



OPEN

DSF inactivator RpfB homologous FadD upregulated in *Bradyrhizobium japonicum* under iron limiting conditions

Kunal Dutta^{1✉}, Sergey Shityakov^{1✉} & Fumito Maruyama^{2✉}

Phytopathogenic bacteria *Xanthomonas campestris* pv. *campestris* (*Xcc*) causes black rot and other plant diseases. *Xcc* senses diffusible signal factor (DSF) as a quorum-sensing (QS) signal that mediates mainly iron uptake and virulence. RpfB deactivates DSF in this DSF–QS circuit. We examined differential gene expression profiles of *Bradyrhizobium japonicum* under low versus high iron conditions and found that *fadD* and *irr* were upregulated under low iron (log₂ fold change 0.825 and 1.716, respectively). In addition to having similar protein folding patterns and functional domain similarities, FadD shared 58% sequence similarity with RpfB of *Xcc*. The RpfB–DSF and FadD–DSF complexes had SWISSDock molecular docking scores of –8.88 kcal/mol and –9.85 kcal/mol, respectively, and the 100 ns molecular dynamics simulation results were in accord with the docking results. However, significant differences were found between the binding energies of FadD–DSF and RpfB–DSF, indicating possible FadD-dependent DSF turnover. The protein–protein interaction network showed that FadD connected indirectly with ABC transporter permease (ABCtp), which was also upregulated (log₂ fold change 5.485). We speculate that the low iron condition may be a mimetic environmental stimulus for *fadD* upregulation in *B. japonicum* to deactivate DSF, inhibit iron uptake and virulence of DSF-producing neighbors. This finding provides a new option of using *B. japonicum* or a genetically improved *B. japonicum* as a potential biocontrol agent against *Xcc*, with the added benefit of plant growth-promoting properties.

Iron is the most ubiquitous element in the Earth's crust after oxygen, silicon, and aluminum. Iron usually exists in the ferrous (Fe²⁺) or ferric (Fe³⁺) oxidation state¹. Ferric iron has relatively low solubility, although it is the predominant form in Earth's oxygenated environment, and this poses a challenge for bacteria with aerobic metabolism. At low pH, the more soluble ferrous iron is the most abundant in an anaerobic or microaerobic environment². Iron is essential for the activity of regulatory and metabolic enzymes, which is why it is essential for bacterial physiology. However, iron can be toxic to cells because of its potential to transport electrons at physiological pH. Bacteria have evolved mechanisms to actively gather iron from the environment when iron is scarce and to control the availability of intracellular free iron when iron is abundant³. This iron homeostasis is crucial for both virulence and normal cellular processes⁴. Ferric uptake regulator (Fur) is a global transcriptional regulator of iron homeostasis, virulence, and oxidative stress⁴. Fur has been widely researched in *Escherichia coli*⁵ and has been shown to regulate iron-dependent expression of over 90 genes⁶. Fur functions as a positive repressor because when it interacts with iron (its co-repressor) it suppresses transcription and when iron is absent it depresses transcription⁶. Fur is a homodimer DNA-binding protein that binds one ferrous ion per subunit. At low iron concentrations, the DNA binding affinity of apo-Fur (Fur without iron) is approximately 1000 times lower than that of Fur and its activity is depressed, resulting in iron absorption, suppression of RyhB small RNA, decreased iron storage, and decreased iron-protein synthesis⁶. Siderophores are extracellular ferric iron chelating molecules that are synthesized by bacteria for iron acquisition⁷. Bacterial strains that are incapable of siderophores production often utilize siderophores produced by other bacterial strains, which gives them a competitive advantage inter-species competition or iron-limiting conditions⁸. Bacteria have other iron acquisition mechanisms, including heme absorption and a ferrous iron transport system^{2,9}. However, *fur* mutant strains

¹Laboratory of Chemoinformatics, Infochemistry Scientific Center, ITMO University, Saint Petersburg, Russian Federation. ²Microbial Genomics and Ecology, The IDEC Institute, Hiroshima University, Higashihiroshima, Japan. ✉email: kunal_lifesc@mail.vidyasagar.ac.in; kdutta@itmo.ru; shityakoff@itmo.ru; fumito@hiroshima-u.ac.jp

were found to lack appropriate iron homeostasis, and a high concentration of intracellular iron was detected as a result of constantly high siderophore production⁴. In host plants, the *fur* mutant strains had diminished oxidative stress resistance and decreased virulence traits^{4,10}. These findings highlight the importance of iron and Fur in the pathogenicity of phytopathogenic bacteria.

Rhizobacteria are soil-dwelling α -proteobacteria that can build symbiotic relationships with leguminous plants (e.g., soybean)¹¹. In a symbiotic relationship, plant growth-promoting rhizobacteria promote plant development through a variety of mechanisms, including but not limited to nitrogen fixation, phosphate solubilization, phytohormone production, antibiotic production, and quorum-sensing (QS) interference¹². Most bacteria use QS as a cell-density-dependent communication mechanism¹³. QS is mediated mainly by N-acyl homoserine lactone, and the social life of bacteria is partly dictated by QS. In phytopathogenic bacteria such as *Xanthomonas campestris* pv. *campestris* (*Xcc*), QS is mediated by family members of the diffusible signal factor (DSF), such as DSF, *cis*-2-dodecenoic acid (BDSF), *cis,cis*-11-methyldodeca-2,5-dienoic acid (IDSF), *trans*-2-decenoic acid (SDSF) and resuscitation-promoting factor RpfB regulates DSF turnover¹⁴. Long-chain fatty acid CoA ligase (FadD) sequences of *Escherichia coli*, *Sinorhizobium meliloti*, and *Agrobacterium tumefaciens* are homologous with the RpfB sequence of *Xcc*¹⁵. Disruption of QS signals by enzymatic conversion of homoserine lactone or DSF is called quorum quenching, and QQ often improves the competitive fitness of bacterial species¹⁶. *Bradyrhizobium japonicum* is a rhizobacteria that promotes plant growth¹². In *B. japonicum* strain LO, iron homeostasis and iron-regulated genes are regulated by iron response regulator (Irr) protein and not restricted to heme biosynthesis⁷. In this work, we analyzed differential gene expression levels in *B. japonicum* strain LO under low and high iron conditions, using microarray expression profiles from the NCBI Gene Expression Omnibus dataset (GEO: GSE4143) and protein–protein interactions using STRING. We found that expression of the *fadD* gene of *B. japonicum* was upregulated under the low iron condition (log₂ fold change (FC) 0.825). We also found that the amino acid sequences of *B. japonicum* FadD and *Xcc* RpfB shared 58% identity, and that the binding energy of FadD–DSF was significantly higher than that of RpfB–DSF, indicating potential FadD-dependent DSF turnover. We speculate that the low iron condition may be a mimetic environmental stimulus for *B. japonicum* that upregulates *fadD* to deactivate DSF, inhibit iron uptake, and the virulence of DSF-producing neighbors.

Results

Differentially expressed genes in the GEO dataset. Analysis of the GEO dataset GSE4143 showed that 642 genes were significantly differentially expressed (DEGs) under the low iron condition (low versus high iron) (Fig. 1). Only 457 of the 642 DEGs were used for further analyses (Supplementary Table S1). Among the 642 DEGs, 384 were upregulated and 258 were downregulated (Fig. 2a). The *fadD* and *irr* genes of *B. japonicum* were upregulated (log₂ FC 0.825 and 1.716, respectively) (Fig. 2b). Notably, the ABC transporter permease gene (*ABCtp*, *blr3355*) was also upregulated (log₂ FC 5.485).

Gene list functional enrichment analysis. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis with the Benjamini–Hochberg FDR ($P < 0.05$) of the DEGs showed that most of the downregulated genes were involved in flagellar assembly, ribosome, oxidative phosphorylation, metabolic pathways, carbon fixation in photosynthetic organisms, porphyrin and chlorophyll metabolism, carbon metabolism, microbial metabolism in diverse environments, glyoxylate, dicarboxylate metabolism, and biosynthesis of secondary metabolites (Supplementary Table S2), whereas most of the upregulated genes were involved in sulfur relay (*moaE*, *moaD*, *mmmA*) and beta-lactam resistance (*acrB*, *ampC*, *acrA*) pathways (Supplementary Table S3). In addition, we used the KEGG Orthology Based Annotation System (KOBAS) also without the Benjamini–Hochberg FDR filter setting ($p < 0.05$) and identified a QS pathway that had one gene (*fadD*) and 19 background genes (Fig. 2c).

