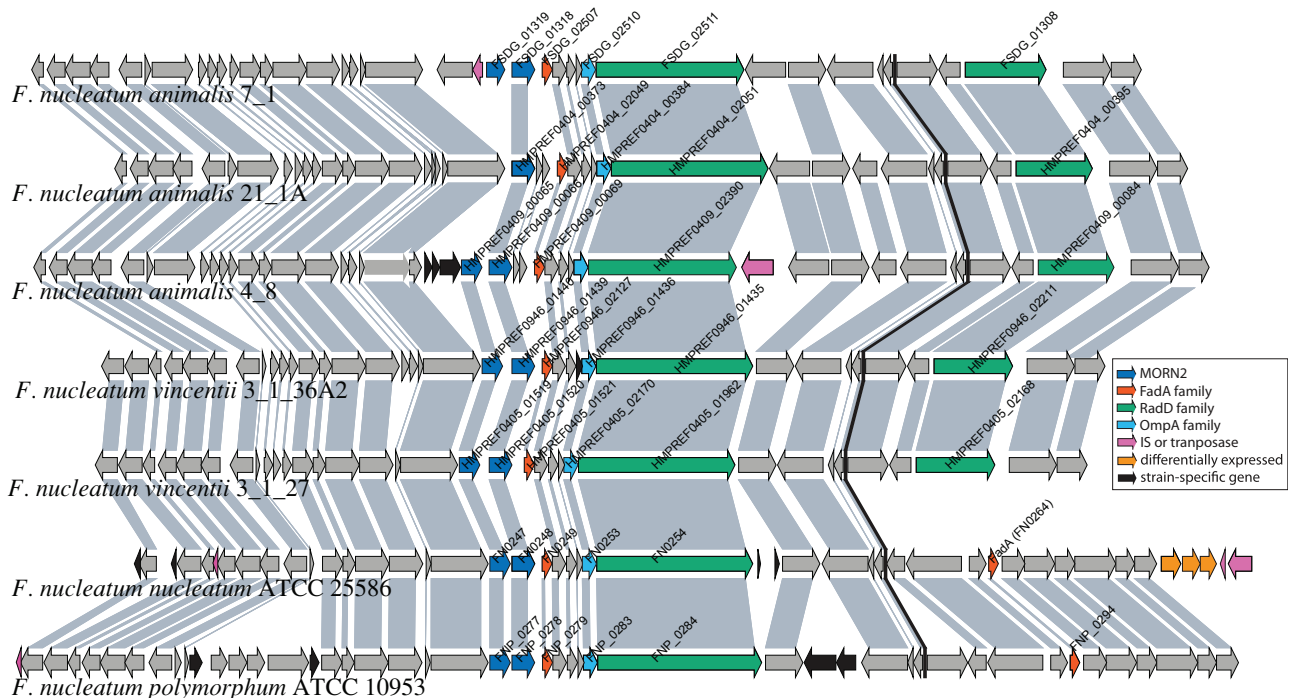
**FIG 2** Gene categories expanded in active and passive invaders ( $Q < 0.0005$ ). (A) The top 12 GO terms expanded in the active invaders were largely membrane related. (B) Pfam domains expanded in the active invaders include the known virulence-related adhesins FadA and RadD, as well as a massive expansion of MORN2 domains of unknown function. (C) A different set of adhesins, including the trimeric autotransporter adhesion protein YadA, is expanded in the passive invaders. (See Table S5 in the supplemental material for a full listing of categories expanded in both active and passive invaders for  $Q < 0.05$ .) (D) Clade-specific summary of expanded Pfam domains. The number of dots represents the average number of domains present in a genome from each of the five clades. For MORN2, there can be multiple domains per gene.

domains (see Table S2c in the supplemental material) within our data set had known function, although 87% were predicted by SignalP (54) to have signal peptides targeting them for export into the extracellular environment or membrane insertion, highlighting a potential role for these domains at the host-pathogen interface. To further validate these results, we verified that all the domains enriched in the active invaders were also enriched in six more recently sequenced *F. nucleatum* genomes isolated from cancerous tumors (see Text S1 in the supplemental material). We also verified that all 44 of the orthogroups exclusive to active invader strains were also present (see Text S1).

Enriched features in passive invaders included two types of YadA domain proteins ( $Q < 1.0e-5$  [9.6 versus 3.4 copies] and  $Q < 2.0e-6$  [9.7 versus 3.1 copies]). These proteins are found in trimeric autotransporter adhesins implicated in host cell adherence (55). Passive invaders also exclusively encoded proteins with the domain DUF2147 (domain of unknown function) ( $Q < 1.1e-4$  [2 copies per species]). Genes with DUF2147 also had predicted signal peptides and were located near other genes that encoded predicted membrane proteins, suggesting roles at the host-microbe interface.

**Evolution of active invasion.** With few exceptions (see Table S5 in the supplemental material), orthogroups unique to active invading strains were scattered around the genome, with little evidence of having been acquired together. However, genes containing predicted Pfam structural domains enriched among active invaders tended to cluster together (Fig. 3; see Text S1 in the supplemental material), suggesting that they evolved together for related function, such as novel surface feature assembly. This observation was supported by neighborhood analysis, which showed a statistically significant association among orthogroups containing the expanded Pfam families—MORN2, FadA, and RadD—occurring near each other and also near genes encoding other virulence-related proteins and membrane proteins (see Table S6 and Text S1 in the supplemental material). While IS elements and proteins related to phage and transposition were found near regions containing Pfam families expanded in active invaders (see Table S6 and Text S1), there was little evidence of them being part of recognizable prophages or recently acquired through lateral transfer (see Materials and Methods), although one gene containing MORN2 domains was present on a predicted plasmid in *F. nucleatum* 3\_1\_27 (see Table S1 in the supplemental material).





