



Article

Comparative Genome Analyses Reveal the Genomic Traits and Host Plant Adaptations of *Flavobacterium akiainvivens* IK-1^T

Xuehua Wan ^{1,2,3}

¹ TEDA Institute of Biological Sciences and Biotechnology, Nankai University, TEDA, Tianjin 300071, China; xuehua.wan@hotmail.com

² The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, Nankai University, Tianjin 300071, China

³ Tianjin Key Laboratory of Microbial Functional Genomics, Nankai University, Tianjin 300071, China

Received: 30 August 2019; Accepted: 2 October 2019; Published: 3 October 2019



Abstract: The genus *Flavobacterium* contains a large group of commensal bacteria identified in diverse terrestrial and aquatic habitats. We compared the genome of a new species *Flavobacterium akiainvivens* IK-1^T to public available genomes of *Flavobacterium* species to reveal the genomic traits and ecological roles of IK-1^T. Principle component analysis (PCA) of carbohydrate-active enzyme classes suggests that IK-1^T belongs to a terrestrial clade of *Flavobacterium*. In addition, type 2 and type 9 secretion systems involved in bacteria-environment interactions were identified in the IK-1^T genome. The IK-1^T genome encodes eukaryotic-like domain containing proteins including ankyrin repeats, von Willebrand factor type A domain, and major royal jelly proteins, suggesting that IK-1^T may alter plant host physiology by secreting eukaryotic-like proteins that mimic host proteins. A novel two-component system *FaRpfC-FaYpdB* was identified in the IK-1^T genome, which may mediate quorum sensing to regulate global gene expressions. Our findings suggest that comparative genome analyses of *Flavobacterium* spp. reveal that IK-1^T has adapted to a terrestrial niche. Further functional characterizations of IK-1^T secreted proteins and their regulation systems will shed light on molecular basis of bacteria-plant interactions in environments.

Keywords: comparative genomics; pan genome; synteny; evolution direction; plant-associated bacteria

1. Introduction

The genus *Flavobacterium* belongs to the phylum *Bacteroidetes* and has been known for almost one hundred years, which was proposed in the first edition of *Bergey's Manual of Determinative Bacteriology* in 1923 [1]. More than 130 species have been identified in the genus *Flavobacterium*, which are isolated from terrestrial and aquatic environments. Three species, including *F. psychrophilum*, *F. columnare*, and *F. branchiophilum*, can cause severe fish diseases worldwide. However, most *Flavobacterium* spp. are commensal bacteria isolated with greater abundances from rivers and oceans [2]. Aquatic *Flavobacteria* played important roles in algae decomposition [2–4], while terrestrial *Flavobacteria* are adapted to plant carbohydrate metabolism and may promote plant growth and development [5,6].

Microorganisms are natural resources of chemicals such as biopigments and biosurfactants that have promising potentials in biotechnological and industrial applications. The genus *Flavobacterium* yields carotenoids, a huge family of pigments, displaying various colors, including yellow, orange, pink, red, and brown. Several species of *Flavobacterium* are able to degrade organic contaminants, such as pentachlorophenol, nylon oligomers, polyaromatics, and pesticides [7–12]. Screening for biosurfactant-producing microorganisms identified one *Flavobacterium* strain as a

biosurfactant-producing microbe [13]. Structure analyses revealed the *Flavobacterium* strain produced at least 37 flavolipids as a new class of biosurfactants [7]. Identification and characterization of *Flavobacterium* species and strains may help reveal biotechnological applications of this genus.

F. akiainvivens IK-1^T, isolated from decaying shrub *Wikstroemia oahuensis* in Hawaii, has been formally described with a validly published name [14]. IK-1^T was Gram-negative and non-motile, and its gliding motility was not observed [14]. A colony formed by IK-1^T showed translucent and off-white [14]. Based on the comparisons of the 16S rRNA gene sequences, the nearest phylogenetic neighbors of IK-1^T were *F. rivuli* WB 3.3-2^T and *F. subsaxonicum* WB 4.1-42^T, which were isolated from a hard water rivulet of North Germany and characterized to be non-motile and non-gliding [14–16]. Analysis of the *F. rivuli* genome has revealed 151 genes encoding carbohydrate-active enzymes (CAZy) and 177 genes encoding peptidases, implying a role of *F. rivuli* in the decomposition potential of organic matter in the environment [15]. In addition, although *F. rivuli* was isolated from stream water, it was affiliated with genomes of terrestrial clade-affiliated *Flavobacteria* [5]. A fingerprinting-based study on flavobacterial community structure also supported that *F. rivuli* is soil-derived [17].

The assembled genome sequence of IK-1^T was briefly reported [18]. However, the genome has not been analyzed in details yet. In this study, we compared the IK-1^T genome to genomes of fifty-eight members of its related *Flavobacterium* species and characterized IK-1^T at a whole genome level. Our findings reveal that IK-1^T belongs to a terrestrial clade of *Flavobacteria* and has adapted to plant host.

2. Results

2.1. Genome Features of IK-1^T

The assembled genome sequence of IK-1^T contains a single chromosome spanning 4.5 Mb [18]. To provide deep analyses of this new species, we annotated the IK-1^T genome in details. Prokka predicted 50 tRNA, and 4067 protein-coding genes (Table S1). A total of 1978 proteins were annotated with functions, and 2089 proteins were annotated as hypothetical proteins with unknown functions. We further identified 20,351 bp of repeats in the genome using RepeatMasker (Table S2). Most of the repeats belong to simple repeats (9336 bp) and low complexity regions (2258 bp). The general genome features of IK-1^T are shown in Figure 1.

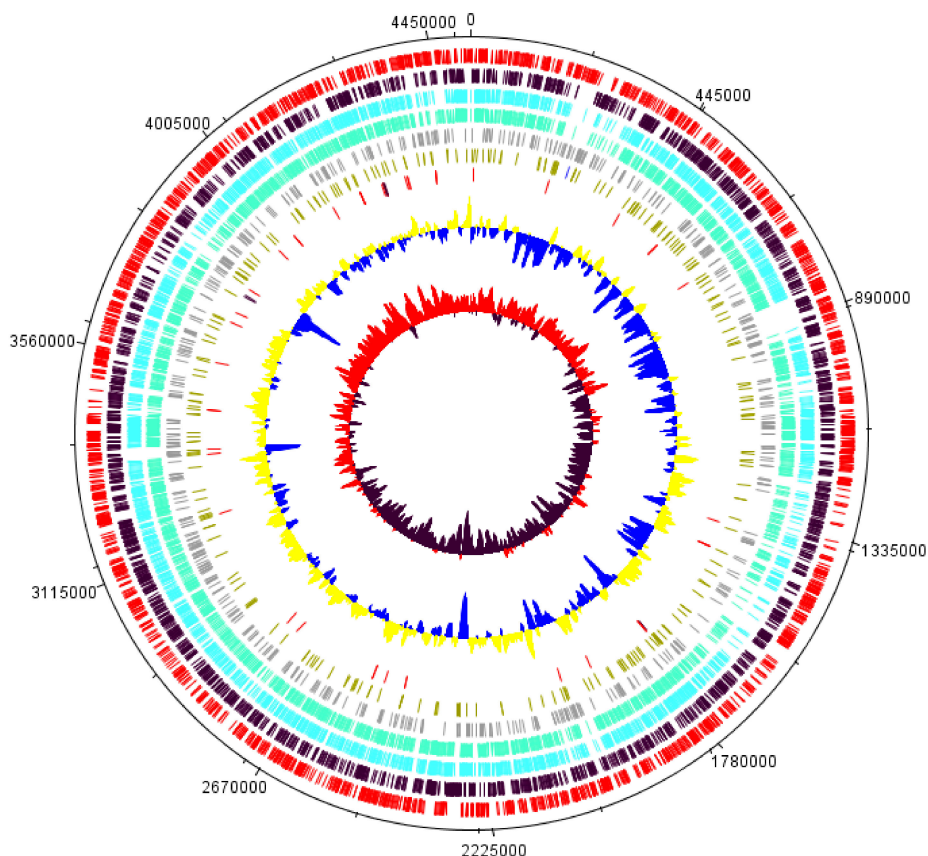


Figure 1. Genome features of *F. akiainvivens* IK-1^T. From the outside to the center: 1. tick marks, 2. coding sequences in forward strand (red), 3. coding sequences in reverse strand (black), 4. IK-1^T-*F. subsaxonicum* orthologs (light blue), 5. IK-1^T-*F. rivuli* orthologs (green), 6. IK-1^T singleton (grey), 7. repeats (light green) and CRISPR (blue), 8. tRNA (red) and rRNA (black), 9. GC plot, below average (43.8%) (blue) and above average (yellow), 10. GC skew $[(G - C)/(G + C)]$, below average (0.028%) (black) and above average (red).