Protein–protein interaction network. We constructed a large protein–protein interaction (PPI) network of the DEGs (Supplementary Fig. S1) and found that the encoded proteins in the top 50 hub nodes were involved in two important KEGG pathways, namely the ribosome and flagellar assembly pathways. The rpmC protein in the ribosome pathway and node 2734684 in the flagellar assembly pathways connect these two metabolic pathways (Fig. 3a). Notably, the top 15 bottleneck nodes included the Irr protein, whereas FadD was not among the top 50 hub or 15 bottleneck nodes (Fig. 3b). The PPI network is composed of subnetworks, including FadD, flagellar assembly, ribosome, and root nodulation clusters (Supplementary Fig. S1). FadD interacts with specific nodes, namely 27354240, 27354241, 27354239, 27354242, 27352300, 27349302, 27351614, 27351617, 27348347 (Fig. 3c). A separate STRING¹⁷-based analysis of the FadD PPI cluster identified significant QS and ABC transporter pathways with FDR-adjusted p values of $7.64e^{-09}$ and $2.82e^{-05}$, respectively (Table 1). Furthermore, FadD and 27354240 participated exclusively in the QS pathway, whereas 27354239, 27354241, 27354242, 27352300, 27349302, and 27351614 participated in the QS and ABC transporter pathways. Additionally, nodes 27351614 and 27348347 participated in various pathways, including cellular iron homeostasis (Fig. 3c). The results also showed that FadD was indirectly connected to ABCtp, which was upregulated (log₂ FC 5.485) (Fig. 3d).

Comparison of FadD and RpfB sequences and structures. Alignment of the *B. japonicum* FadD and *Xcc* RpfB amino acid sequences showed that they shared 58% similarity (Supplementary Fig. S2) and had identical protein family domains in very similar positions in their sequences (Supplementary Table S4). A multiple sequence alignment and phylogenetic tree showed that *B. japonicum* FadD significantly resembled other bacterial FadD amino acid sequences already known to be homologous to the *Xcc* RpfB sequence (Fig. 4). Structural assessment of the AlphaFold¹⁸ predicted protein structures of RpfB (AF-Q8P9K5-F1) and FadD

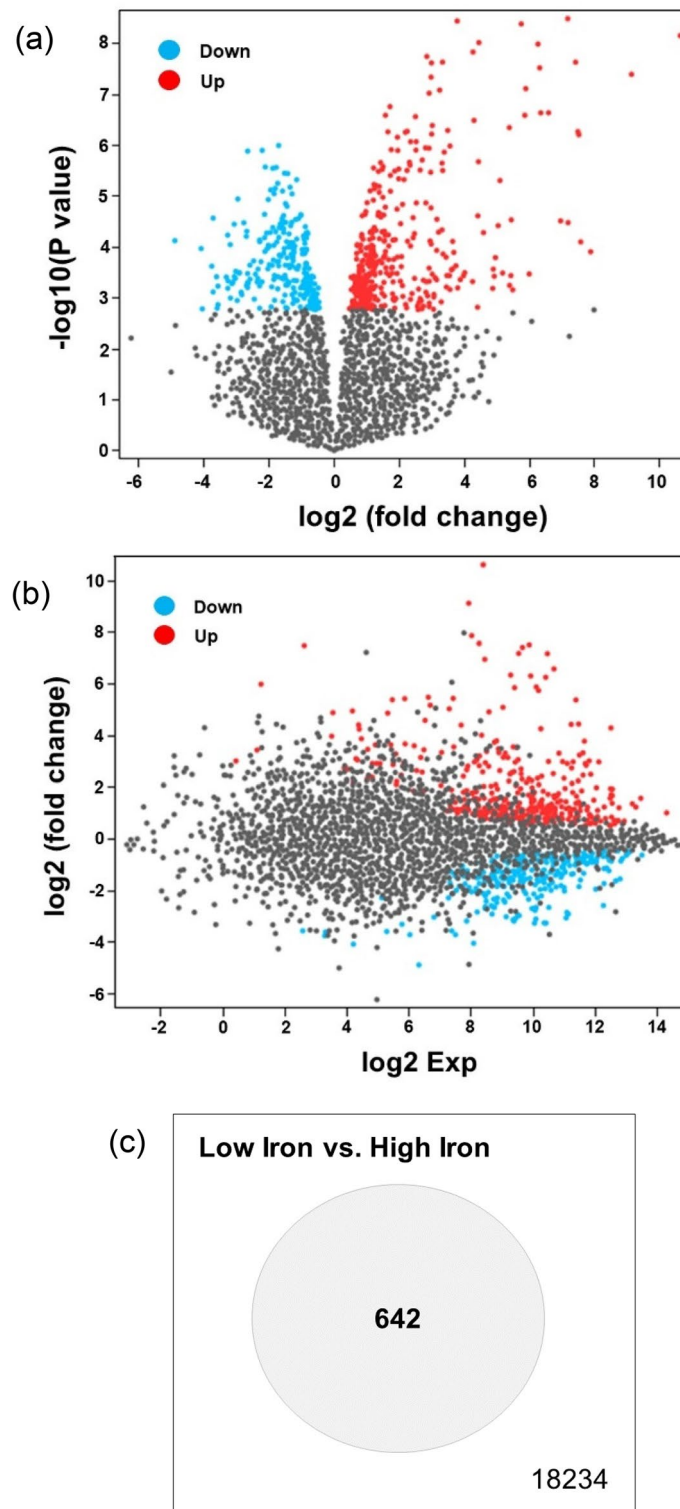


Figure 1. Differentially expressed genes between *Bradyrhizobium japonicum* grown in low and high iron conditions. (a) Volcano plot. (b) Mean difference plot showing \log_2 fold change (x-axis) versus average \log_2 expression values (y-axis). Genes were considered to be significantly differentially expressed when the Benjamini–Hochberg false discovery rate-adjusted p value was < 0.05 . (c) Venn diagram of the overlap in significantly differentially expressed genes between the low and high iron conditions. The Gene Expression Omnibus GSE4143 dataset was used in this analysis. Graphical plots were generated using the GEO2R analysis tool and adapted for better readability.

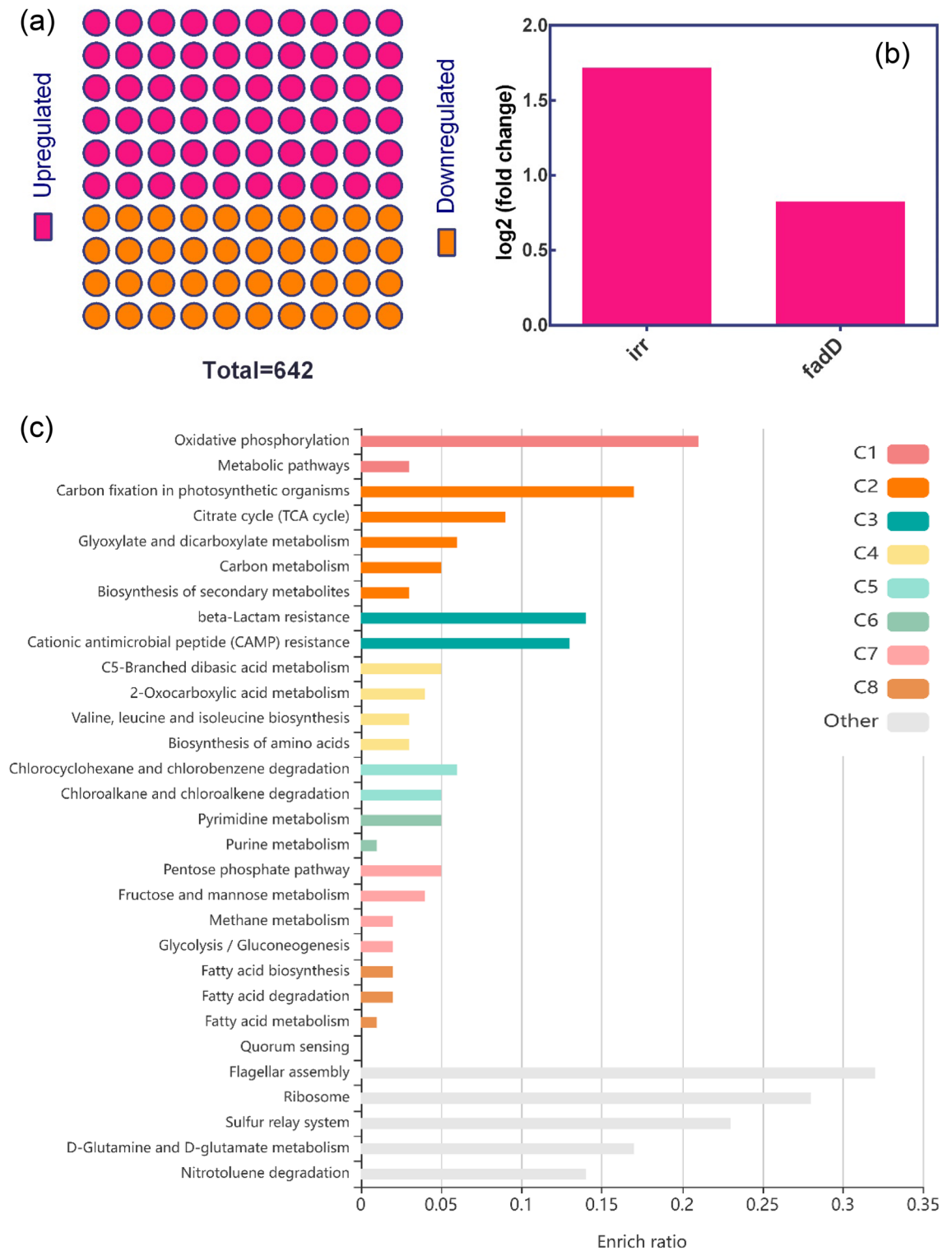


Figure 2. Differentially expressed genes (DEGs) and enriched KEGG metabolic pathways. **(a)** Contingency plot of the 642 DEGs between the low and high iron conditions. **(b)** Differential expression levels of *irr* and *fadD* in low iron compared with high iron conditions. **(c)** KEGG metabolic pathways of the DEGs with no cutoff filter. The quorum sensing pathway had a very low enrichment ratio of 0.01 with one input gene (*fadD*) and 19 background genes. The KEGG pathways were divided into eight main clusters (C1–C8) based on their functions.