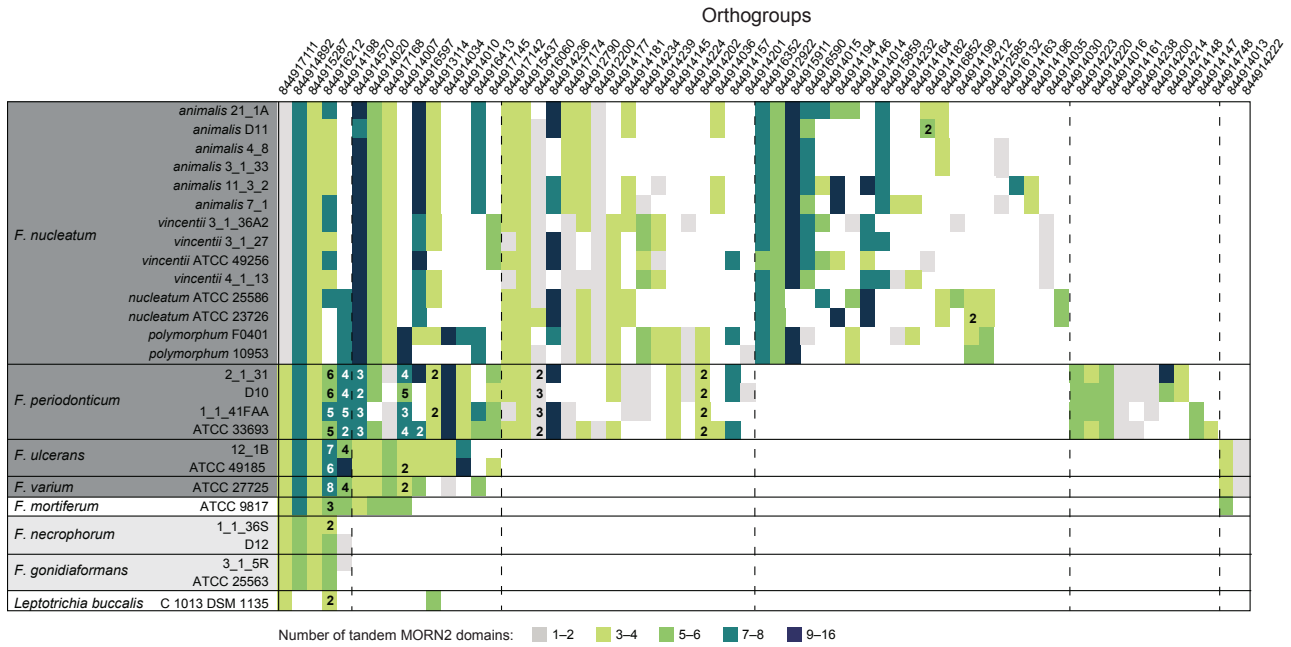
**FIG 3** Gene families expanded in active invaders clustered in the genome. This multiple alignment of our set of seven finished *F. nucleatum* genomes shows an example of a region with close physical association among FadA, RadD, and MORN2 family proteins. The gray-shaded connectors represent orthologous protein pairs. This genomic region also contains an *ompA* gene shown to be involved in biofilm formation in *Fusobacterium* (69), as well as numerous strain-specific genes, transposases, IS elements, and genes previously identified as being differentially expressed under aggregation conditions (11). In two genomes (*F. nucleatum* subsp. *nucleatum* ATCC 25586 and *F. nucleatum* subsp. *polymorphum* ATCC 10953), a syntenic break (indicated with a black line) placed the canonical *fadA* gene (FN0264) in this region rather than a second *radD* gene. Several regions similar to this are located in the genome.

To gain evidence for whether the active invader-specific orthogroups were acquired in the adaptive radiation (Fig. 1), we examined the occurrence of the 44 active invader-specific orthogroups in the closely related outgroup *Leptotrichia buccalis*, the presence of which would suggest that the *Fusobacterium* ancestor possessed them prior to the radiation and that they were lost by passive invaders. A minority (17 of 44 [39%]) of active invader-specific orthogroups were found in *L. buccalis*. Those predicted from this analysis to have been lost in the radiation of passive invaders included genes encoding a membrane-bound two-component sensor/histidine kinase, a META domain-containing protein, a hypothetical transporter, and a secreted peptidoglycan catabolism protein. Genes not present in the outgroup and likely to have been acquired by active invaders in the radiation included four of the six orthogroups previously implicated in virulence in other species, including those encoding branched-chain amino acid transport and type IV pili (40, 41, 44–46), as well as 15 of the 19 genes encoding predicted membrane-bound or secreted proteins, highlighting acquired changes in surface organization of the active invaders.

Interestingly, the complement of MORN2 domains remained relatively unchanged between *Leptotrichia* and the passive invaders but was greatly expanded in the active invader clades (Fig. 2D). In contrast, FadA and BMC domain-containing genes, both present in *Leptotrichia*, appeared to have been lost in the passive invaders, as well as in *F. mortiferum*. Yada was not observed in *Leptotrichia* but may not be necessary for active invasion, since the active invader clade C (*F. ulcerans*/*F. varium*) lacked genes con-

taining these domains. That *F. mortiferum* contained no Yada, FadA, or BMC domains and only a modest expansion of MORN2 domains suggested that it might not be capable of active invasion. Overall, our data predict an evolutionary model whereby (i) the last common ancestor of *Leptotrichia* and *Fusobacterium* shared active invasion-associated FadA, BMC, RadD, and MORN2 domain-containing genes, (ii) FadA and BMC were lost during the evolution of *F. mortiferum*, *F. necrophorum*, and *F. gonidiaformans*, and (iii) MORN2- and, to some degree, RadD domain-containing genes underwent expansion in the active invaders. In support of this model, we observed that at least three *Leptotrichiaceae* genomes in the Pfam database also shared FadA, RadD, and BMC domains and a small set of ancestral MORN2 domains (data not shown).