2.2. IK-1^T is a Member of Terrestrial Clade-Affiliated *Flavobacteria*

Comparative gene analysis suggested that the terrestrial and aquatic *Flavobacteria* can be hierarchically clustered into distinct clades [5]. The genomes of members in the terrestrial clade-affiliated *Flavobacteria* encode glycoside hydrolase (GH) families GH78 and GH106 for utilization of rhamnogalacturonan in plant cell wall [5]. To investigate whether IK-1^T is a terrestrial bacterium associated with plant, we identified the gene families of carbohydrate-active enzymes (CAZy) in the genome, according to the previously reported method [5]. We surveyed 16 CAZy families including GH, carbohydrate esterases (CE), polysaccharide lyases (PL), and carbohydrate-binding module (CBM). The abundance and distribution of CAZy families in IK-1^T indicated that it is a terrestrial bacterium, compared to those in seven terrestrial and eight aquatic *Flavobacterium* spp. (Table S3). We found that the IK-1^T genome encodes GH78 and GH106 genes, suggesting that IK-1^T belongs to the terrestrial clade of *Flavobacteria*. Likewise, 17 GH43 and two GH27 involved in metabolism of glucans containing arabinose were found in the IK-1^T genome, which are associated with the terrestrial clade-affiliated *Flavobacteria*. Principle component analysis (PCA) of the CAZy gene families further statistically clustered IK-1^T with the terrestrial *Flavobacterium* spp., indicating that IK-1^T has adapted to a terrestrial environment (Figure 2).

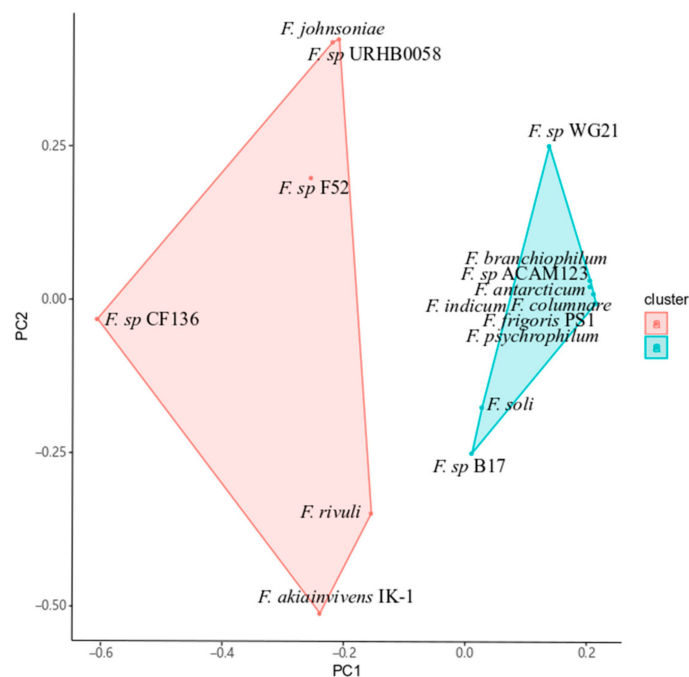


Figure 2. Two dimensional scatter plot based on carbohydrate-active enzyme (CAZY) classes in the terrestrial and aquatic flavobacteria. The matrix representing the presence and absence of the CAZY classes was imported into the R program. IK-1^T was clustered with members from the terrestrial clade of *Flavobacteria* (red). Members from the aquatic clade of *Flavobacteria* (blue) was clustered separately. The terrestrial and aquatic clades of *Flavobacteria* showed distinct distances in this two dimensional plot. The percentages of PC1 and PC2 are 84.9% and 8.49%, respectively. Red: terrestrial *flavobacteria*, blue: aquatic *flavobacteria*.

2.3. An Open Pan-Genome of *Flavobacteria*

We downloaded the genome sequences of 58 members of the genus *Flavobacterium* and re-annotated them using Prokka. The detailed information was summarized in Supplementary Table S4. A total of 215,727 protein sequences were loaded for all-vs.-all BLASTP. We removed any alignments at or below 80% of aligned length and 30% identity. We performed OrthoMCL analysis to identify 13,583 orthologous groups (Table S5). Only seven orthologous groups were conserved in the 59 genomes of *Flavobacterium* species. Because the genomes of three strains (SCGC AAA536-P05, MS220-5C, and JGI 0001001-D01) contained fewer than 2000 genes, we excluded them and found that 541 orthologous groups were conserved in the 56 genomes of *Flavobacteria*. These results indicated that gene gain-and-loss events occurred in *Flavobacteria*, which contribute to the genome dynamics and biodiversity of *Flavobacteria*.

The pan-genome of *Flavobacteria* contained 34,769 genes (Figure 3a). By applying power law regression using gene number medians, the gene accumulation curve fitted Heaps' law and indicated an open pan-genome which is far from saturation ($\alpha = 0.57$, $R^2 = 0.998$) [19,20]. We recorded the presence and absence of the orthologous groups in a binary matrix. PCA analysis of the presence and absence of the orthologous groups showed close relationships between IK-1^T and *F. rivuli*/*F. subsaxonicum* (Figure 3b), consistent with the comparison result of 16S rRNA gene sequences. Likewise, phylogenetic analysis using concatenated single-ortholog groups also clustered IK-1^T with *F. rivuli* and *F. subsaxonicum* (Figure S1). These data suggest that *Flavobacteria* occupy diversified niches that lead to an open and dynamic pan-genome. In addition, the nearest phylogenetic neighbors of IK-1^T are *F. rivuli* and *F. subsaxonicum*.

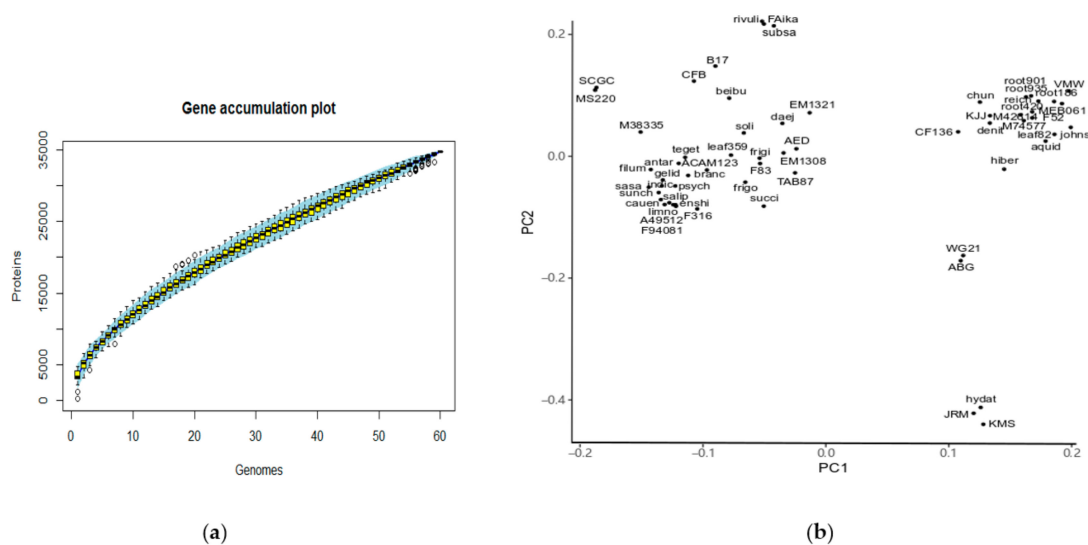


Figure 3. Pan-genome of *Flavobacterium*. (a) Gene accumulation plot of pan-genome of *Flavobacterium*. Each box represents the values for any possible combination of gene clusters in the respective number of genomes. (b) Two dimensional scatter plot based on the presence and absence of gene families in the *Flavobacterium* genus. The binary matrix representing the presence and absence of gene families was imported into the R program. IK-1^T showed close relationship to *F. rivuli* and *F. subsaxonicum*, and showed distinct distances to other species in the *Flavobacterium* genus. The percentages of PC1 and PC2 in Figure 3b are 15.95% and 4.88%, respectively.