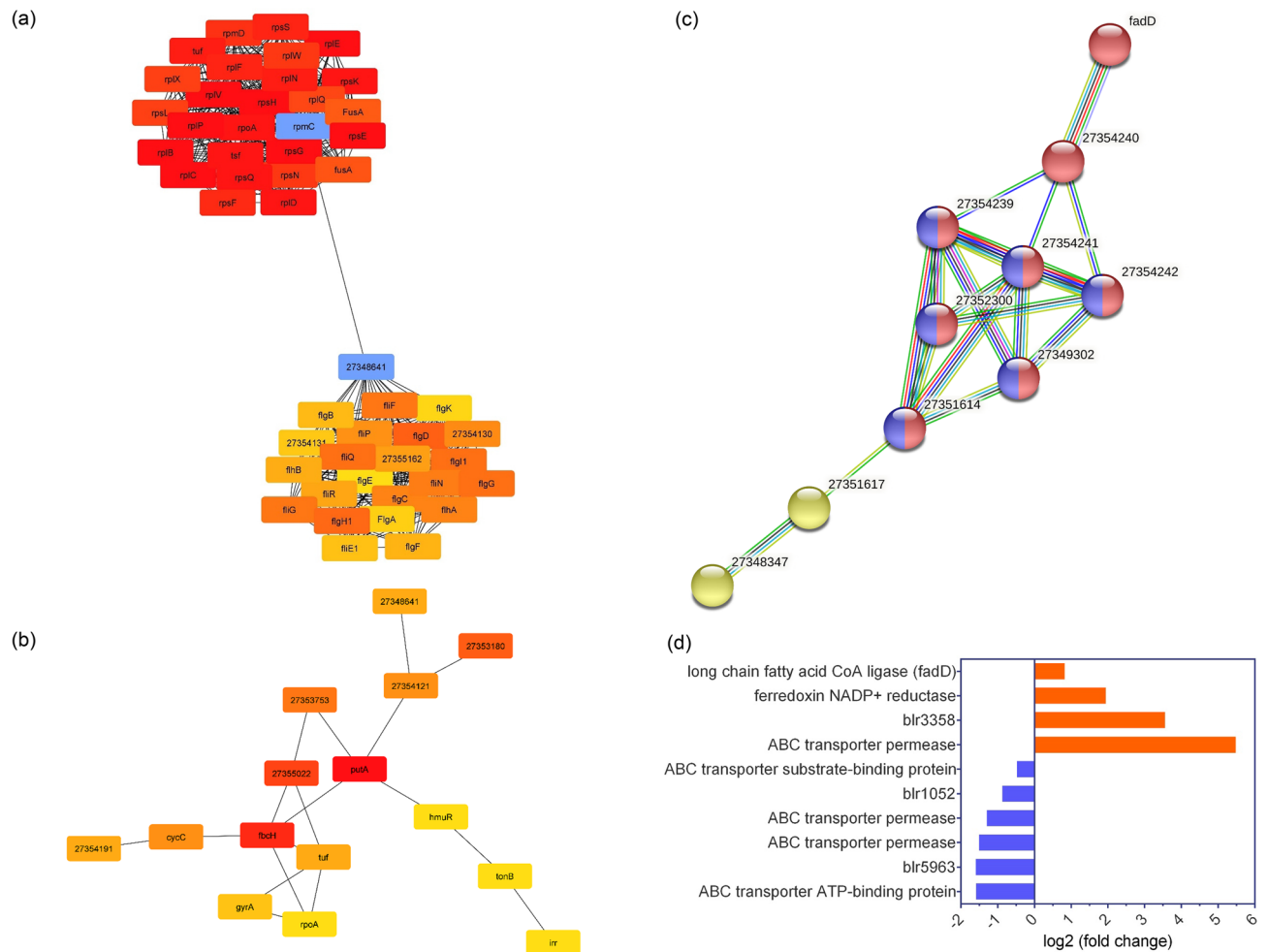


Figure 3. Protein–protein interaction (PPI) network of the differentially expressed genes, and hub and bottleneck nodes. **(a)** Top 50 hub nodes in the PPI network. Connecting hub nodes are in blue boxes. **(b)** Top 15 bottleneck nodes in the PPI network. The Irr protein is bottom right. **(c)** The FadD cluster in the PPI network was separated out and studied using STRING. Nodes are in multiple colors according to the pathways involved: red, quorum sensing (bj02024)^{58–60}, blue, ABC transporters (bj02010)^{58–60}, mixed, including ferredoxin NADP reductase and bacteria; and yellow, cellular iron ion homeostasis (Local network cluster (STRING), CL: 1531). **(d)** Differential expression levels of node genes involved in various pathways, including quorum sensing and ABC transporter pathways.

Description	KEGG pathway	Count in network	Strength	FDR (<i>p</i> -value)
Quorum sensing	bj02024	8 of 273	1.39	7.64e ⁻⁰⁹
ABC transporter	bj02010	6 of 305	1.21	2.82e ⁻⁰⁵

Table 1. KEGG pathways in the FadD cluster from the protein–protein interaction network in Fig. 3c. Count in network: The first number indicates how many proteins in the network are annotated with a particular term. The second number indicates how many proteins in total have this term assigned. Strength: Log₁₀(observed/expected). This measure describes how large the enrichment effect is. It’s the ratio between (i) the number of proteins in the network that are annotated with a term and (ii) the number of proteins that expect to be annotated with this term in a random network of the same size. False Discovery Rate (FDR): This measure describes how significant the enrichment is. Shown are *p* values corrected for multiple testing within each category using the Benjamini–Hochberg procedure.

(AF-A0A0A3XRM6-F1) using the RRDistMaps tool in UCSF Chimera¹⁹ detected similarities in their three-dimensional structures (Supplementary Fig. S2), and *Xcc* RpfB and *B. japonicum* FadD had TM-align²⁰ scores of 0.95995 and 0.95828 respectively, confirming the structures were very similar.

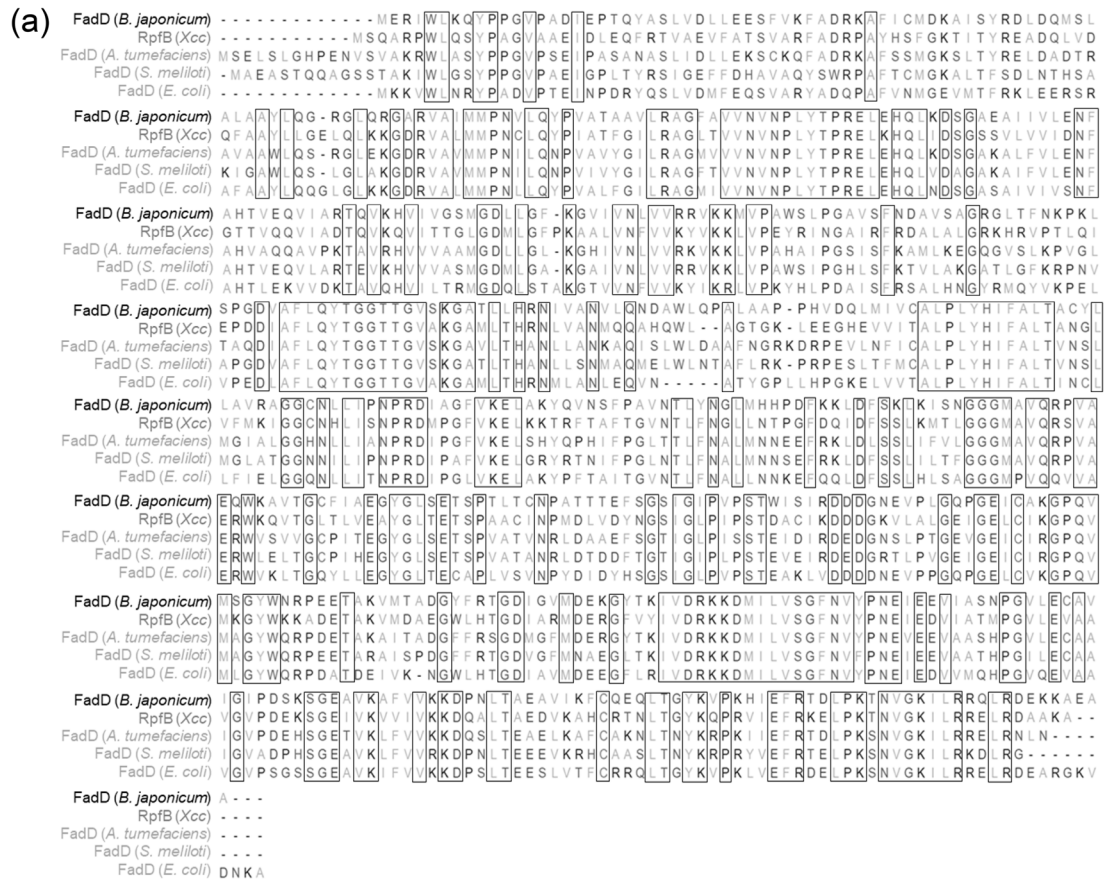


Figure 4. Comparison of the FadD amino acid sequences of *Bradyrhizobium japonicum*, related bacterial species and RpfB of *Xcc*. (a) Multiple sequence alignment, and (b) phylogenetic tree. The bootstrap values are shown next to the branches.