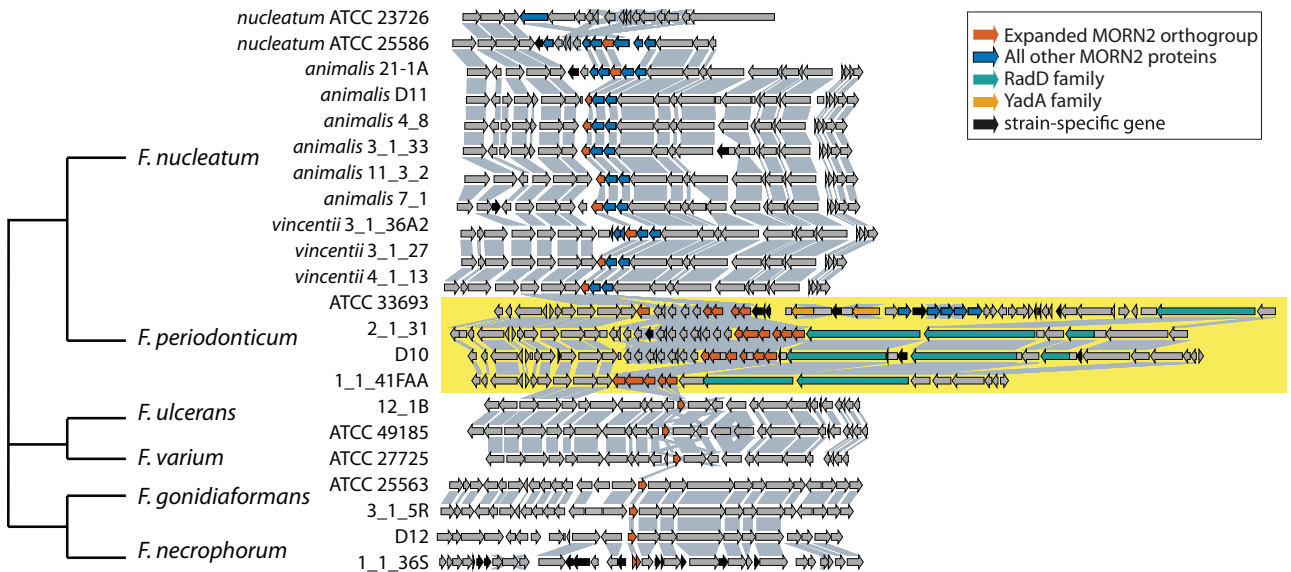
**MORN2 evolution and function.** Of the expanded Pfam domains, MORN2 was among the most intriguing because it represented the most frequent domain in active invader genomes (115 to 250 MORN2 domain copies per genome), and nothing was known about the function of proteins containing this domain. The MORN2 domain itself is 22 to 23 amino acids long, often found in multiple copies per gene, and was highly variable across the strains in our study. MORN2-containing proteins are incredibly diverse in the organization and grouping of domains, which is comprehensively illustrated at <http://pfam.xfam.org/family/PF07661>. The number of genes containing MORN2 domains (MORN2 genes) was also variable, with members of the same subspecies differing in their MORN2 gene content by as much as 25% (Fig. 4). The small number of MORN2 genes found in the



**FIG 4** Species distribution of MORN2 orthogroups. Each column represents an orthogroup. The boxes are colored according to the average number of MORN2 domains per protein, and for each orthogroup containing two or more paralogs, the numbers indicate the number of paralogs. There is tremendous variation in the number and structure of proteins containing MORN2 domains across genomes. Active invaders (species names shaded in dark gray) contain longer, more complex MORN2-containing proteins than passive invaders (species names shaded in light gray). Passive invaders contain only a small complement of short, “ancestral” orthogroups, which are present across all species. Expansions of active invader orthogroups, especially in *F. periodonticum*, can be seen by the presence of paralogs.

passive invaders was also present in nearly all active invaders, suggesting that MORN2 genes evolved and expanded in active invader species from an ancestral set of MORN2 genes held by the last common ancestor at the time of the adaptive radiation. Supporting this, we observed (i) no evidence that MORN2 genes were

recently acquired on genomic islands (using Islandviewer [56]), (ii) MORN2 genes clustered in active invader genomes (Fig. 3), with evidence of recent local duplication in some species (Fig. 5), and (iii) there was significantly greater sequence identity between pairs of MORN2 domains that were spatially close but not in the



**FIG 5** Expansion of an “ancestral” MORN2 orthogroup in *F. periodonticum*. This orthogroup is present in all species, including the passive invaders, but it is greatly expanded in *F. periodonticum*. In *F. periodonticum*, this expanded cluster is also located near other known virulence-related adhesins, including RadD and YadA family members. All members of this orthogroup are present in this small region of the genome. Arrows represent syntenic orthogroup members. The orthogroup identification no. for this expanded group is 844916212 (see Table S2 in the supplemental material).

same protein (38.3%), compared to those >10 kb away (31.3%) ( $P < 2.2e-16$ ). We also observed an unusually high level of overall chromosomal rearrangements in actively invading *Fusobacterium* strains, which could be linked to the high rate of genome rearrangements within MORN2 regions (see Fig. S3 and Text S1 in the supplemental material).

Since nothing was previously known about the function of MORN2 genes, we sought clues to their function by comparing the ecologies and lifestyles of other bacteria that exhibited expanded sets of MORN2 domains. Remarkably, although MORN2 domains were present in a diverse set of bacterial genomes (see Text S1 in the supplemental material), the only other bacterium that contained >100 MORN2 domains, like active invader *Fusobacterium*, was *Helicobacter bilis* ATCC 43879, which colonizes the bile, liver, and intestines (24). The closely related species *Helicobacter hepaticus* also possessed a relatively large number of MORN2 domains, 28 copies, which was higher than the number in the passively invading fusobacteria. *H. bilis* and the related *H. hepaticus* are both involved in colitis, hepatitis, and, similar to actively invading *Fusobacterium* spp. exhibit carcinogenic potential and association with preterm birth in animals (24–26). Although it is unknown whether *H. bilis* or *H. hepaticus* actively invades host cells, given (i) the striking similarity in pathology associated with these two species and actively invading *Fusobacterium* and (ii) the fact that MORN2 genes colocalize with the expanded FadA family and RadD family genes known to be involved in adhesion and virulence in *Fusobacterium*, we propose a testable model that predicts a related role for MORN2-containing proteins in adhesion and pathogenicity and possibly in promoting cancer or cancer progression.