To reveal species-specific genes in IK-1^T, orthoMCL predicted 596 singletons in the IK-1^T genome, and 491 singletons (82.4%) were annotated as hypothetical proteins. The 105 singletons were annotated with functions including hydrolases, proteases, transferases, oxidoreductases, histidine kinases, transcriptional regulators/activators, transmembrane proteins, and transporters. The singletons were re-annotated using InterProScan and plotted for Gene Ontology (GO) annotation (Figure S2). These IK-1^T-specific genes contribute to the large size of the open pan-genome of *Flavobacterium*.

2.4. Synteny Between IK-1^T and *F. rivuli*/*F. subsaxonicum*

Genome rearrangement and synteny provide evolutionary relationships between genomes. We used SyMAP to determine syntenic blocks in IK-1^T-*F. rivuli* and IK-1^T-*F. subsaxonicum*. Comparison of the genomes of IK-1^T and *F. rivuli* revealed eight syntenic blocks, while comparison of the genomes of IK-1^T and *F. subsaxonicum* revealed eleven syntenic blocks, spanning 1,614,043 bp (36%) and 2,026,781 bp (45%) in the IK-1^T genome, respectively (Figure 4a,c, Table S6). Synteny blocks and dot plots were shown in Figure 4b and d. Majority of the genome regions were not in the syntenic blocks. By contrast, comparison of the genomes of *F. rivuli* and *F. subsaxonicum* revealed sixty-eight syntenic blocks, covering 74% and 75% of the genome sequences of *F. rivuli* and *F. subsaxonicum*, respectively (Figures S3 and S4). These findings suggest that the level of synteny between *F. rivuli* and *F. subsaxonicum* is much higher than that between IK-1^T and *F. rivuli*/*F. subsaxonicum*. In addition, we compared the genomes of IK-1^T and an outgroup member *Elizabethkingia meningoseptica* for synteny analysis. *E. meningoseptica* was originally described as *Flavobacterium meningoseptica*. Both *Elizabethkingia* and *Flavobacterium* belong to the Family *Flavobacteriaceae*. The genome of *E. meningoseptica* (GenBank: CP016378.1) contains 4 Mb sequences and encodes 3647 genes. Only one syntenic block was identified between the genomes of *E. meningoseptica* (2%) and IK-1^T (5%) (Figures S5 and S6). These data suggest that extreme low level of synteny was found between IK-1^T and the outgroup member. Taken together, the above data suggest that dramatic rearrangements occurred in the genomes of IK-1^T and *F. rivuli* or *F. subsaxonicum* after they diverged from their most recent common ancestor (MRCA). Moreover, extensive synteny

and conserved gene orders exist between these genomes. This synteny may be due to incomplete dislocation of gene order or selective aggregation of genes under evolutionary pressures.

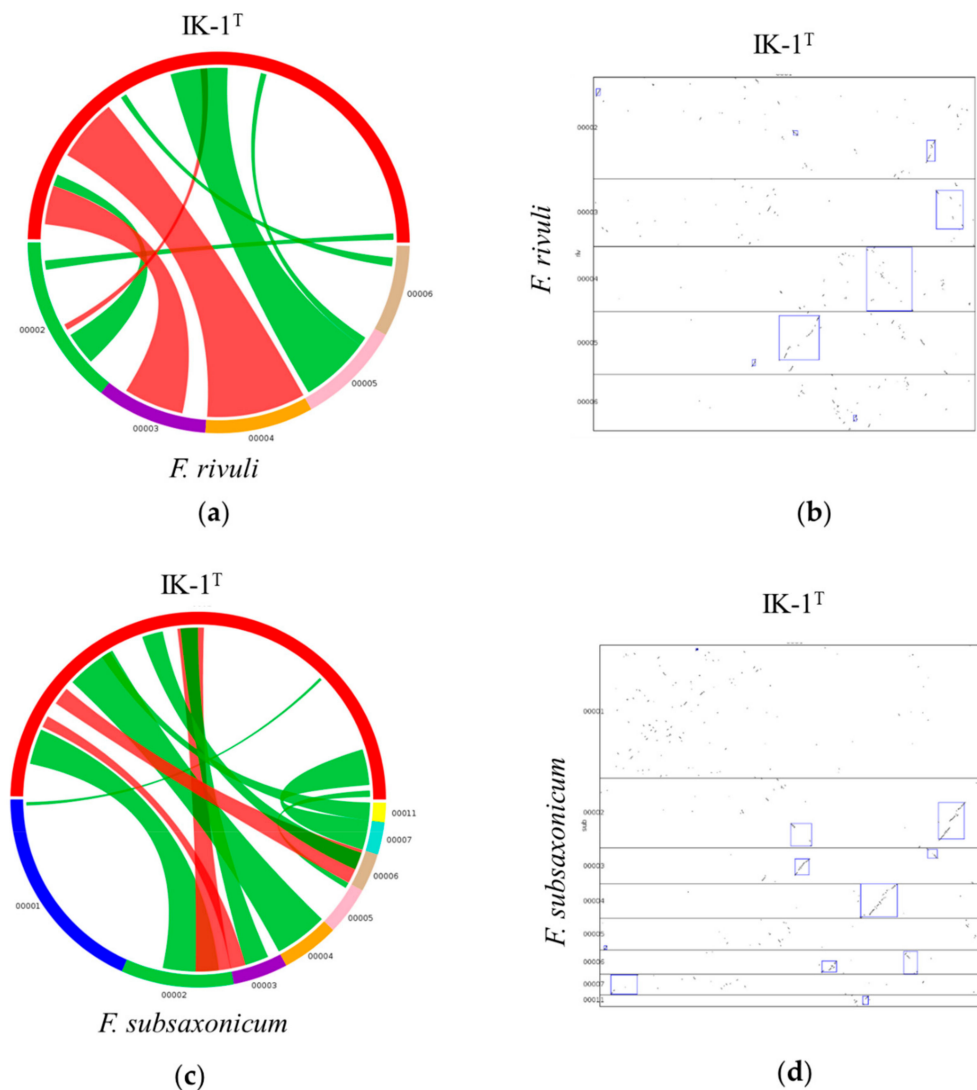


Figure 4. Syntenic blocks shared between IK-1^T and its nearest phylogenetic neighbors. (a) Syntenic blocks shared between IK-1^T and *F. rivuli*. (b) Syntenic regions shared between IK-1^T and *F. rivuli* visualized by dotplot. (c) Syntenic blocks shared between IK-1^T and *F. subsaxonicum*. (d) Syntenic regions shared between IK-1^T and *F. subsaxonicum* visualized by dotplot. Red: forward synteny, green: reverse synteny. Red in the circle: IK-1^T, other colors in the circle: scaffolds of *F. rivuli* (a) or *F. subsaxonicum* (c).