Molecular docking and rescoring. Molecular docking scores of RpfB–DSF and FadD–DSF using CB-Dock2²¹ and AutoDock Vina²² were not largely different (Fig. 5), whereas the SWISSDock²³ and fastDRH²⁴ scores differed significantly. Specifically, RpfB–DSF had a SWISSDock score of -8.88 kcal/mol and fastDRH MM-PBSA and MM-GBSA scores of -22.27 and -34.07 kcal/mol respectively, whereas FadD–DSF had a SWISSDock score of -9.85 kcal/mol and fastDRH MM-PBSA and MM-GBSA scores of -34.02 kcal/mol, and -38.07 kcal/mol respectively (Table 2). The contact residues between RpfB and DSF were PRO84, ASN85, PRO108, LEU109,

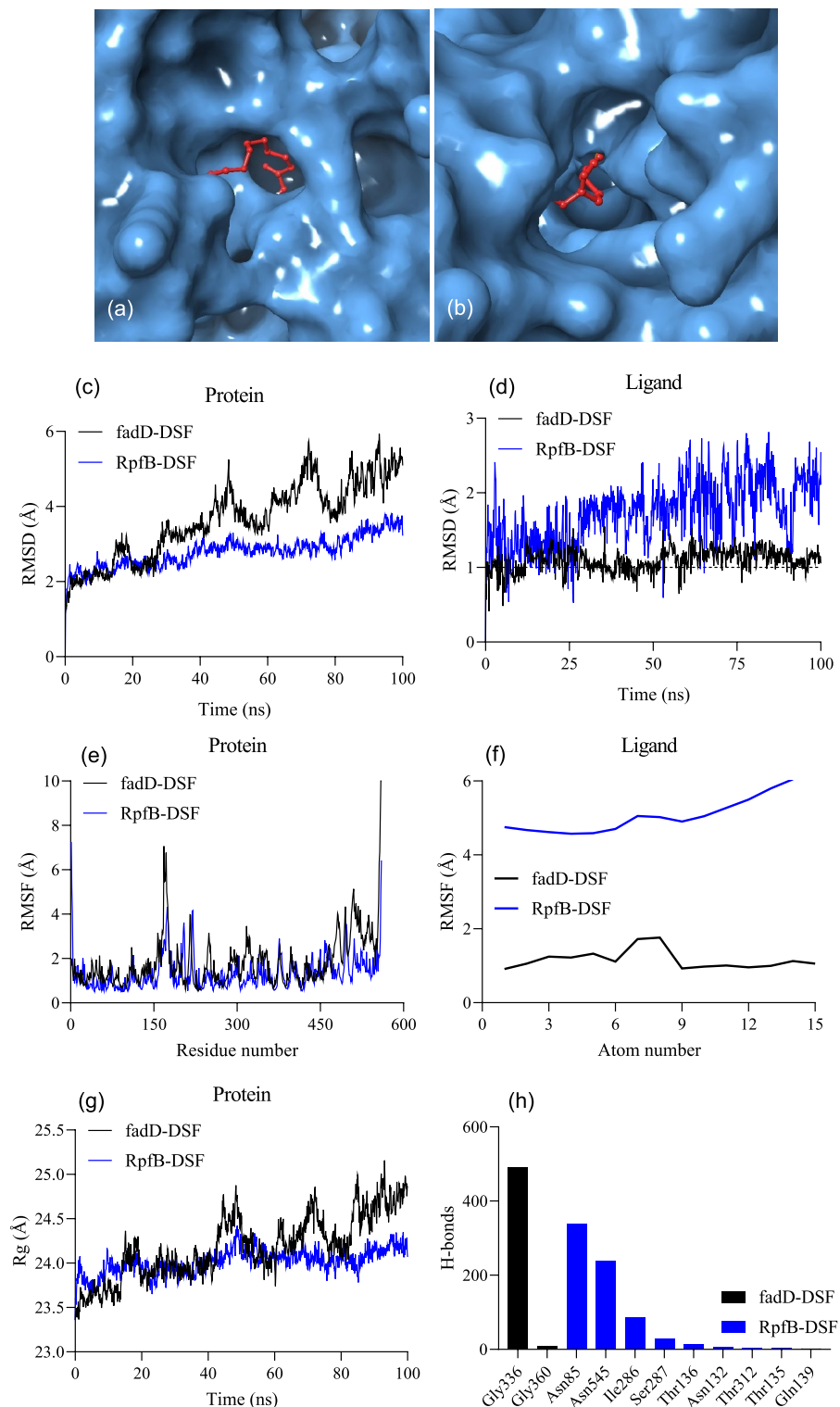


Figure 5. Biophysical interactions between receptors RpfB and FadD and ligand DSF by molecular docking and molecular dynamics (MD) simulations. (a,b) Molecular docking between DSF and RpfB of *Xcc* (a), and between DSF and FadD of *Bradyrhizobium japonicum* (b). DSF is indicated by the red ball-and-stick structures. (c,d) RMSD of the receptors and ligands in the complexes. (e,f) RMSF of the receptors and ligands in the complexes. (g) Radius of gyration (Rg) of the receptor proteins. (h) Numbers of H-bonds and the amino acid residues involved in H-bonding calculated using the receptor–ligand complexes generated during the 100-ns molecular dynamics simulations. The threshold is depicted as a dashed line.

Energy terms	FadD-DSF	RpfB-DSF
SWISS-Dock	− 9.85	− 8.88
CB-Dock2	− 5.74	− 5.73
AD-Vina	− 5.65	− 5.52
MM-PBSA ^a	− 27.42	− 22.27
MM-GBSA ^a	− 38.07	− 34.02
ΔE_{vdW}	− 42.92	− 28.93
ΔE_{el}	− 9.07	− 10.01
ΔG_{gas}	− 51.98	− 38.94
ΔG_{solv}	45.33 ^b /10.93 ^c	37.94 ^b /15.42 ^c
ΔG_{pbsa}	− 6.65	− 1.66
ΔG_{gbsa}	− 41.06	− 23.52

Table 2. Molecular docking energy scores (kcal/mol) for ligand DSF bound to receptors FadD and RpfB and MM-PBSA/GBSA energy terms (kcal/mol) calculated for 100 ns. ^aFastDRH, ^bMM-PBSA; ^cMM-GBSA.

ILE130, ASN132, PHE133, LEU154, LEU176, THR259, ALA260, LEU261, PRO262, LEU263, TYR264, ILE286, SER287, ASN288, PRO289, and ARG290, whereas the contact residues between FadD and DSF were ASN236, TRP239, LEU240, PHE265, ALA269, LEU273, ILE331, ASN333, GLY334, GLY335, GLY336, MET337, GLY358, TYR359, GLY360, LEU361, PRO366, THR367, THR369, and CYS370 (Supplementary Fig. S3). Additionally, the top 10 potential hotspot residues and the top 30 heatmap residues were identified (Supplementary Fig. S4) based on per-residue energy decomposition analysis of multiple docking poses of the ligand²⁴. Therefore, they (Supplementary Fig. S4) are maybe pretty useful for future drug–designing endeavors against RpfB *Xcc*.

Molecular dynamics simulations. To confirm the molecular docking results, the free energy of binding was estimated based on implicit solvation models for the FadD and RpfB proteins complexed with DSF. The MM-PBSA/GBSA calculations using 100 ns MD trajectories showed that the binding affinity for FadD–DSF was much higher than that for RpfB–DSF (Table 2), which confirmed the molecular docking results. van der Waals forces were the main contributors to the protein–ligand affinity because of the lipophilic nature of the ligand molecule.

To explore the movements of the two complexes during the simulations, root-mean-square deviation (RMSD) and fluctuation (RMSF) values, and the radius of gyration (Rg) were plotted versus time (Fig. 5c–g). The FadD–DSF RMSD values were higher (RMSD = 3.64 Å) than those for RpfB–DSF (RMSD = 2.78 Å), indicating FadD deviated more than RpfB from the initial state (Fig. 5c). The ligand RMSD in the FadD–DSF complex remained unchanged at the 1.0 Å threshold, whereas the ligand RMSD in the RpfB–DSF complex changed greatly probably because DSF was released from the binding site (Fig. 5d). The receptor RMSF values identified amino acid residues associated with the high flexibility of the N- and C-terminal regions of RpfB (Fig. 5e). The RMSF values of DSF showed similar flexibility patterns as the RMSF values for the receptors; that is, DSF was more flexible in the RpfB–DSF complex than it was in the FadD–DSF complex (Fig. 5f). The Rg values showed that RpfB was more compact than FadD in the complexes; that is, Rg values were higher for FadD than they were for RpfB (Fig. 5g).