## DISCUSSION

This study represents, by far, the largest comparative genome analysis of *Fusobacterium* strains, and includes 26 strains of 7 species, sourced from diverse human body sites and disease states. We find that *Fusobacterium* species exhibit a high degree of heterogeneity. Even within subspecies, which our ANI analysis indicates should actually be called separate “species,” there is substantial variation in genome size, architecture, and content. Our analyses highlighted features unique to individual species and subspecies, providing clues to how this genetically intractable genus of opportunistic pathogens diverged and the selective forces that likely drove divergence. Notably, we observed a large number of species-specific genes encoding membrane proteins and adhesins consistent with *Fusobacterium* having an unusual ability to adhere to various ligands (57), including to host cells (58), and highlighting the host-microbe interface as an important driver of diversity within this genus (35).

**Evolution of active invasion.** The most prominent features of an active invader genome are its large size and the fact that it contains twice as many genes encoding membrane-associated proteins as passive invader genomes. In addition, active invader genomes are distinguished by encoding (i) a large and expanded set of proteins containing MORN2 domains, and, to a lesser extent, RadD family adhesin domains, (ii) FadA family adhesin proteins and proteins containing BMC domains, previously implicated in host cell invasion and auto-aggregation, respectively (11), (iii) additional adhesion-related proteins, including type IV pili, which encode extracellular adhesive appendages, and (iv) proteins containing domains previously implicated in virulence, such as

AzIC, AzID, META, chorismate mutase, and patatin-like phospholipases.

Our data set contained more active invaders than passive invaders. However, our observations were consistent across the passive invaders, including representatives from two species. Passive invader genomes consistently contained fewer membrane proteins, a different complement of virulence adhesins, and strikingly fewer MORN2 domains, despite their variation in size and gene count. Additionally, all MORN2 domains found in each of the four passive invaders had orthologs in the active invaders, providing a strong indication that the MORN2 genes found in the passive invaders were present before the adaptive radiation event and evolution of active invasion. We used strict Q value cutoffs for our Pfam domain analysis, to eliminate noise due to the small number of passive invaders in our data set. Additional genomes from these species would help to validate our results.

Our analyses revealed that the *Fusobacterium* genus likely underwent an adaptive radiation event, in which three lineages diverged from a common ancestor at a similar point in time (Fig. 1). Lineage 1, containing active invader clades, split into *F. nucleatum* and *F. periodonticum*. Lineage 3 split into *F. ulcerans* and *F. varium* (both of which can actively invade host cells), and *F. mortiferum* (the invasion phenotype of which is unknown). Lineage 2 (containing *F. necrophorum* and *F. gonidiaformans*) contained primarily species requiring additional factors to invade cells, such as coinfection with other microbes (“passive invaders”). Our data suggest that, in this adaptive radiation, clades within two of the three lineages acquired (or retained) the ability for active invasion, whereas lineage 2, containing *F. necrophorum* and *F. gonidiaformans*, did not, due to losing virulence-associated factors (such as FadA and BMC) or never acquiring these factors and/or undergoing the required expansion of other genes, like MORN2 genes, to become active invaders. It is unclear whether *F. mortiferum* should be classified as an active or passive invader. *F. mortiferum* lacks FadA and BMC domains. It contains 59 MORN2 domains, which is intermediate between the active invaders (>100 MORN2 domains per genome) and the passive invaders (17 to 20 MORN2 domains per genome), and it also contains more copies of RadD domain-containing genes than the passive invaders (8 copies rather than the 0 to 3 copies observed in the passive invaders). Our data suggest that *F. mortiferum*'s invasion strategy is different from those of other *Fusobacterium* spp.

In addition to the large variation in host cell invasion properties between the active and passive invaders, differences in invasion potential have also been observed within the *F. nucleatum* species (5). *F. nucleatum* strains from several different subspecies isolated from diseased tissue tended to be more invasive than examples from the same subspecies isolated from healthy tissue (5). We were unable to find any clear difference in gene family content or metabolic pathway composition between these two sets of *F. nucleatum* strains (data not shown). Further investigations are needed to find an explanation for these differences in invasiveness among active invader species.

**Diversity and evolution of genes containing MORN2 domains.** There appears to have been an ancient event leading to a striking expansion and diversification of MORN2 domains in the active invaders. Further diversity-yielding events, potentially mediated by the repetitive nature of the tandem domain repeats themselves, as well as mobile elements, such as IS elements and transposases, have further diversified and expanded MORN2 do-



mains within each clade. The resulting diversity of MORN2 genes may contribute to *Fusobacterium*'s ability to adapt to new environments and infect many types of cells.

We see a spatial correlation between MORN2 genes and mobile elements (IS elements and transposases; see Text S1 in the supplemental material), indicating that mobile genetic elements could be involved in their diversification; however, there was no evidence to suggest that MORN2 genes were recently acquired. Instead, our data indicate that the expanded set of MORN2 genes in active invaders arose primarily through local duplication (Fig. 5) from an ancestral set of genes containing relatively few repeats. Tandem domain repeats themselves are often facilitators of rapid evolution, often imparting useful phenotypic consequences, including rapid variation in microbial cell surface (59). Recombination between repeat sequences has been proposed to explain expansions and contractions in the number of repetitive elements within genes (59) and may explain why MORN2 genes from active invaders possess more copies of the MORN2 domain than passive invaders.

**Function of proteins containing MORN2 domains.** In addition to genes with well-documented roles, many of the active invader-specific genes have no known function, including the MORN2 genes. With only one exception in *H. bilis*, the massive expansion of MORN2 domains observed in the active invaders is a feature highly specific to the *Fusobacterium* genus (see Text S1 in the supplemental material). There are two related families of MORN domains found in *Fusobacterium*: MORN (Pfam identifier PF02493), found primarily in eukaryotes, and MORN2 (Pfam identifier PF07661), found primarily in bacteria, and concentrated within the genus *Fusobacterium*. Previous work on MORN domains in eukaryotes implicates them in the mediation of interactions between cytoplasmic membranes and other intracellular structures, such as the cytoskeleton, endoplasmic reticulum, and kinases (60–64). In contrast, nothing is known about the function of prokaryotic MORN or MORN2 domains, and, unlike eukaryotic MORN-containing proteins, MORN2 proteins appear to function outside the cell, due to the presence of signal peptide sequences. There are only a few examples of MORN domains present in actively invading *Fusobacterium* species (an average of 4 domains in 1.6 proteins per genome), and these are often found in the same gene as MORN2 domains. Because of the close similarity between MORN and MORN2 and the fact that no MORN domains were found in *Leptotrichia*, we propose that fusobacterial MORN domains are misclassified and are, in fact, a variant of MORN2 domains that evolved from the closely related MORN2 within the active invaders. Besides MORN, the only other domains that occasionally coincide with MORN2 in the same gene are DnaJ and the chorismate mutase domain. The chorismate mutase domain has been implicated in virulence in other bacteria, including *Mycobacterium tuberculosis*, and is often found on pathogenicity islands (49, 50).