2.5. Evolution Direction in the IK-1^T Genome

We next investigated the evolution direction of genes in IK-1^T. We retrieved pairwise orthologs from IK-1^T-*F. rivuli* (2318 orthologous gene pairs), IK-1^T-*F. subsaxonicum* (2353 orthologous gene pairs), IK-1^T-*F. psychrophilum* (1463 orthologous gene pairs), IK-1^T-*F. johnsoniae* (2016 orthologous gene pairs), IK-1^T-*F. antarcticum* (psychrotolerant bacterium, 1645 orthologous gene pairs), IK-1^T-*F. beibuense* (marine bacterium, 2091 orthologous gene pairs), IK-1^T-*F. sp. Leaf359* (1892 orthologous gene pairs), IK-1^T-*F. sp. root186* (1993 orthologous gene pairs), and IK-1^T-*F. soli* (1893 orthologous gene pairs). Although these selected *Flavobacterium* species were isolated from different terrestrial or aquatic environments, the peaks of kernel density plots of Ka/Ks values (the ratio of nonsynonymous substitution rate (Ka) to synonymous substitution rate (Ks)) were centered less than 0.1, indicating a purifying selection occurred (Figure 5a,b,

data not shown for IK-1^T-*F. psychrophilum*, IK-1^T-*F. johnsoniae*, IK-1^T-*F. antarcticum*, IK-1^T-*F. sp.* Leaf359, IK-1^T-*F. sp.* root186, and IK-1^T-*F. soli*).

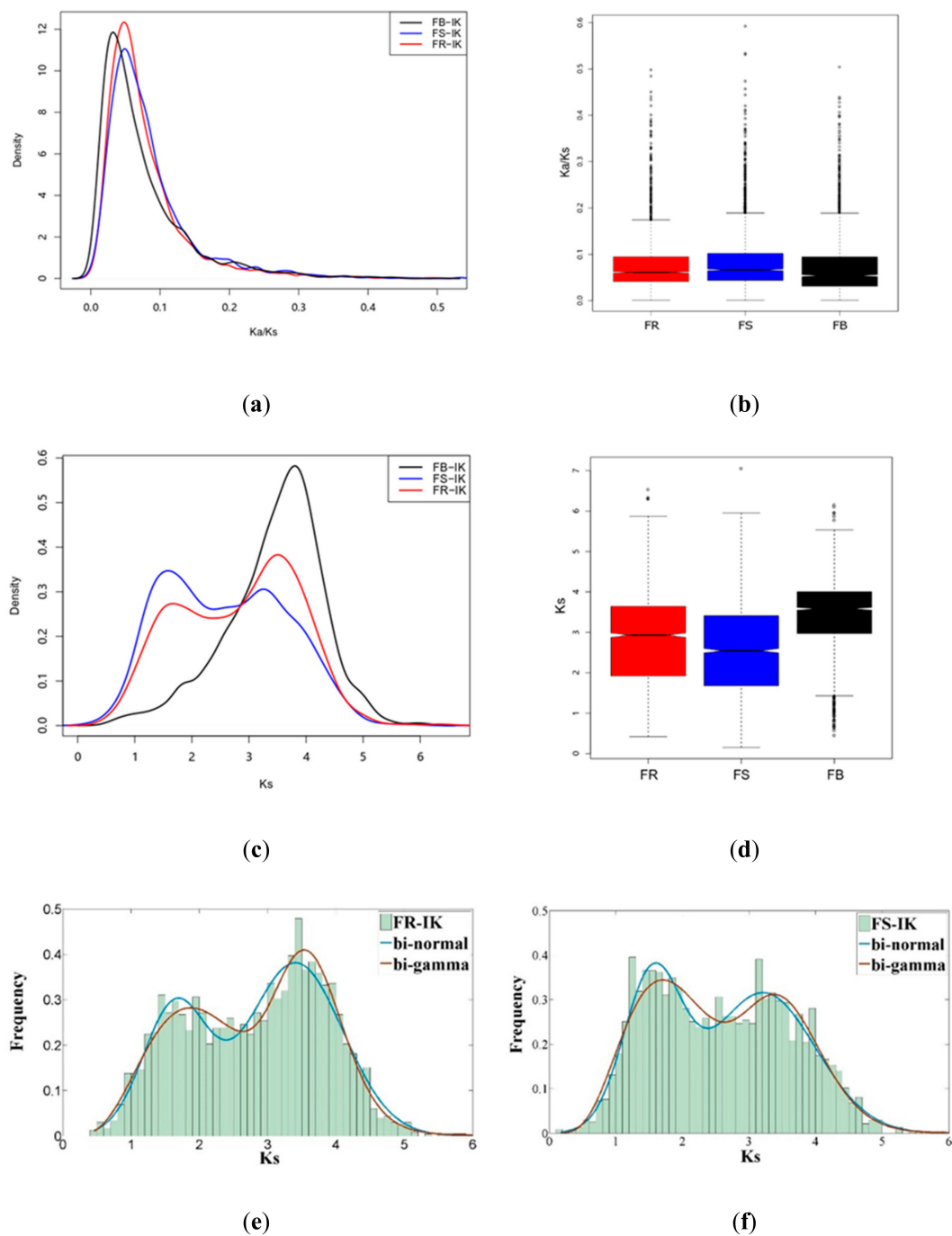


Figure 5. The comparison of ortholog sequence divergence between IK-1^T and its phylogenetic neighbors. (a) The distribution of Ka/Ks values of orthologous genes between IK-1^T and its phylogenetic neighbors. (b) The comparison of Ka/Ks values of orthologous genes between IK-1^T and its phylogenetic neighbors by boxplot. (c) The distribution of Ks values of orthologous genes between IK-1^T and its phylogenetic neighbors. (d) The comparison of Ks values of orthologous genes between IK-1^T and its phylogenetic neighbors by boxplot. (e) Fitting of Ks frequency of orthologous genes between IK-1^T and *F. rivuli*. (f) Fitting of Ks frequency of orthologous genes between IK-1^T and *F. subsaxonicum*. IK: *F. akiainvivens* IK-1^T, FB: *F. beibuense* F44-8, FS: *F. subsaxonicum* DSM 21790, FR: *F. rivuli* DSM 21788.

Intriguingly, the density plots of Ks distributions of orthologous genes between IK-1^T and *F. rivuli* or *F. subsaxonicum* displayed two peaks (Figure 5c,d). This bimodal pattern of density plot of Ks values of orthologous genes fitted to a bi-normal or bi-gamma distribution (Figure 5e,f). In contrast, the density plots of Ks distributions of pairwise orthologous genes in IK-1^T-*F. beibuense*, IK-1^T-*F. psychrophilum*, IK-1^T-*F. johnsoniae*, or IK-1^T-*F. antarcticum*, formed a single peak (Figure 5c,d, data not shown for IK-1^T-*F. psychrophilum*, IK-1^T-*F. johnsoniae*, IK-1^T-*F. antarcticum*, IK-1^T-*F. sp. Leaf359*, IK-1^T-*F. sp. root186*, and IK-1^T-*F. soli*). Annotations of the orthologous genes with those lower Ks values ($K_s < 2$) identified the genes involved in central dogma, such as 30S and 50S ribosomal proteins, DNA polymerases, and DNA-directed RNA polymerases. These data suggest that more flexible relaxed molecular clock model may apply for bacteria with close phylogenetic relationships.

2.6. Gliding Motility and Type 9 Secretion System in IK-1^T

To examine the genes absent from the gliding system in the IK-1^T genome, we searched the IK-1^T genome sequence using both BLASTP and TBLASTN and identified 20 genes involved in gliding motility. We identified the presence of orthologs of 17 gliding genes in IK-1^T (Table S7), and found the absence of orthologs of three adhesion-like protein coding genes (*remA*, *remB*, and *sprB*). In addition, the IK-1^T genome also encodes type 9 secretion system (T9SS) that are components of gliding motility system (Table S7).

2.7. Type 2 Secretion System in IK-1^T

We next examined the type 2 secretion system in IK-1^T. IK-1^T encodes the core components of a Sec-SRP (signal recognition particle) secretion pathway, containing SecD/F, SecY, YajC, YidC, SecA, FtsY, and SRP54 ffh. We confirmed that IK-1^T lacks SecE, SecG, SecM, and SecB by performing BLASTP and TBLASTN analyses. SecY, SecE, and SecG form a transporter complex across the inner membrane, but SecE and SecG are not obligatorily required for their translocation activity [21,22]. In addition, chaperone SecB and secretion monitor SecM are not obligatorily required for the sec-SRP pathway [22]. Thus, IK-1^T encodes the essential components of Sec-SRP pathway for a potentially functional type 2 secretion system.