Numbers of H-bonds between the receptors and their ligands were assessed to determine the contribution of H-bonds to the binding affinity. The binding affinity was high between FadD and DSF although only two amino acids (Gly336 and Gly360) were involved in forming H-bonds. However, nine amino acid residues were involved in forming H-bonds between RpfB and DSF, the binding affinity was low probably because the ligand was released from the binding site (Fig. 5h).

Discussion

Bacteria are usually surrounded by various strains and species that compete for the limited life resources²⁵. Therefore, bacteria have developed strategies to decimate and dislodge their neighbors, including but not limited to secretions to gather resources (e.g., siderophores for iron), injuring or poisoning neighboring cells²⁶, and establishing microcolonies while preventing other organisms from doing so²⁷. Biofilms are aggregated microcolonies of surface-associated microbial cells enclosed by an extracellular polymeric substance (EPS) and governed mostly by QS²⁸. QS also has a crucial role in regulating virulence and biofilm formation in *Xcc*²⁹. Conversely, plant growth-promoting rhizobacteria such as *B. japonicum* can promote plant growth by QS interference or quorum quenching of adjacent phytopathogenic bacteria^{10,12}. Recently, there has been a major focus on QS signals, including DSF family signals³⁰, and QS was found to be associated with advantages or competitive interactions²⁶. DSF and BDSF are QS process mediators in *Xcc*¹⁴, and RpfB can deactivate DSF by β -oxidation via its fatty acyl-CoA ligase activity³¹. Furthermore, DSF may potentially impair the development, growth, and immunity of host plants³¹. In *Xcc*, high extracellular concentrations of DSF activate the RpfC–RpfG system at high cell densities, relieving the inhibition of c-di-GMP on the global transcription factor Clp that regulates the expression of several virulence factors and iron absorption³². *Xcc rpfF* produced DSF and *rpfF* homolog mutant *Xanthomonas oryzae* pv. *oryzae* (*Xoo*) strains that were deficient in growth and DSF production³³. *Xoo* is a phytopathogenic bacteria that causes bacterial blight in rice plants. *Xoo rpfF* homolog mutant strains showed unusual tetracycline susceptibility under

low iron conditions and tetracycline resistance upon iron supplementation³³. The iron-dependent pathogenicity of *Xcc* in rice plants has been reported³⁴. *Xcc* can produce xanthoferrin, a siderophore that is necessary for maximum virulence and growth under low iron conditions³⁵. Xanthoferrin can bind with ferric iron and enhance growth of *Xcc* in host plants³⁵. These findings show the importance of iron and the DSF-dependent QS system for the manifestation of virulence traits of phytopathogenic bacteria.

FadD sequences of *E. coli*, *S. meliloti*, and *A. tumefaciens* are homologs of the *Xcc* RpfB sequence¹⁵. However, homology between *B. japonicum* FadD and *Xcc* RpfB has not been reported previously. Alignment of *B. japonicum* FadD and *Xcc* RpfB amino acid sequences showed that they shared 58% similarity (Supplementary Fig. S2). Sequence similarity offers insights into the functions of proteins³⁶. TM-align scores of the three-dimensional structures of two model proteins can be obtained by comparing their residue equivalencies³⁰. *Xcc* RpfB and *B. japonicum* FadD had TM-align scores of 0.95995 and 0.95828 respectively, which is indicative of similar protein folding patterns. Thus, in addition to sharing sequence similarity, they shared protein folding patterns and protein family domains (Supplementary Table S4), indicating potential functional similarities between the two proteins. Furthermore, the RMSD of RpfB in the RpfB–DSF was lower than the RMSD of FadD in FadD–DSF throughout the MD trajectories. RMSD measures the difference between the initial and final structures of a protein. Discrepancies during the MD trajectory of a protein structure may indicate its conformational stability³⁷. For FadD, the RMSD fluctuations did not exceed approximately 4–6 Å (approximately 40–100 ns), indicating a small fluctuation range. For the ligand in the two complexes, the RMSD fluctuations showed opposite trends. The RMSD fluctuations for DSF were higher in the RpfB–DSF complex than they were in the FadD–DSF throughout the MD trajectories (Fig. 5). Small RMSD fluctuations are indicative of stable ligand binding³⁸. RMSF quantifies the average displacements of amino acid residues from a reference point over an MD trajectory, and therefore RMSF can identify structural regions that deviate most from their initial structure³⁹. Notably, some protein regions of RpfB–DSF and FadD–DSF, approximately 150–160 and 455 to the C-terminal end, exhibited similar fluctuations and, on average, the RMSF of RpfB–DSF and FadD–DSF were broadly similar, indicating satisfactory receptor–ligand interactions. The Rg indicates the distribution of atoms around the axis of a protein, and is a measure of the distance between the point while it is spinning and the point where the transfer of energy has the greatest possible impact⁴⁰. For RpfB in RpfB–DSF, the Rg remained between 23.5 Å and 24.5 Å. For FadD in FadD–DSF, the Rg began to fluctuate after 40 ns then stayed close to 25 Å up to 100 ns. The more than 500 H-bonds that were predicted to be formed by GLY336 of FadD–DSF are indicative of the single dominance of GLY336 in the receptor–ligand interactions, whereas numerous amino acid residues were predicted to form H-bonds in RpfB–DSF. Because H-bonds help create and stabilize protein structures⁴¹, the formation of H-bonds throughout the MD trajectory suggests stable receptor–ligand interactions in RpfB–DSF and FadD–DSF.

Statistically significant differences in gene expression levels between two experimental conditions are used to identify DEGs⁴². We identified 642 DEGs in *B. japonicum* by comparing gene expression levels under low and high iron conditions (Fig. 1). Irr protein and FadD are important for iron homeostasis and QS or quorum quenching. We found that *fadD* and *irr* of *B. japonicum* were upregulated under the low iron condition (log₂ FC 0.825 and 1.716, respectively). Furthermore, report indicated that Irr protein was required for an adequate sense of cellular iron status⁵. And, absence of a normal iron response and iron-associated genes in an *irr* mutant strain was found to decrease the total cellular iron content⁵. These characteristics of the Irr protein explain the upregulation of *irr* under the low iron condition. However, the significance of *fadD* upregulation in *B. japonicum* under low iron conditions has not been explained in previous studies.

In the present study, the gene list functional enrichment analysis identified the flagellar assembly and ribosome pathways as the top two pathways with FDR-adjusted *p* values < 0.05 (Supplementary Table S2), and genes associated with these two metabolic pathways were among the top 50 hub nodes in the PPI network (Fig. 3a). Furthermore, the hub nodes that connect these two pathways, rpmC from the ribosome network and 27348641 from the flagellar assembly network, were found to be critically important (Fig. 3a). However, only Irr protein was among the top 15 bottleneck nodes of the entire PPI network (Fig. 3b), although FadD was not among the top 50 bottleneck/hub nodes. Notably, the expression level of *fadD* (log₂ FC 0.825) was almost half that of Irr protein (log₂ FC 1.716). This finding raises the question of why high *fadD* expression occurs in *B. japonicum* when iron availability was restricted.

In bacteria, iron plays crucial physiological roles in DNA replication, transcription, metabolism, energy production through respiration, and pathogenicity³. Low iron conditions can trigger processes that prevent iron uptake by other microorganisms, thereby ensuring optimal conditions for the survival of self. The ferric iron binding transcription factor XibR, which coordinately controls the expression of genes related to iron metabolism, chemotaxis, and flagellar motility, was found to be necessary for optimal virulence of *Xcc*⁴³. In low-iron conditions, *XibR* mutant strains showed decreased growth and decreased intracellular iron concentrations⁴³. The significance of iron metabolism in the pathophysiology of bacteria is further supported by the finding that the strong bactericide Xinjunan (dioctyldiethylenetriamine) acted by altering cellular iron metabolism⁴⁴.