The domain organization of MORN2 genes points to a possible role for their products in adhesion. Examples of diverse expanded families of adhesins have been observed in other organisms, where variable numbers of repeats in cell wall proteins allow for rapid modulation of adhesive properties, adaptation to the environment, or evasion of the host immune system (59, 65, 66). Often these domains represent subunits, which oligomerize to form a large variable structure. In *H. pylori*, genes encoding Sel1-like repeats (SLRs) are involved in adaptation of *H. pylori* to specific hosts,

due to strain-specific variations in the number of SLRs (67). SLR genes are similar to MORN2 genes, in that they are of similar length, are poorly conserved, and have repetitive domains, with a similar pattern of conserved residues. In fact, Sel1 domains are found in the same proteins as MORN2 domains in some strains of *Escherichia coli*, *Shigella*, *Salmonella*, and *Yersinia*. The repetitive, modular structure of MORN2 genes points to their involvement in adhesion and in promoting rapid adaptation to diverse environmental conditions.

In addition, the spatial organization of MORN2 genes within the genome points to a possible role in virulence, invasion, and adhesion. Genes containing MORN2 domains are found clustered with genes encoding known virulence adhesins (FadI, FadA, and RadD, as well as other FadA and RadD family proteins), membrane-associated pathogenicity factors (such as OmpA proteins), and virulence factors (including chorismate mutase domain-containing proteins) (49, 68), as well as secreted and membrane-associated proteins of unknown function. That expansion of MORN2 is so particular to actively invading species of *Fusobacterium*, in addition to a distantly related organism also implicated in pathologies similar to those of actively invading *Fusobacterium* species, suggests involvement of MORN2 genes in pathogenicity in the active invaders. The fact that these expansions are present in all *F. nucleatum*, including those resident in the healthy mouth as well as those in cancerous tumors, indicates that these invasive capabilities may be characteristic of all *F. nucleatum*. Targeting MORN2 proteins, as well as other active invader-specific proteins of unknown function, in future research may reveal functions in adhesion and invasion and indicate good anti-virulence targets, which would be of high interest, considering the association between *F. nucleatum* and human maladies, such as colorectal cancer (16, 17).

**Conclusions.** *Fusobacterium* species have steadily gained attention as important bacterial pathogens, now implicated in a diverse range of human pathologies, including colorectal cancer and preterm birth. However, largely due to their genetic intractability, little is known about the mechanisms that have allowed some species to become such pervasive pathogens. In the largest comparison of fusobacterial genomes, our work has helped to close this knowledge gap by (i) constructing the highest-resolution phylogeny of *Fusobacterium* to date, (ii) characterizing the gene content and genomic architecture of member species, (iii) identifying genetic features and molecular pathways that distinguish the most invasive forms of *Fusobacterium*, and (iv) explaining the evolution of active forms of host cell invasion. Importantly, we have discovered a class of genes of unknown function that strongly associate with active forms of host cell invasion and likely represent new strategies for bacterial adherence to and invasion of host cells. The insights gained represent an important step forward in unraveling the mechanisms of *Fusobacterium* pathogenesis and will enable development of diagnostic and therapeutic strategies for the detection and treatment of fusobacterial disease.

## MATERIALS AND METHODS

The strains selected for genome analysis are described in Table S1 in the supplemental material. Methods for DNA sequencing, genome assembly and annotation, orthogroup clustering, phylogenetic analysis, renaming of strains, ANI and shared-gene analysis, multiple alignments, and other bioinformatics analyses are described in the supplemental material.



## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01864-14/-/DCSupplemental>.

Text S1, DOCX file, 0.1 MB.  
Figure S1, PDF file, 0.5 MB.  
Figure S2, PDF file, 0.8 MB.  
Figure S3, JPG file, 2.5 MB.  
Table S1, PDF file, 0.2 MB.  
Table S2, XLSX file, 0.8 MB.  
Table S3, XLSX file, 0.1 MB.  
Table S4, DOCX file, 0.2 MB.  
Table S5, XLSX file, 0.1 MB.  
Table S6, XLSX file, 0.1 MB.

## ACKNOWLEDGMENTS

This work was undertaken on behalf of the Human Microbiome Project Consortium and generously supported by the NIH, NHGRI, and NIAID (U54-HG004969 to the Broad Institute). This project was also funded in part by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. HHSN272200900018C (A.M.-M. and A.M.E.).

We thank the members of the Broad Institute's COBRA Team, including Christopher Desjardins, Paul Godfrey, Varun Mazumdar, Gustavo Cerqueira, and Jennifer Wortman, as well as Yiping Han at Case Western Reserve University, for helpful discussions. We acknowledge the Broad Institute's Genome Sequencing Platform, including Susanna Hamilton, Sequencing Project Manager for HMP, the Broad Institute's Assembly Teams, including Sarah Young, Theresa Hepburn, Sakina Saif, Michael Fitzgerald, Harindra Arachchi, and Peg Priest, the Broad Institute's Annotation Team, including Narmada Shenoy, Teena Mehta, Chandri Yandava, and Lucia Alvarado, and the Broad Institute's Finishing Team, including Pendexter MacDonald, Alma Imamovic, Annie Lui, Amr Abouelleil, Gary Gearin, Anna Montmayeur, and Caryn McCowan. We thank Leslie Gaffney (Broad Institute) for help with figures.