Peptidases can recycle unfolded or misfolded proteins, remove signal peptides, and function as exotoxins when they are secreted. We used the MEROPS database 10.0 [23] to search for genes encoding peptidase in IK-1^T, and identified 159 peptidases and their inhibitors grouped in 66 clans. These data reveal that IK-1^T encodes fewer peptidases and inhibitors than *F. rivuli* that encodes 177 peptidases and inhibitors [15].

Next, we examined the signal peptides in proteins that may be exported outside IK-1^T. PrediSi server and SignalP 4.1 server, respectively, predicted 1310 and 754 proteins containing signal peptide cleavage sites. A total of 720 proteins were predicted to contain signal peptide cleavage sites annotated by both programs. Among these proteins, 15 proteins were predicted to contain transmembrane regions, including 10 hypothetical proteins, collagenase precursor, lipid A 1-phosphatase, TonB-dependent receptor plug domain protein, Y_Y_Y domain protein, and carboxylesterase N1hH.

2.8. Eukaryotic-Like Domain Containing Proteins in IK-1^T

Because bacterial homologues of eukaryotic signaling domains were found in a number of plant pathogenesis-related proteins [24], we further searched the IK-1^T genome for eukaryotic-like domains using the effective eukaryotic-like domains (EffectiveELD) database [25]. We predicted 115 proteins containing eukaryotic-like domains (ELD) ($Z\text{-score} \geq 4$). Thirty-seven proteins containing ELDs also contained T2SS signal peptide, suggesting that they may be secreted by T2SS. Besides GH families and redoxins, ELD proteins encoded in the IK-1^T genome include three ankyrin repeat (ANK) containing

proteins, four von Willebrand factor type A domain (vWFA) containing proteins, and three major royal jelly proteins (MRJP), which are mainly enriched in symbiont or pathogenic bacteria.

ANK-containing proteins have been identified in eukaryotes, eukaryotic viruses, and various bacterial taxa [26]. Bacterial ANK-containing proteins can be secreted into host cell, and then hijack host immune responses or modulate global gene transcriptions by mimicking host proteins [26]. IK-1^T has three genes encoding ANK-containing proteins. OrthoMCL clustered one IK-1^T ANK-containing protein as singleton, and clustered the other two ANK-containing proteins into two different orthologous gene families. The neighbor-joining tree of *Flavobacterium* ANK-containing proteins was constructed using MEGA X based on Dayhoff model and pairwise matching (Figure S7). Both IK-1^T ANK-containing proteins were closely clustered with those from *F. rivuli* and *F. subsaxonicum*. A BLASTP search of the NCBI non-redundant sequence database with IK-1^T ANK-containing singleton as query revealed significant similarities with the ANK-containing protein from soil bacterium *Solitalea Canadensis*, suggesting that this singleton gene may be acquired by horizontal gene transfer (HGT).

Von Willebrand factor in metazoans mediates ligand-binding events [27]. Although functions of bacterial vWFA domains remain largely unknown, studies suggest that bacterial vWFA domains may be secreted and have similar ligand and/or protein-binding functions to their eukaryotic homologues [24]. The IK-1^T genome encodes four vWFA domain-containing proteins that were clustered into four different orthologous gene families in *Flavobacterium*. Three vWFA domain-containing proteins in IK-1^T were clustered with those from *F. rivuli* and *F. subsaxonicum*, but one IK-1^T vWFA domain-containing protein was clustered with that from *F. tegetincola* (Figure S8). They contain additional domains such as a BatA domain or a secreted endopeptidase domain that may contribute to IK-1^T pathogenesis.

Members of Major Royal Jelly Protein (MRJP) family in insects play diverse roles in physiology, development, and reproductive maturation [28]. MRJP domain-containing proteins were scattered in various bacterial species, suggesting that MRJP family genes with important roles for niche adaptation may be retained in bacterial genomes. IK-1^T genome encodes three MRJP genes and OrthoMCL clustered them into three different orthologous gene families in *Flavobacterium*. Two IK-1^T MRJP proteins containing a sole MRJP domain were clustered with MRJPs from *F. rivuli* and/or *F. subsaxonicum* (Figure S9). The third IK-1^T MRJP protein with an additional N-terminal cupin domain was clustered with the MRJP from *F. hydatis*.

2.9. Quorum Sensing Mechanism

Unicellular bacteria can produce and/or respond to signal molecules, named as autoinducers, to cooperate their gene expressions via quorum sensing pathway. However, the quorum sensing system hasn't been systematically proposed in the genus *Flavobacterium*. To determine the quorum sensing system in IK-1^T, we performed BLASTP to search the IK-1^T proteome using Rpf protein sequences from plant-pathogenic *Xanthomonas campestris* as queries. We identified homologous proteins of PrfB, LysS, RpfG, RpfC, RpfF, RpfB, and PrfA in IK-1^T. Homologous proteins of RpfE or RpfH were not found by either BLASTP or TBLASTN. Unlike in plant pathogen *Xanthomonas campestris* and *Xylella fastidiosa* where these genes form an *rpf* operon [29], most of these homologous proteins were distributed widely in the IK-1^T genome (Table S8). The homologs of RpfG and RpfC still locate adjacent with 212 bp apart and probably form an operon. The homolog of RpfC, *FaRpfC*, contains HisKA, HATPase_c, REC, and HPT domains in the same order as the RpfC in *X. campestris* (Figure 6a). However, *FaRpfC* contains an additional N-terminal CHASE3 (Cyclases/Histidine kinases Associated Sensory Extracellular) domain, which is found in transmembrane receptor from bacteria but absent in *X. campestris* RpfC (Figure 6a). Interestingly, the homolog of RpfG with the best e-value is a transcriptional regulatory protein YpdB, *FaYpdB*. Unlike RpfG which contains a C-terminal diguanylate cyclase domain and regulates bacterial homeostasis of the second messenger c-di-GMP, *FaYpdB* contains an N-terminal REC domain and a C-terminal LytTR domain, a winged helix-turn-helix (wHTH) which receives the phosphorelay signal to regulate DNA transcription (Figure 6b). This putative two-component sensing signaling pathway, *FaRpfC-FaYpdB*, is consistent with the lacking of putative diguanylate

cyclase or c-di-GMP phosphodiesterase in IK-1^T, which is confirmed by our BLASTP and TBLASTN analyses. When performing BLASTP using *FaRpfC* and *FaYpdB* as the query sequences to search against the NCBI non-redundant sequence database, the ranked top 5 homologous proteins were found in *Chryseobacterium* spp. or *Epilithonimonas* spp., which belong to the Family *Flavobacteriaceae*. This provides evidence that this two-component sensing signaling system may have evolved convergently between IK-1^T and species outside of the genus *Flavobacterium*.

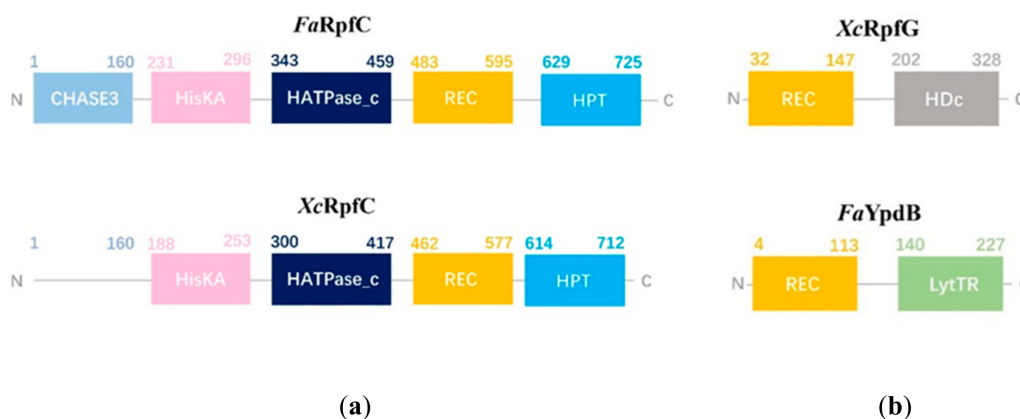


Figure 6. Comparison of domain arrangements of RpfC (a) and its cognate response regulator (b) in IK-1^T and *Xanthomonas campestris*.