In addition to *fadD*, other genes such as *fadE* and *fadR* have been shown to play critical roles in β-oxidation in *E. coli*⁴⁵. Interestingly, *fadE*, *fadF*, *fadG*, *fadH*, *fadL*, and the transcriptional regulator *fadR* were also found to be involved in bacterial lipid metabolism⁴⁶, but none of these genes were differentially expressed in *B. japonicum* when compared low versus high iron conditions (Supplementary Table S1). During resource-limiting conditions, bacteria generally opt for energy-efficient pathways and nonessential transcription is typically prevented by operons such as the lactose operon in *E. coli*⁴⁷. We found that *fadD* was upregulated in *B. japonicum* under the low iron condition (Fig. 3c), implying that the high expression of *fadD* may be attributable to other factors. The STRING-based analysis of the FadD cluster in the PPI network (Supplementary Fig. S1) found two crucial KEGG pathways: the QS and ABC transporter pathways (Table 1). However, the STRING- and KEGG-based pathway predictions have limitations because these methods are not optimized or adapted for multi-species interactions. Indeed, a homogeneous or isogenic bacterial community is rare because a wide range of microorganisms are

typical inhabitants of bacteria⁴⁸. Moreover, bacteria can injure and outcompete their neighbors when resources are scarce through their cell signaling or metabolic pathways²⁷. Among the other DEGs, *ABCtp* (blr3355) was upregulated (log₂ FC 5.485), and was indirectly connected to FadD (Fig. 3d). Permeases are membrane transporter proteins⁴⁶ that can transport proteins out of cells. The high expression of *ABCtp* (blr3355) under the low iron condition indicates that it was required and probably linked to the extracellular transport of FadD. Furthermore, the ABC transporter was found to be upregulated in *Streptococcus pneumoniae* by two different QS mechanism⁴⁹.

QS plays an important role in inter-species rivalry⁵⁰ and cooperation⁵¹, and QS inhibition across species has been reported⁵². RpfB deactivates DSF in *Xcc*¹⁴, which can disrupt the cyclic di-GMP- and Clp-mediated signaling circuit of cellular iron absorption^{31,32,53–56} similar to *Xanthomonas campestris* (Fig. 6a). Furthermore, *B. japonicum* FadD and *Xcc* RpfB are homologous (Fig. 4) and a significant difference in the binding energies of FadD–DSF and RpfB–DSF was detected (Supplementary Table S3), which indicates potential FadD-dependent DSF turnover (Fig. 6b). Therefore, we speculate that the low iron condition may be a mimetic environmental stimulus for *fadD* upregulation in *B. japonicum* to deactivate DSF and inhibit iron uptake and virulence of DSF-producing neighbors. This may explain why *fadD* and *ABCtp* were upregulated in *B. japonicum* under the low iron condition.

Conclusions

We analyzed differential gene expression of *B. japonicum* under low and high iron conditions and found that FadD, Irr, and ABCtp genes were upregulated (log₂ FC 0.852, 1.724, and 5.485, respectively). The PPI network indicated that FadD was indirectly connected with ABCtp. We identified similar protein folding patterns, functional domains, and 58% sequence similarity between FadD of *B. japonicum* and RpfB of *Xcc*. We also identified QS and ABC transporter pathways by STRING-based analysis of the DEGs. Molecular docking showed a significant difference in binding energies between FadD–DSF and RpfB–DSF, indicating a potential FadD-dependent DSF turnover that was supported by the MD simulations. We speculate that low iron conditions may be a mimetic environmental stimulus for *fadD* upregulation in *B. japonicum* to deactivate DSF-mediated iron uptake and the virulence of DSF-producing neighbors. Our results provide a new option of using *B. japonicum* or genetically modified *B. japonicum* as a biocontrol agent against plant diseases caused by *Xcc*. However, to help assure agricultural production in the face of climate change, exploratory *FadD* gene editing research may be an important focus for future studies.

Methods

Microarray data and literature search. A literature search was performed with keyword combinations such as plant growth-promoting rhizobacteria, PGPR, *Bradyrhizobium japonicum*, quorum sensing, iron, and nodulation using Google Scholar (<https://scholar.google.com/>) and PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). We also searched the microarray datasets in the NCBI Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>) and selected the GSE4143 dataset for analysis (Table 3). GEO samples GSM94778, GSM94780, and GSM94781 were grown in low iron conditions, and GEO samples GSM94783, GSM94784, and GSM94785 were grown in high iron conditions. Briefly, *B. japonicum* LO strain was grown in a modified GSY medium with no exogenous iron source to yield 0.3 μM iron concentration in the medium (low iron condition), and in a modified GSY medium supplemented with 12 μM FeCl₃·6H₂O (high iron condition)⁵.

Differential gene expression. Differentially expressed genes (DEGs) between *B. japonicum* grown in the low and high iron conditions were identified using GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>)⁵⁷. Genes within the cutoff criteria of FDR-adjusted *P*-value < 0.05 (Benjamin–Hochberg) were considered DEGs. Log transformation of the data was set to “auto-detect” mode with force normalization and limma precision weights (Vooma). NCBI-generated category of platform annotation was selected to display on results. The results were obtained in tab-delimited format and exported to Microsoft Excel for further analysis.

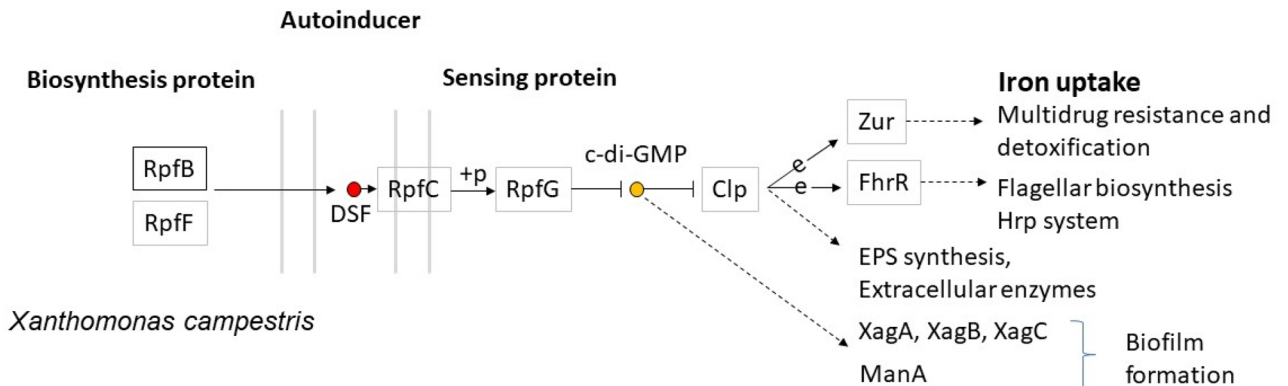
Functional enrichment and ontology of DEGs. The gene symbols of the significant DEGs were used for gene set functional enrichment using KOBAS–Kyoto Encyclopedia of Genes and Genomes (KEGG)^{58–60} pathways (<http://kobas.cbi.pku.edu.cn/>)⁶¹ with FDR-adjusted *P* value < 0.05 as the cutoff for significance. The results were also saved without an FDR-adjusted *P* value of < 0.05 for comparative analysis.

PPI network construction and identification of hub and bottleneck nodes. The gene symbols of the DEGs were used to construct the entire PPI network using STRING (<https://string-db.org/>). The DEGs were annotated with gene ontology terms under the three main categories, biological process, molecular function, and cellular component, as well as local network cluster (STRING), keywords (Uniport), and protein domains and features (InterPro)⁶² with medium confidence settings (i.e., a combined score > 0.4), and saved as short tabular text output. Subsequently, the PPI network topology was exported directly from the STRING website to Cytoscape (v3.9.1)⁶³ (<http://www.cytoscape.org/>) for visualization. The top 50 hub nodes and top 15 bottleneck nodes were identified using the cytoHubba (v0.1) plugin app with the Maximal Clique Centrality (MCC) scoring method⁶⁴. Nodes that interacted with *fadD* were identified and studied separately.

Comparisons of FadD and RpfB. Protein sequences of FadD of *B. japonicum* (UniProt: A0A0A3XRM6), RpfB of *Xcc* (UniProt: Q8P9K5), FadD of *E. coli* (GenBank: UBF39181.1), *Agrobacterium tumefaciens* (GenBank: CAD0207327.1), *Sinorhizobium meliloti* SM11 (GenBank: AEH77411.1) were aligned using MUSCLE (MEGA v11)⁶⁵. The evolutionary history was inferred using the maximum likelihood method and JTT matrix-based

(a)

QUORUM SENSING



(b)