## REFERENCES

- Citron DM. 2002. Update on the taxonomy and clinical aspects of the genus *Fusobacterium*. *Clin. Infect. Dis.* 35:S22–S27. <http://dx.doi.org/10.1086/340983>.
- Munoz R, Yarza P, Ludwig W, Euzéby J, Amann R, Schleifer KH, Glöckner FO, Rosselló-Móra R. 2011. Release LTPs104 of the all-species living tree. *Syst. Appl. Microbiol.* 34:169–170. <http://dx.doi.org/10.1016/j.syapm.2011.03.001>.
- Conrads G, Claros MC, Citron DM, Tyrrell KL, Merriam V, Goldstein EJ. 2002. 16S–23S rDNA internal transcribed spacer sequences for analysis of the phylogenetic relationships among species of the genus *Fusobacterium*. *Int. J. Syst. Evol. Microbiol.* 52:493–499. <http://dx.doi.org/10.1099/ijs.0.01972-0>.
- Huggan PJ, Murdoch DR. 2008. Fusobacterial infections: clinical spectrum and incidence of invasive disease. *J. Infect.* 57:283–289. <http://dx.doi.org/10.1016/j.jinf.2008.07.016>.
- Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, Devinney R, Lynch T, Allen-Vercoe E. 2011. Invasive potential of gut mucosa-derived *Fusobacterium nucleatum* positively correlates with IBD status of the host. *Inflamm. Bowel Dis.* 17:1971–1978. <http://dx.doi.org/10.1002/ibd.21606>.
- Han YW, Shi W, Huang GT, Kinder Haake S, Park NH, Kuramitsu H, Genco RJ. 2000. Interactions between periodontal bacteria and human oral epithelial cells: *Fusobacterium nucleatum* adheres to and invades epithelial cells. *Infect. Immun.* 68:3140–3146. <http://dx.doi.org/10.1128/IAI.68.6.3140-3146.2000>.
- Ikegami A, Chung P, Han YW. 2009. Complementation of the fadA mutation in *Fusobacterium nucleatum* demonstrates that the surface-exposed adhesin promotes cellular invasion and placental colonization. *Infect. Immun.* 77:3075–3079. <http://dx.doi.org/10.1128/IAI.00209-09>.
- Xu M, Yamada M, Li M, Liu H, Chen SG, Han YW. 2007. FadA from *Fusobacterium nucleatum* utilizes both secreted and nonsecreted forms for functional oligomerization for attachment and invasion of host cells. *J. Biol. Chem.* 282:25000–25009. <http://dx.doi.org/10.1074/jbc.M611567200>.
- Fardini Y, Wang X, Témoins S, Nithianantham S, Lee D, Shoham M, Han YW. 2011. *Fusobacterium nucleatum* adhesin FadA binds vascular endothelial cadherin and alters endothelial integrity. *Mol. Microbiol.* 82:1468–1480. <http://dx.doi.org/10.1111/j.1365-2958.2011.07905.x>.
- Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. 2013. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/ $\beta$ -catenin signaling via its FadA adhesin. *Cell Host Microbe* 14:195–206. <http://dx.doi.org/10.1016/j.chom.2013.07.012>.
- Merritt J, Niu G, Okinaga T, Qi F. 2009. Autoaggregation response of *Fusobacterium nucleatum*. *Appl. Environ. Microbiol.* 75:7725–7733. <http://dx.doi.org/10.1128/AEM.00916-09>.
- Bachrach G, Ianculovici C, Naor R, Weiss EI. 2005. Fluorescence based measurements of *Fusobacterium nucleatum* coaggregation and of fusobacterial attachment to mammalian cells. *FEMS Microbiol. Lett.* 248:235–240. <http://dx.doi.org/10.1016/j.femsle.2005.05.055>.
- Shanitzki B, Hurwitz D, Smorodinsky N, Ganeshkumar N, Weiss EI. 1997. Identification of a *Fusobacterium nucleatum* PK1594 galactose-binding adhesin which mediates coaggregation with periopathogenic bacteria and hemagglutination. *Infect. Immun.* 65:5231–5237.
- Han YW, Wang X. 2013. Mobile microbiome: oral bacteria in extra-oral infections and inflammation. *J. Dent. Res.* 92:485–491. <http://dx.doi.org/10.1177/0022034513487559>.
- Wang X, Buhimschi CS, Témoins S, Bhandari V, Han YW, Buhimschi IA. 2013. Comparative microbial analysis of paired amniotic fluid and cord blood from pregnancies complicated by preterm birth and early-onset neonatal sepsis. *PLoS One* 8:e56131. <http://dx.doi.org/10.1371/journal.pone.0056131>.
- Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Taberner J, Baselga J, Liu C, Shivdasani RA, Ogino S, Birren BW, Huttenhower C, Garrett WS, Meyerson M. 2012. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* 22:292–298. <http://dx.doi.org/10.1101/gr.126573.111>.
- Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA, Holt RA. 2012. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* 22:299–306. <http://dx.doi.org/10.1101/gr.126516.111>.
- Wright WF, Shiner CN, Ribes JA. 2012. Lemierre syndrome. *South. Med. J.* 105:283–288. <http://dx.doi.org/10.1097/SMJ.0b013e31825581ef>.
- Tadepalli S, Stewart GC, Nagaraja TG, Narayanan SK. 2008. Human *Fusobacterium necrophorum* strains have a leukotoxin gene and exhibit leukotoxic activity. *J. Med. Microbiol.* 57:225–231. <http://dx.doi.org/10.1099/jmm.0.47598-0>.
- George WL, Kirby BD, Sutter VL, Citron DM, Finegold SM. 1981. Gram-negative anaerobic bacilli: their role in infection and patterns of susceptibility to antimicrobial agents. II. Little-known *Fusobacterium* species and miscellaneous genera. *Rev. Infect. Dis.* 3:599–626. <http://dx.doi.org/10.1093/clinids/3.3.599>.
- Ohkusa T, Yoshida T, Sato N, Watanabe S, Tajiri H, Okayasu I. 2009. Commensal bacteria can enter colonic epithelial cells and induce proinflammatory cytokine secretion: a possible pathogenic mechanism of ulcerative colitis. *J. Med. Microbiol.* 58:535–545. <http://dx.doi.org/10.1099/jmm.0.005801-0>.
- Adriaans B, Shah H. 1988. *Fusobacterium ulcerans* sp. nov. from tropical ulcers. *Int. J. Syst. Evol. Microbiol.* 37:447–448.
- Han YW, Ikegami A, Chung P, Zhang L, Deng CX. 2007. Sonoporation is an efficient tool for intracellular fluorescent dextran delivery and one-step double-crossover mutant construction in *Fusobacterium nucleatum*. *Appl. Environ. Microbiol.* 73:3677–3683. <http://dx.doi.org/10.1128/AEM.00428-07>.
- Fox JG, Yan LL, Dewhirst FE, Paster BJ, Shames B, Murphy JC, Hayward A, Belcher JC, Mendes EN. 1995. *Helicobacter bilis* sp. nov., a novel *Helicobacter* species isolated from bile, livers, and intestines of aged, inbred mice. *J. Clin. Microbiol.* 33:445–454.
- Fox JG, Ge Z, Whary MT, Erdman SE, Horwitz BH. 2011. *Helicobacter* hepatic infection in mice: models for understanding lower bowel inflammation and cancer. *Mucosal Immunol.* 4:22–30. <http://dx.doi.org/10.1038/mi.2010.61>.
- Dewhirst FE, Fox JG, Mendes EN, Paster BJ, Gates CE, Kirkbride CA, Eaton KA. 2000. “*Flexispira rappini*” strains represent at least 10 *Helicobacter* taxa. *Int. J. Syst. Evol. Microbiol.* 50:1781–1787.
- Tunnick R, Jackson L. 1925. *Bacillus gonidiaformans* (N.SP.): an hith-