3. Discussion

Bacteria have been playing important roles in altering environments on Earth for a billion years. For the past decade, massive parallel sequencing technologies and bioinformatic analyses have made tremendous progresses in surveying potential new microbial species. As many as one trillion (10^{12}) microbial species have been predicted to exist on Earth, suggesting a much greater biodiversity of microbes than anticipated [30]. Identifications of the isolated cultures of new bacterial species by genome sequencing and physiological studies will help improve our understanding on bacterial ecological roles and the genetic basis of bacterial evolution in the tree of life. Specifically, studies on plant-associated bacteria will shed light on their effects on plant and environmental health. For example, decaying wood inhabiting bacteria help convert organic matters back to usable chemicals, acting as decomposers to re-shape a forest.

More than 130 species in the genus *Flavobacterium* have been identified and 59 of them have their genome sequences available in the NCBI database. By comparing orthologous gene families in *Flavobacteria*, we found that the presence and absence of gene families revealed close relationships between IK-1^T and *F. rivuli* or *F. subsaxonicum*, which is consistent with the comparison result of the 16S rRNA gene sequences. *F. rivuli* and *F. subsaxonicum* were isolated from water rivulet of North Germany. However, the detailed analysis of *F. rivuli* suggested that it belongs to the terrestrial clade [5,17]. Synteny analysis revealed extensive synteny relationships exist between the genomes of IK-1^T and *F. rivuli*/*F. subsaxonicum*. Although these species were isolated from two spots, far apart on different continents, they showed dramatic conserved genome features. Notably, the density plots of Ks distributions of orthologous genes between IK-1^T and *F. rivuli* or *F. subsaxonicum* displayed a bimodal pattern which fitted to a bi-normal or bi-gamma distribution. In our unpublished work, we also found a bi-normal or bi-gamma Ks distribution of orthologous genes in close phylogenetic neighbors in the genus *Luteimonas*. These findings indicate that a more flexible relaxed molecular clock model may apply for bacteria with close phylogenetic relationships. In addition, orthologs of central dogma genes and other house-keeping genes from IK-1^T and *F. rivuli* or *F. subsaxonicum* evolve with lower synonymous mutation rates. Taken together, our analyses of gene gain and loss events, conserved genome regions, and synonymous mutation events consistently suggest that *F. rivuli* and *F. subsaxonicum* are the closest phylogenetic neighbors of IK-1^T.

Our analyses of gene families of CAZy domains in IK-1^T suggest that it belongs to a terrestrial clade of *Flavobacteria*. The nearest phylogenetic neighbor of IK-1^T, *F. rivuli*, lacks GH10 xylanase and GH115 α -glucuronidase, which act on xylans in the secondary wall of angiosperms and lignified monocots. By contrast, IK-1^T encodes one GH10 and one GH115, suggesting that compared to *F. rivuli*, IK-1^T obtained GH10 and GH115 to be more active on xylans. In addition, IK-1^T encodes one more GH28 than *F. rivuli*, which is involved in degradation of the major structural component of a plant cell wall. IK-1^T also encodes four more GH43 endoarabinanases than *F. rivuli*, which degrade important plant polysaccharide arabinan. Taken together, IK-1^T has gained extra CAZy gene families during evolution and adapted to plant hosts.

In the genus *Flavobacterium*, motility system and T9SS share a range of components. Motilities of *F. psychrophilum*, *F. johnsoniae*, and *F. columnare* are based on gliding mechanism. Comparisons of 37 genomes of members of the phylum *Bacteroidetes* defined a core set of orthologs for 11 of the gliding motility genes [31]. In contrast, no gliding motility was observed for IK-1^T [14]. The lacks of the three adhesion-like proteins (RemA, RemB, and SprB) were reported in the draft genome sequence of *F. rivuli*, which is also a non-gliding *Flavobacterium* species [15]. These results suggest that like *F. rivuli*, the gliding system encoding genes are maintained in the IK-1^T genome but the adhesion-like proteins may be required for gliding. Since the gliding system can serve as a type 9 secretion system, also known as the Por secretion system (PorSS) [31–33], the conserved gliding motility genes in IK-1^T may function to secrete its T9SS effectors involved in bacteria-host interactions.

Bacterial secretion systems export proteins or virulence factors out of bacterial membrane to extracellular environment or host cells. Up to date, species harboring T2SS are extracellular pathogens. The proportions of type 2 secretomes in the proteomes of Gram-negative bacteria ranged from 12.6% to 42.4% [34]. Based on the fitted regression equation, secretome percentage = 0.0017 \times proteome size + 23.8 [34], the secretome of IK-1^T (predicted to be ~30.7%, ~1248 proteins) is consistent with the secretome range in Gram-negative bacteria. Eukaryotic signaling domains shared between eukaryotes and bacteria can be either retained in genomes evolved from last universal common ancestor (LUCA) or acquired by direct HGT [24]. The eukaryotic-like domains in bacteria may have variable functions and mimic eukaryotic counterparts to regulate host physiology. Their important unknown functions restricted the conserved amino acid sequences during evolution. IK-1^T genome encodes three or four copies of ANK-, vWFA-, and MRJP-containing proteins, respectively. Sequence comparisons showed that most of the eukaryotic-like domains of IK-1^T were closely clustered with orthologous genes from its nearest phylogenetic neighbors *F. rivuli* and *F. subsaxonicum*. One ANK-containing protein is a singleton in *Flavobacteria*, whereas one vWFA- and one MRJP-containing protein was clustered with other *Flavobacteria* species. These suggest that the ANK singleton may be obtained by a unique HGT event, and vWFA- and MRJP-containing proteins were retained from MRCA. Taken together, these proteins containing putative eukaryotic signaling domains may play pivotal roles in bacterial interactions with plant host. Further functional characterizations of these proteins will help determine their roles in bacteria-host interactions.

Plant-associated bacteria use a quorum sensing mechanism to communicate within the community to regulate gene expressions and population behaviors. In this study, we identified a novel putative quorum sensing system, *FaRpfC-FaYpdB*, in IK-1^T. *FaRpfC* with the N-terminal CHASE3 domain represents a new type of the hybrid sensory kinase RpfC. The CHASE3 domain is usually found in MCP (methyl-accepting chemotaxis protein), and the homology model of CHASE3 domain structure suggested a remote similarity to the periplasmic ligand-binding domain of the aspartate chemoreceptor Tar from *Escherichia coli* and *Salmonella enterica* serovar Typhimurium [35,36]. The CHASE3 domain from *Sphingomonas melonis* Fr1 senses heat shock and different salts in the general stress response [37]. Further functional characterization of this novel quorum sensing system will provide insights to the sensing-signaling mechanism in IK-1^T.

Taken together, our analyses of IK-1^T genomic traits suggest that IK-1^T is a plant-associated bacterium and has adapted to terrestrial environment. Using IK-1^T as an example, we showed that

HGT events occurred and may help IK-1^T survive in the plant host. Further functional characterization of IK-1^T unique genes will improve our understanding in specific ecological roles and evolution of these new genes. Moreover, large-scale analysis of bacterial orthologous genes may reveal a ubiquitous bimodal distribution of synonymous mutation rate in phylogenetic neighbors. Of note, surveying genomes of environmental microbes will contribute to the gene bank of Earth and provide the genetic basis for evolution analyses.

4. Materials and Methods

4.1. Genome Functional Annotation

The genome sequence of IK-1^T was annotated in prokka 1.11 [38]. Repeat regions were annotated using RepeatMasker [39]. Protein functions and Gene Ontology (GO) terms were annotated using InterProScan [40]. RNA families were further annotated using Rfam database and perl script rfam_scan [41]. Pathways were annotated using Kyoto Encyclopedia of Genes and Genomes (KEGG) website [42–44]. Synteny analysis was carried out using Symap 4.0 [45,46]. Scaffolds above 100 kb were compared in Symap 4.0. Neighbor-joining trees were constructed using MEGA X based on Dayhoff model and pairwise matching [47]. Protein domains were analyzed using SMART (Simple Modular Architecture Research Tool) [48] and NCBI CDD (Conserved Domain Database).