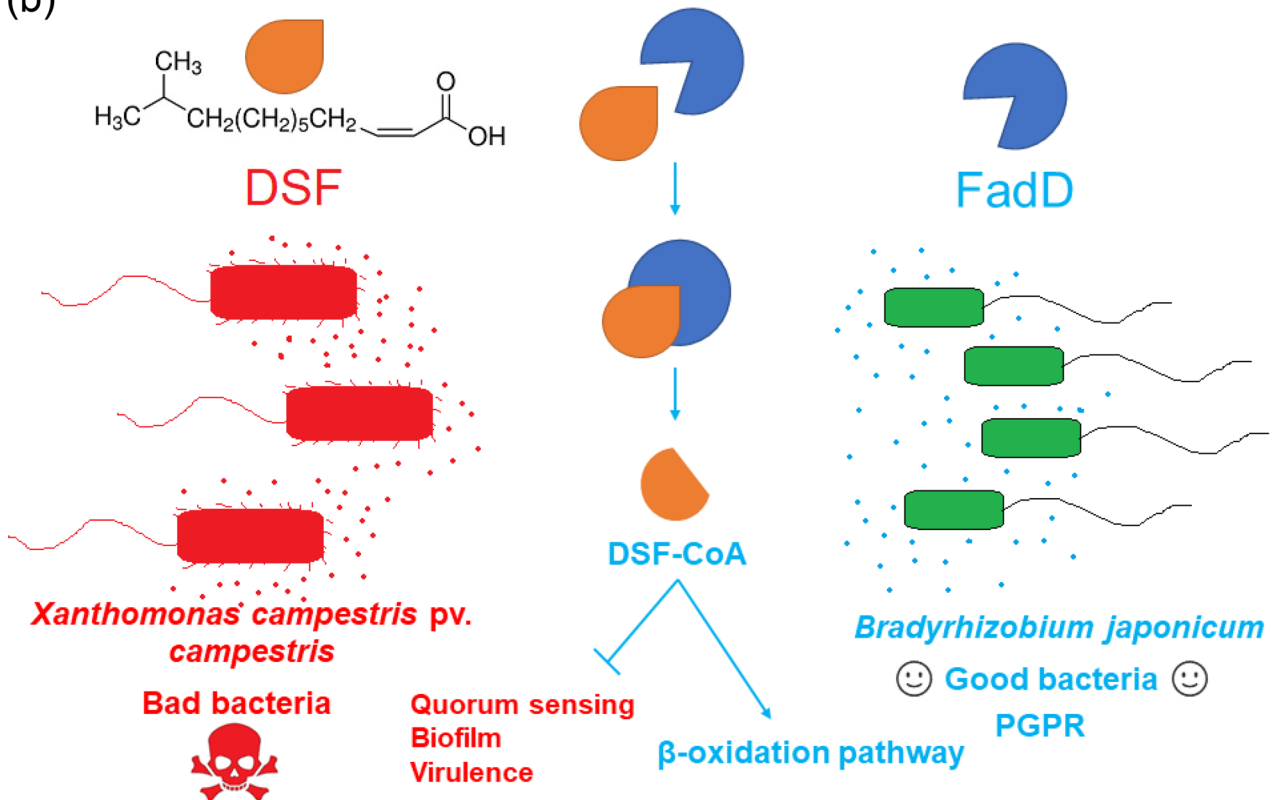


Figure 6. Proposed molecular signaling cascade of DSF-mediated cellular responses. (a) DSF-mediated quorum sensing in *Xanthomonas campestris*, obtained from KEGG (bja02024)^{58–60} (b) Possible mode of action of inter-species communication via DSF of *Xcc* and the RpfB homolog FadD of *Bradyrhizobium japonicum*.

model⁶⁶, and nearest neighbor interchange distance was used as a tree inference option and tested by 1000 bootstrap replications. Evolutionary analyses were conducted in MEGA v11⁶⁵. The protein domains of FadD of *B. japonicum* and RpfB of *Xcc* were annotated using InterProScan⁶⁷. The three-dimensional protein structures were compared using TM-align²⁰ and the RRDistMaps tool in UCSF Chimera¹⁹.

Molecular docking and rescoring. The AlphaFold¹⁸ protein structure files of RpfB (AF-Q8P9K5-F1) and FadD (AF-A0A0A3XRM6-F1) were used as the receptors, and the SDF file of diffused signaling factor (DSF),

Datasets	Sample platform	Samples	Platform organism	Sample organism		
GSE4143	GPL3401	Low iron	<i>Bradyrhizobium diazoefficiens</i> USDA 110	<i>Bradyrhizobium japonicum</i> strain LO		
		GSM94778				
		GSM94780				
		GSM94781				
High iron	GSM94783	GSM94784	GSM94785			
				GSM94783	GSM94784	GSM94785
Process	Grown under an iron limitation or under iron sufficiency and compared to each other by whole genome microarray analysis ⁵					

Table 3. Details of the NCBI Gene Expression Omnibus dataset used in this study.

cis-11-methyl-2-dodecenoic acid (PubChem ID: 11469920) was obtained from the PubChem database and used as the ligand. Auto blind docking was performed using CB-Dock2 (<https://cadd.labshare.cn/cb-dock2/php/index.php>)²¹ with 10 repeats with unaltered parameters, and averages were used for the interpretations. The best-hit ligand–receptor structures were downloaded, and screenshots of the contact residues were saved for further analysis. The receptor–ligand interactions were visualized using free Maestro v13.2 (Maestro, Schrödinger, LLC, New York, NY, 2021)⁶⁸. SWISSDock (<http://www.swissdock.ch/docking>)²³ was also used with the same receptor and ligand. We also used fastDRH (<http://cadd.zju.edu.cn/fastdrh/overview>)²⁴ for high-speed docking with the AutoDock Vina docking engine, and calculated the structure-truncated MM/PB(GB)SA energy and per-residue energy decomposition based on multiple poses. The receptor–ligand complex obtained from CB-Dock2 was used as the binding pocket reference, and then 10 pose numbers were selected. For pose rescoring, receptor forcefield ff99SB (with TIP3P water model) and ligand force field GAFF2 were selected, and the truncation radius setting was kept at the default value with all rescoring procedures. The hotspot predictions were carried out with an unaltered ff99SB forcefield and default truncation radius. To evaluate the docking poses, MM/PBSA and MM/GBSA (to yield an energy decomposition on a per-residue basis) were submitted separately. The results obtained were downloaded and used for further analyses.

Molecular dynamics simulation. All molecular dynamics (MD) simulations were performed using the AMBER 16 package with the ff99SB and GAFF force fields for the receptors RpfB and FadD, and the ligand DSF⁶⁹. The antechamber module of AmberTools was used to calculate the partial charges of the ligand using the semi-empirical AM1-BCC function according to the standard protocol⁷⁰. The complexes were solvated with the TIP3P water model and neutralized by adding Na⁺ ions using the tLeap input script from the AmberTools package. Long-range electrostatic interactions were modeled using the particle-mesh Ewald method⁷¹. The SHAKE algorithm⁷² was applied to constrain the length of covalent bonds, including the hydrogen atoms. Langevin thermostat was implemented to equilibrate the temperature of the system at 310 K. A 2.0-fs time step was used in all the MD setups. For the minimization and equilibration (NVT and NPT ensembles) phases, 100,000 steps and a 1-ns period were used, respectively. Finally, 100 ns classical MD simulations, with no constraints as NPT ensemble, were performed for each of the receptor–ligand complexes using molecular mechanics combined with the Poisson–Boltzmann (MM-PBSA) or generalized Born (MM-GBSA) method augmented with the hydrophobic solvent-accessible surface area term^{73,74}. The MM-PBSA/GBSA solvation models were applied as a post-processing end-state method to calculate the free energies (ΔG_{pbsa} and ΔG_{gsa}).

Data availability

The dataset GSE4143 analyzed during the current study is available at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4143>.

Received: 31 December 2022; Accepted: 18 May 2023

Published online: 29 May 2023

References

1. Franza, T. & Expert, D. Role of iron homeostasis in the virulence of phytopathogenic bacteria: An ‘a la carte’ menu. *Mol. Plant Pathol.* **14**, 429–438 (2013).
2. Andrews, S. *et al.* *Metallomics and the Cell* 203–239 (Springer, 2013).
3. Frawley, E. R. & Fang, F. C. The ins and outs of bacterial iron metabolism. *Mol. Microbiol.* **93**, 609–616 (2014).
4. Jittawattipoka, T., Sallabhan, R., Vattanaviboon, P., Fuangthong, M. & Mongkolsuk, S. Mutations of ferric uptake regulator (fur) impair iron homeostasis, growth, oxidative stress survival, and virulence of *Xanthomonas campestris* pv. *campestris*. *Arch. Microbiol.* **192**, 331–339 (2010).
5. Yang, J. *et al.* *Bradyrhizobium japonicum* senses iron through the status of haem to regulate iron homeostasis and metabolism. *Mol. Microbiol.* **60**, 427–437 (2006).
6. Andrews, S. C., Robinson, A. K. & Rodríguez-Quiriones, F. Bacterial iron homeostasis. *FEMS Microbiol. Rev.* **27**, 215–237 (2003).
7. Raymond, K. N., Müller, G. & Matzanke, B. F. Complexation of iron by siderophores: a review of their solution and structural chemistry and biological function. *Struct. Chem.* 49–102 (1984).
8. Niehus, R., Picot, A., Oliveira, N. M., Mitri, S. & Foster, K. R. The evolution of siderophore production as a competitive trait. *Evolution* **71**, 1443–1455 (2017).
9. Zughaier, S. & Cornelis, P. Role of Iron in bacterial pathogenesis. *Front. Cell. Infect. Microbiol.* **8**, 344 (2018).