- erto undescribed organism. *J. Infect. Dis.* 36:430–438. <http://dx.doi.org/10.1093/infdis/36.5.430>.
28. Eggerth AH, Gagnon BH. 1933. The bacteroides of human feces. *J. Bacteriol.* 25:389–413.
  29. Harris NM. 1905. *Bacillus mortiferus* (nov. spec.). *J. Exp. Med.* 6:519–547. <http://dx.doi.org/10.1084/jem.6.4-6.519>.
  30. Kapatral V, Anderson I, Ivanova N, Reznik G, Los T, Lykidis A, Bhattacharyya A, Bartman A, Gardner W, Grechkin G, Zhu L, Vasieva O, Chu L, Kogan Y, Chaga O, Goltsman E, Bernal A, Larsen N, D'Souza M, Walunas T, Pusch G, Haselkorn R, Fonstein M, Kyrpides N, Overbeek R. 2002. Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. *J. Bacteriol.* 184:2005–2018. <http://dx.doi.org/10.1128/JB.184.7.2005-2018.2002>.
  31. Karpathy SE, Qin X, Gioia J, Jiang H, Liu Y, Petrosino JF, Yerrapragada S, Fox GE, Haake SK, Weinstock GM, Highlander SK. 2007. Genome sequence of *Fusobacterium nucleatum* subspecies polymorphum—a genetically tractable *Fusobacterium*. *PLoS One* 2:e659. <http://dx.doi.org/10.1371/journal.pone.0000659>.
  32. Kapatral V, Ivanova N, Anderson I, Reznik G, Bhattacharyya A, Gardner WL, Mikhailova N, Lapidus A, Larsen N, D'Souza M, Walunas T, Haselkorn R, Overbeek R, Kyrpides N. 2003. Genome analysis of *F. nucleatum* sub spp. *vincentii* and its comparison with the genome of *F. nucleatum* ATCC 25586. *Genome Res.* 13:1180–1189. <http://dx.doi.org/10.1101/gr.566003>.
  33. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57: 81–91. <http://dx.doi.org/10.1099/ijs.0.64483-0>.
  34. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102: 2567–2572. <http://dx.doi.org/10.1073/pnas.0409727102>.
  35. Bolstad AI, Jensen HB, Bakken V. 1996. Taxonomy, biology, and periodontal aspects of *Fusobacterium nucleatum*. *Clin. Microbiol. Rev.* 9:55–71.
  36. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27:29–34. <http://dx.doi.org/10.1093/nar/27.20.e29>.
  37. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR, Sonnhammer EL, Bateman A. 2008. The Pfam protein families database. *Nucleic Acids Res.* 36:D281–D288. <http://dx.doi.org/10.1093/nar/gkm960>.
  38. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <http://dx.doi.org/10.1093/bioinformatics/bti610>.
  39. Han YW, Ikegami A, Rajanna C, Kawsar HI, Zhou Y, Li M, Sojar HT, Genco RJ, Kuramitsu HK, Deng CX. 2005. Identification and characterization of a novel adhesin unique to oral fusobacteria. *J. Bacteriol.* 187: 5330–5340. <http://dx.doi.org/10.1128/JB.187.15.5330-5340.2005>.
  40. Basavanna S, Khandavilli S, Yuste J, Cohen JM, Hosie AH, Webb AJ, Thomas GH, Brown JS. 2009. Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJ/HMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. *Infect. Immun.* 77:3412–3423. <http://dx.doi.org/10.1128/IAI.01543-08>.
  41. Omer S. 2013. Role of branched-chain amino acid transporters in *Staphylococcus aureus* virulence. Electronic Thesis and Dissertation Repository, University of Western Ontario, London, Ontario, Canada.
  42. Desvaux M, Khan A, Beatson SA, Scott-Tucker A, Henderson IR. 2005. Protein secretion systems in *Fusobacterium nucleatum*: genomic identification of type 4 piliation and complete type V pathways brings new insight into mechanisms of pathogenesis. *Biochim. Biophys. Acta* 1713:92–112. <http://dx.doi.org/10.1016/j.bbamem.2005.05.002>.
  43. Korotkov KV, Sandkvist M, Hol WG. 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nat. Rev. Microbiol.* 10:336–351. <http://dx.doi.org/10.1038/nrmicro2762>.
  44. Craig L, Pique ME, Tainer JA. 2004. Type IV pilus structure and bacterial pathogenicity. *Nat. Rev. Microbiol.* 2:363–378. <http://dx.doi.org/10.1038/nrmicro885>.
  45. Strom MS, Lory S. 1993. Structure-function and biogenesis of the type IV pili. *Annu. Rev. Microbiol.* 47:565–596. <http://dx.doi.org/10.1146/annurev.mi.47.100193.003025>.
  46. Wall D, Kaiser D. 1999. Type IV pili and cell motility. *Mol. Microbiol.* 32:1–10. <http://dx.doi.org/10.1046/j.1365-2958.1999.01339.x>.
  47. Banerji S, Flieger A. 2004. Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? *Microbiology* 150:522–525. <http://dx.doi.org/10.1099/mic.0.26957-0>.
  48. Uliana SR, Goyal N, Freymüller E, Smith DF. 1999. Leishmania: over-expression and comparative structural analysis of the stage-regulated meta 1 gene. *Exp. Parasitol.* 92:183–191. <http://dx.doi.org/10.1006/expr.1999.4410>.
  49. Sasso S, Ramakrishnan C, Gamper M, Hilvert D, Kast P. 2005. Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *FEBS J.* 272:375–389. <http://dx.doi.org/10.1111/j.1742-4658.2004.04478.x>.