4.2. Annotation of Carbohydrate-Active Enzymes and Peptidases

Both the CAZy Analysis Toolkit (<http://mothra.ornl.gov/cgi-bin/cat.cgi>) and the dbCAN prediction web server (<http://csbl.bmb.uga.edu/dbCAN/>) were used to annotate genes encoding carbohydrate binding and metabolic enzymes. Only genes were identified by both programs were selected for principle component analysis (PCA) using R. The MEROPS database 10.0 manually curated for peptidase was used to annotate peptidases and the families [23].

4.3. Clustering of Orthologous Groups

The Genome sequences of 59 members of *Flavobacterium* spp. were downloaded from NCBI (Table S2) and annotated using Prokka 1.11. Orthologous groups of proteins were identified by filtering all-vs-all BLASTP results and orthoMCL [49]. If the aligned length was less than 80% of the longer length of two protein sequences, the BLASTP result was filtered out. Then if the aligned length was equal to or above 150 amino acids, the BLASTP result with pairwise identity below 30% was filtered out; if the aligned length was less than 150 amino acids, the BLASTP result was filtered by the formula $(100 \times (0.06 + 4.8 \times L^{-(-0.32 \times (1 + e^{-L/1000}))))$, in which L stands for the aligned length [50]. Gene accumulation curve was plotted using R package vegan [51]. PCA was carried out using R command ggfortify [52,53]. The percentages of PC1 and PC2 in Figure 2 are 84.9% and 8.49%, respectively. *Flavobacterium* species were clustered using R package cluster. The percentages of PC1 and PC2 in Figure 3b are 15.95% and 4.88%, respectively.

4.4. Analysis of Synonymous/Nonsynonymous Substitution Rates

Synonymous substitution rate (Ks) and the ratio of nonsynonymous substitution rate to synonymous substitution rate (Ka/Ks ratio) were calculated by ParaAT and KaKs_Calculator packages [54,55]. Density plots of Ka and Ks values were created using R commands. Fitting of Ks distribution frequency was carried out using in-house MATLAB program.

4.5. Analysis of *Flavobacterium Akiainvovens* IK-1^T Secretion System

PrediSi server (<http://www.predisi.de/home.html>) and SignalP 4.1 server were used to predict signal peptides in the proteome [56]. EffectiveELD database was used to predicted potential secreted proteins [25].

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/19/4910/s1>.

Author Contributions: Conceptualization, X.W.; methodology, X.W.; validation, X.W.; formal analysis, X.W.; investigation, X.W.; resources, X.W.; data curation, X.W.; writing—original draft preparation, X.W.; writing—review and editing, X.W.; visualization, X.W.; funding acquisition, X.W.

Funding: This work was supported by the Fundamental Research Funds for the Central Universities, Nankai University (grant number 63191440).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

IK-1 ^T	<i>Flavobacterium akiainvivens</i> IK-1 ^T
LUCA	last universal common ancestor
MRCA	most recent common ancestor
HGT	horizontal gene transfer
ELD	eukaryotic-like domains
ANK	ankyrin repeat
vWFA	von Willebrand factor type A domain
MRJP	major royal jelly proteins

References

1. Bergey, D.H.; Harrison, F.C.; Breed, R.S.; Hammer, B.W.; Huntoon, F.M. *Bergey's Manual of Determinative Bacteriology*, 1st ed.; The Williams and Wilkins Co.: Baltimore, MD, USA, 1923; pp. 1–442.
2. Kirchman, D.L. The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *Fems Microbiol. Ecol.* **2002**, *39*, 91–100. [[CrossRef](#)]
3. Fernández-Gómez, B.; Richter, M.; Schüler, M.; Pinhassi, J.; Acinas, S.G.; González, J.M.; Pedrós-Alió, C. Ecology of marine *Bacteroidetes*: A comparative genomics approach. *ISME J.* **2013**, *7*, 1026–1037. [[CrossRef](#)] [[PubMed](#)]
4. Teeling, H.; Fuchs, B.M.; Becher, D.; Klockow, C.; Gardebrecht, A.; Bennke, C.M.; Kassabgy, M.; Huang, S.; Mann, A.J.; Waldmann, J.; et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **2012**, *336*, 608–611. [[CrossRef](#)] [[PubMed](#)]
5. Kolton, M.; Sela, N.; Elad, Y.; Cytryn, E. Comparative Genomic Analysis Indicates that Niche Adaptation of Terrestrial Flavobacteria Is Strongly Linked to Plant Glycan Metabolism. *PLoS ONE* **2013**, *8*, e76704. [[CrossRef](#)] [[PubMed](#)]
6. Soltani, A.A.; Khavazi, K.; Asadi-Rahmani, H.; Omidvari, M.; Dahaji, P.A.; Mirhoseyni, H. Plant growth promoting characteristics in some *Flavobacterium* spp. isolated from soils of Iran. *J. Agric. Sci.* **2010**, *2*, 106–115. [[CrossRef](#)]
7. Bodour, A.A.; Guerrero-barajas, C.; Jiorle, B.V.; Malcomson, M.E.; Paull, A.K.; Somogyi, A.; Trinh, L.N.; Bates, R.B.; Maier, R.M. Structure and characterization of Flavolipids, a novel class of biosurfactants produced by *Flavobacterium* sp. strain MTN11. *Appl. Environ. Microbiol.* **2004**, *70*, 114–120. [[CrossRef](#)] [[PubMed](#)]
8. Lee, J.Y.; Xun, L. Purification and characterization of 2,6-di-chloro-p-hydroquinone chlorohydrolase from *Flavobacterium* sp. strain ATCC 39723. *J. Bacteriol.* **1997**, *179*, 1521–1524.
9. Lo, K.V.; Zhu, C.M.; Cheuk, W. Biodegradation of pentachloro-phenol by *Flavobacterium* species in batch and immobilized continuous reactors. *Env. Technol.* **1998**, *19*, 91–96. [[CrossRef](#)]
10. McAllister, K.A.; Lee, H.; Trevors, J.T. Microbial degradation of pentachlorophenol. *Biodegradation.* **1996**, *7*, 1–40. [[CrossRef](#)]
11. Negoro, S. Biodegradation of nylon oligomers. *Appl. Microbiol. Biotechnol.* **2000**, *54*, 461–466. [[CrossRef](#)]
12. Wang, S.Y.; Vipulanandan, C. 2001 Biodegradation of naphthalene-contaminated soils in slurry bioreactors. *J. Env. Eng. Asce.* **2000**, *127*, 748–754. [[CrossRef](#)]
13. Bodour, A.A.; Drees, K.P.; Maier, R.M. Distribution of Biosurfactant-Producing Bacteria in Undisturbed and Contaminated Arid Southwestern Soils. *Appl. Env. Microbiol.* **2003**, *69*, 3280–3287. [[CrossRef](#)] [[PubMed](#)]