  50. Degrassi G, Devescovi G, Bigirimana J, Venturi V. 2010. *Xanthomonas oryzae* pv. *oryzae* XKK.12 contains an AroQgamma chorismate mutase that is involved in rice virulence. *Phytopathology* 100:262–270. <http://dx.doi.org/10.1094/PHYTO-100-3-0262>.
  51. Kaplan CW, Lux R, Haake SK, Shi W. 2009. The *Fusobacterium nucleatum* outer membrane protein RadD is an arginine-inhibitable adhesin required for inter-species adherence and the structured architecture of multispecies biofilm. *Mol. Microbiol.* 71:35–47. <http://dx.doi.org/10.1111/j.1365-2958.2008.06503.x>.
  52. Kaplan CW, Ma X, Paranjpe A, Jewett A, Lux R, Kinder-Haake S, Shi W. 2010. *Fusobacterium nucleatum* outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. *Infect. Immun.* 78: 4773–4778. <http://dx.doi.org/10.1128/IAI.00567-10>.
  53. Kerfeld CA, Heinhorst S, Cannon GC. 2010. Bacterial microcompartments. *Annu. Rev. Microbiol.* 64:391–408. <http://dx.doi.org/10.1146/annurev.micro.112408.134211>.
  54. Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785–786. <http://dx.doi.org/10.1038/nmeth.1701>.
  55. Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA. 2006. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* 14:264–270. <http://dx.doi.org/10.1016/j.tim.2006.04.005>.
  56. Langille MG, Brinkman FS. 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* 25:664–665. <http://dx.doi.org/10.1093/bioinformatics/btp030>.
  57. Kolenbrander PE. 2000. Oral microbial communities: biofilms, interactions, and genetic systems. *Annu. Rev. Microbiol.* 54:413–437. <http://dx.doi.org/10.1146/annurev.micro.54.1.413>.
  58. Edwards AM, Grossman TJ, Rudney JD. 2006. *Fusobacterium nucleatum* transports noninvasive *Streptococcus cristatus* into human epithelial cells. *Infect. Immun.* 74:654–662. <http://dx.doi.org/10.1128/IAI.74.1.654-662.2006>.
  59. Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. 2012. Beyond junk—variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes (Basel)* 3:461–480. <http://dx.doi.org/10.3390/genes3030461>.
  60. Lorestani A, Sheiner L, Yang K, Robertson SD, Sahoo N, Brooks CF, Ferguson DJ, Striepen B, Gubbels MJ. 2010. A *Toxoplasma MORN1* null mutant undergoes repeated divisions but is defective in basal assembly, apicoplast division and cytokinesis. *PLoS One* 5:e12302. <http://dx.doi.org/10.1371/journal.pone.0012302>.
  61. Takeshima H, Komazaki S, Nishi M, Iino M, Kangawa K. 2000. Junctophilins: a novel family of junctional membrane complex proteins. *Mol. Cell* 6:11–22. [http://dx.doi.org/10.1016/S1097-2765\(05\)00005-5](http://dx.doi.org/10.1016/S1097-2765(05)00005-5).
  62. Liang Q, Lu X, Jiang L, Wang C, Fan Y, Zhang C. 2010. EMB1211 is required for normal embryo development and influences chloroplast biogenesis in *Arabidopsis*. *Physiol. Plant* 140:380–394. <http://dx.doi.org/10.1111/j.1399-3054.2010.01407.x>.
  63. Choi YJ, Hwang KC, Park JY, Park KK, Kim JH, Park SB, Hwang S, Park H, Park C, Kim JH. 2010. Identification and characterization of a novel mouse and human MOP1 gene containing MORN-motif protein in testis. *Theriogenology* 73:273–281. <http://dx.doi.org/10.1016/j.theriogenology.2009.09.010>.
  64. Mikami K, Saavedra L, Sommarin M. 2010. Is membrane occupation and recognition nexus domain functional in plant phosphatidylinositol phosphate kinases? *Plant Signal. Behav.* 5:1241–1244. <http://dx.doi.org/10.4161/psb.5.10.12922>.
  65. Oh SH, Cheng G, Nuessen JA, Jajko R, Yeater KM, Zhao X, Pujol C, Soll DR, Hoyer LL. 2005. Functional specificity of *Candida albicans* Als3p proteins and clade specificity of ALS3 alleles discriminated by the number

- of copies of the tandem repeat sequence in the central domain. *Microbiology* 151:673–681. <http://dx.doi.org/10.1099/mic.0.27680-0>.
66. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37:986–990. <http://dx.doi.org/10.1038/ng1618>.
67. Mittl PR, Schneider-Brachert W. 2007. Sel1-like repeat proteins in signal transduction. *Cell. Signal.* 19:20–31. <http://dx.doi.org/10.1016/j.cellsig.2006.05.034>.
68. Qamra R, Prakash P, Aruna B, Hasnain SE, Mande SC. 2006. The 2.15-Å crystal structure of *Mycobacterium tuberculosis* chorismate mutase reveals an unexpected gene duplication and suggests a role in host-pathogen interactions. *Biochemistry* 45:6997–7005. <http://dx.doi.org/10.1021/bi0606445>.
69. Chew J, Zilm PS, Fuss JM, Gully NJ. 2012. A proteomic investigation of *Fusobacterium nucleatum* alkaline-induced biofilms. *BMC Microbiol.* 12:189. <http://dx.doi.org/10.1186/1471-2180-12-189>.