14. Kuo, I.; Saw, J.; Kapan, D.D.; Christensen, S.; Kaneshiro, K.Y.; Donachie, S.P. *Flavobacterium akiainvivens* sp. nov., from decaying wood of *Wikstroemia oahuensis*, Hawai'i, and emended description of the genus *Flavobacterium*. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 3280–3286. [[CrossRef](#)] [[PubMed](#)]
15. Hahnke, R.L.; Stackebrandt, E.; Meier-Kolthoff, J.P.; Tindall, B.J.; Huang, S.; Rohde, M.; Lapidus, A.; Han, J.; Trong, S.; Hayneset, M.; et al. High quality draft genome sequence of *Flavobacterium rivuli* type strain WB 3.3-2^T (DSM 21788^T), a valuable source of polysaccharide decomposing enzymes. *Stand. Genom. Sci.* **2015**, *10*. [[CrossRef](#)]
16. Ali, Z.; Cousin, S.; Frühling, A.; Brambilla, E.; Schumann, P.; Yang, Y.; Stackebrandt, E. *Flavobacterium rivuli* sp. nov., *Flavobacterium subsaxonicum* sp. nov., *Flavobacterium swingsii* sp. nov. and *Flavobacterium reichenbachii* sp. nov., isolated from a hard water rivulet. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2610–2617. [[CrossRef](#)] [[PubMed](#)]
17. Cousin, S. Flavobacterial community structure in a hardwater rivulet and adjacent forest soil, Harz mountain, Germany. *Curr. Microbiol.* **2009**, *58*, 409–415. [[CrossRef](#)]
18. Wan, X.; Hou, S.; Saito, J.A.; Kaneshiro, K.Y.; Donachie, S.P. Genome Sequence of *Flavobacterium akiainvivens* IK-1^T, Isolated from Decaying *Wikstroemia oahuensis*, an Endemic Hawaiian Shrub. *Genome Announc.* **2015**, *3*. [[CrossRef](#)]
19. Heaps, H.S. *Information Retrieval: Computational and Theoretical Aspects*; Academic Press Inc.: Orlando, FL, USA, 1978.
20. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [[CrossRef](#)]
21. Yang, Y.; Yu, N.; Tai, P.C. SecE-depleted Membranes of *Escherichia coli* Are Active. *J. Biol. Chem.* **1997**, *272*, 13660–13665. [[CrossRef](#)]
22. Beckwith, J. The Sec-dependent pathway. *Res. Microbiol.* **2014**, *164*, 497–504. [[CrossRef](#)]
23. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [[CrossRef](#)] [[PubMed](#)]
24. Ponting, C.P.; Aravind, L.; Koonin, E.V. Eukaryotic Signalling Domain Homologues in Archaea and Bacteria Ancient. *J. Mol. Biol.* **1999**, *289*, 729–745. [[CrossRef](#)] [[PubMed](#)]
25. Eichinger, V.; Nussbaumer, T.; Platzer, A.; Jehl, M.A.; Arnold, R.; Rattei, T. EffectiveDB-updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic Acids Res.* **2016**, *44*, D669–D674. [[CrossRef](#)] [[PubMed](#)]
26. Al-Khodor, S.; Price, C.T.; Kalia, A.; Abu Kwaik, Y. Ankyrin-repeat containing proteins of microbes: A conserved structure with functional diversity. *Trends in Microbiol.* **2010**, *18*, 132–139. [[CrossRef](#)] [[PubMed](#)]
27. Baldwin, E.T.; Sarver, R.W.; Bryant, G.L., Jr.; Curry, K.A.; Fairbanks, M.B.; Finzel, B.C.; Garlick, R.L.; Heinrikson, R.L.; Horton, N.C.; Kelley, L.C.; et al. Cation binding to the integrin CD11b domain and activation model assessment. *Structures* **1998**, *6*, 923–935. [[CrossRef](#)]
28. Drapeau, M.D.; Albert, S.; Kucharski, R.; Prusko, C.; Maleszka, R. Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. *Genome Res.* **2006**, *16*, 1385–1404. [[CrossRef](#)]
29. Da Silva, F.R.; Vettore, A.L.; Kemper, E.L.; Leite, A.; Arruda, P. Fastidian gum: The exopolysaccharide possibly involved in bacterial pathogenicity. *FEMS Microbiol. Lett.* **2001**, *203*, 165–171. [[CrossRef](#)]
30. Locey, K.J.; Lennon, J.T. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5970–5975. [[CrossRef](#)]
31. McBride, M.J.; Zhu, Y. Gliding Motility and Por Secretion System Genes Are Widespread among Members of the Phylum *Bacteroidetes*. *J. Bacteriol.* **2013**, *195*, 270–278. [[CrossRef](#)]
32. Sato, K.; Naito, M.; Yukitake, H.; Hirakawa, H.; Shoji, M.; McBride, M.J.; Rhodes, R.G.; Nakayama, K. A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 276–281. [[CrossRef](#)]
33. Abby, S.S.; Rocha, E.P.C. Identification of protein secretion systems in bacterial genomes using MacSyFinder. *Methods Mol. Biol.* **2017**, *1615*, 1–21. [[PubMed](#)]
34. Song, C.; Kumar, A.; Saleh, M. Bioinformatic Comparison of Bacterial Secretomes. *Genom. Proteom. Bioinform.* **2009**, *7*, 37–46. [[CrossRef](#)]

35. Zhulin, I.B.; Nikolskaya, A.N.; Galperin, M.Y. Common extracellular sensory domains in transmembrane receptors for diverse signal transduction pathways in Bacteria and Archaea. *J. Bacteriol.* **2003**, *185*, 285–294. [[CrossRef](#)]
36. Mascher, T.; Helmann, J.D.; Unden, G. Stimulus Perception in Bacterial Signal-Transducing Histidine Kinases. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 910–938. [[CrossRef](#)] [[PubMed](#)]
37. Kaczmarczyk, A.; Hochstrasser, R.; Vorholt, J.A.; Francez-Charlot, A. Two-Tiered Histidine Kinase Pathway Involved in Heat Shock and Salt Sensing in the General Stress Response of *Sphingomonas melonis* Fr1. *J. Bacteriol.* **2015**, *197*, 1466–1477. [[CrossRef](#)]
38. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)] [[PubMed](#)]
39. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013–2015. Available online: <http://repeatmasker.org/>.
40. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: Protein domains identifier. *Nucleic Acids Res.* **2005**, *33*, 116–120. [[CrossRef](#)]
41. Burge, S.W.; Daub, J.; Eberhardt, R.; Tate, J.; Barquist, L.; Nawrocki, E.P.; Eddy, S.R.; Gardner, P.P.; Bateman, A. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **2012**, *41*, 226–232. [[CrossRef](#)]
42. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [[CrossRef](#)]
43. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)]
44. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [[CrossRef](#)] [[PubMed](#)]
45. Soderlund, C.; Bomhoff, M.; Nelson, W. SyMAP v3.4: A turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* **2011**, *39*, 68. [[CrossRef](#)]
46. Soderlund, C.; Nelson, W.; Shoemaker, A.; Paterson, A. SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* **2006**, *16*, 1159–1168. [[CrossRef](#)] [[PubMed](#)]
47. Kumar, S.; Stecher, G.; Li, M.; Niyaz, C.; Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2006**, *35*, 1547–1549. [[CrossRef](#)] [[PubMed](#)]
48. Letunic, I.; Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **2018**, *46*, D493–D496. [[CrossRef](#)]
49. Li, L.; Stoekert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res.* **2003**, *13*, 2178–2189. [[CrossRef](#)] [[PubMed](#)]
50. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng.* **1999**, *12*, 85–94. [[CrossRef](#)] [[PubMed](#)]
51. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlenn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. vegan: Community Ecology Package. R package version 2.5–5. 2019. Available online: <https://github.com/vegandevs/vegan>.
52. Tang, Y.; Horikoshi, M.; Li, W. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *R. J.* **2016**, *8*, 478–489. [[CrossRef](#)]
53. Horikoshi, M.; Tang, Y. ggfortify: Data Visualization Tools for Statistical Analysis Results. 2016. Available online: <https://CRAN.R-project.org/package=ggfortify> (accessed on 30 August 2019).
54. Zhang, Z.; Li, J.; Zhao, X.Q.; Wang, J.; Wong, G.K.; Yu, J. KaKs Calculator: Calculating Ka and Ks through model selection and model averaging. *Genom. Proteom. Bioinform.* **2006**, *4*, 259–263. [[CrossRef](#)]
55. Zhang, Z.; Xiao, J.; Wu, J.; Zhang, H.; Liu, G.; Wang, X.; Dai, L. ParaAT: A parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **2012**, *419*, 779–781. [[CrossRef](#)]
56. Nielsen, H. Predicting secretory proteins with SignalP. *Protein Funct. Predict.* **2017**, *1611*, 59–73.