10. Vessey, J. K. Plant growth promoting rhizobacteria as biofertilizers. *Plant Soil* **255**, 571–586 (2003).
11. Rivera, M. C. & Izard, J. *Metagenomics for Microbiology* 145–159 (Elsevier, 2015).
12. Bhattacharyya, P. N. & Jha, D. K. Plant growth-promoting rhizobacteria (PGPR): Emergence in agriculture. *World J. Microbiol. Biotechnol.* **28**, 1327–1350 (2012).
13. Bassler, B. L. How bacteria talk to each other: Regulation of gene expression by quorum sensing. *Curr. Opin. Microbiol.* **2**, 582–587 (1999).
14. Baltenneck, J., Reverchon, S. & Hommais, F. Quorum sensing regulation in phytopathogenic bacteria. *Microorganisms* **9**, 239 (2021).
15. Soto, M. J., Fernández-Pascual, M., Sanjuan, J. & Olivares, J. A fadD mutant of *Sinorhizobium meliloti* shows multicellular swarming migration and is impaired in nodulation efficiency on alfalfa roots. *Mol. Microbiol.* **43**, 371–382 (2002).
16. Liu, Y., Qin, Q. & Defoirdt, T. Does quorum sensing interference affect the fitness of bacterial pathogens in the real world?. *Environ. Microbiol.* **20**, 3918–3926 (2018).
17. Szklarczyk, D. *et al.* STRING v10: Protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
18. Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein–sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
19. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
20. Zhang, Y. & Skolnick, J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
21. Liu, Y. *et al.* CB-Dock 2: Improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res.* **50**, W159–W164 (2022).
22. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
23. Grosdidier, A., Zoete, V. & Michielin, O. SwissDock, a protein–small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **39**, W270–W277 (2011).
24. Wang, Z. *et al.* fastDRH: A webserver to predict and analyze protein–ligand complexes based on molecular docking and MM/PB (GB) SA computation. *Brief. Bioinform.* (2022).
25. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: Surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8**, 15–25 (2010).
26. He, Y.-W. *et al.* DSF-family quorum sensing signal-mediated intraspecies, interspecies, and inter-kingdom communication. *Trends Microbiol.* **31**, 36–50 (2022).
27. Ghoul, M. & Mitri, S. The ecology and evolution of microbial competition. *Trends Microbiol.* **24**, 833–845 (2016).
28. Donlan, R. M. Biofilms: Microbial life on surfaces. *Emerg. Infect. Dis.* **8**, 881 (2002).
29. Diab, A. A. *et al.* BDSF is the predominant in-plant quorum-sensing signal used during *Xanthomonas campestris* infection and pathogenesis in Chinese cabbage. *Mol. Plant Microbe Interact.* **32**, 240–254 (2019).
30. Tian, X.-Q., Wu, Y., Cai, Z. & Qian, W. BDSF is a degradation-prone quorum-sensing signal detected by the histidine kinase RpfC of *Xanthomonas campestris* pv. *campestris*. *Appl. Environ. Microbiol.* **88**, e00031–e122 (2022).
31. Song, K. *et al.* The plant defense signal salicylic acid activates the RpfB-dependent quorum sensing signal turnover via altering the culture and cytoplasmic pH in the phytopathogen *Xanthomonas campestris*. *MBio* **13**, e03644–e13621 (2022).
32. Cai, Z. *et al.* Fatty acid DSF binds and allosterically activates histidine kinase RpfC of phytopathogenic bacterium *Xanthomonas campestris* pv. *campestris* to regulate quorum-sensing and virulence. *PLoS Pathog.* **13**, e1006304 (2017).
33. Chatterjee, S. & Sonti, R. V. rpfF mutants of *Xanthomonas oryzae* pv. *oryzae* are deficient for virulence and growth under low iron conditions. *Mol. Plant-Microbe Interact.* **15**, 463–471 (2002).
34. Rai, R., Javvadi, S. & Chatterjee, S. Cell–cell signalling promotes ferric iron uptake in *Xanthomonas oryzae* pv. *oryzicola* that contribute to its virulence and growth inside rice. *Mol. Microbiol.* **96**, 708–727 (2015).
35. Pandey, S. S., Patnana, P. K., Rai, R. & Chatterjee, S. Xanthoferrin, the α -hydroxycarboxylate-type siderophore of *Xanthomonas campestris* pv. *campestris*, is required for optimum virulence and growth inside cabbage. *Mol. Plant Pathol.* **18**, 949–962 (2017).
36. Alberts, B. *et al.* *Molecular Biology of the Cell*, 4th ed. (Garland Science, 2002).
37. Aier, I., Varadwaj, P. K. & Raj, U. Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Sci. Rep.* **6**, 1–10 (2016).
38. Ivanova, L. *et al.* Molecular dynamics simulations of the interactions between glial cell line-derived neurotrophic factor family receptor GFR α 1 and small-molecule ligands. *ACS Omega* **3**, 11407–11414 (2018).
39. Batut, B., Galaxy Training Network, Taylor J, Backofen R, Nekrutenko A, Grüning B. *et al.* Community-driven data analysis training for biology. *Cell Syst.* **6**, 752–758 (2018).
40. Sneha, P. & Doss, C. G. P. Molecular dynamics: New frontier in personalized medicine. *Adv. Protein Chem. Struct. Biol.* **102**, 181–224 (2016).
41. Chikalov, I., Yao, P., Moshkov, M. & Latombe, J.-C. Learning probabilistic models of hydrogen bond stability from molecular dynamics simulation trajectories. *BMC Bioinform.* **12**, 1–6 (2011).
42. Anjum, A. *et al.* Identification of differentially expressed genes in rna-seq data of *Arabidopsis thaliana*: A compound distribution approach. *J. Comput. Biol.* **23**, 239–247 (2016).
43. Pandey, S. S., Patnana, P. K., Lomada, S. K., Tomar, A. & Chatterjee, S. Co-regulation of iron metabolism and virulence associated functions by iron and XibR, a novel iron binding transcription factor, in the plant pathogen *Xanthomonas*. *PLoS Pathog.* **12**, e1006019 (2016).
44. Zang, H.-Y. *et al.* A specific high toxicity of Xinjunan (Diocetyl diethylenetriamine) to *Xanthomonas* by affecting the iron metabolism. *Microbiol. Spectrum*. **11**, e04382–e14322 (2023).
45. Zhang, H., Wang, P. & Qi, Q. Molecular effect of FadD on the regulation and metabolism of fatty acid in *Escherichia coli*. *FEMS Microbiol. Lett.* **259**, 249–253 (2006).
46. Zhang, Y.-M. & Rock, C. O. *Biochemistry of Lipids, Lipoproteins and Membranes* 73–112 (Elsevier, 2016).
47. Müller-Hill, B. *The lac Operon* (de Gruyter, 2011).
48. Barriuso, J. *et al.* Ecology, genetic diversity and screening strategies of plant growth promoting rhizobacteria (PGPR). In *Plant-Bacteria Interactions: Strategies and Techniques to Promote Plant Growth*, 1–17 (2008).
49. Knutsen, E., Ween, O. & Håvarstein, L. S. Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. *J. Bacteriol.* **186**, 3078–3085 (2004).
50. Li, Y.-H. & Tian, X. Quorum sensing and bacterial social interactions in biofilms. *Sensors* **12**, 2519–2538 (2012).
51. Wellington, S. & Greenberg, E. P. Quorum sensing signal selectivity and the potential for interspecies cross talk. *MBio* **10**, e00146–e1119 (2019).
52. Hoang, H. T., Nguyen, T. T. T., Do, H. M., Nguyen, T. K. N. & Pham, H. T. A novel finding of intra-genus inhibition of quorum sensing in *Vibrio* bacteria. *Sci. Rep.* **12**, 1–10 (2022).
53. Ryan, R. P., An, S.-Q., Allan, J. H., McCarthy, Y. & Dow, J. M. The DSF family of cell–cell signals: An expanding class of bacterial virulence regulators. *PLoS Pathog.* **11**, e1004986 (2015).

54. Solano, C., Echeverz, M. & Lasa, I. Biofilm dispersion and quorum sensing. *Curr. Opin. Microbiol.* **18**, 96–104 (2014).
55. Ryan, R. P. & Dow, J. M. Communication with a growing family: Diffusible signal factor (DSF) signaling in bacteria. *Trends Microbiol.* **19**, 145–152 (2011).
56. He, Y. W. *et al.* Xanthomonas campestris cell–cell communication involves a putative nucleotide receptor protein Clp and a hierarchical signalling network. *Mol. Microbiol.* **64**, 281–292 (2007).
57. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets—Update. *Nucleic Acids Res.* **41**, D991–D995 (2012).
58. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
59. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
60. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
61. Bu, D. *et al.* KOBAS-i: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* **49**, W317–W325. <https://doi.org/10.1093/nar/gkab447> (2021).
62. Mering, C. V. *et al.* STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* **31**, 258–261 (2003).
63. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
64. Chin, C.-H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**, 1–7 (2014).
65. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
66. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282 (1992).
67. Zdobnov, E. M. & Apweiler, R. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
68. SchrödingerRelease. 2022-2: Maestro, Schrödinger, LLC, New York, NY. 2021. <https://www.schrodinger.com/products/maestro>.
69. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
70. Marques, S. M. *et al.* Screening of natural compounds as P-glycoprotein inhibitors against multidrug resistance. *Biomedicines* **9**, 357 (2021).
71. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
72. Miyamoto, S. & Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
73. Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).
74. Shityakov, S., Roewer, N., Förster, C. & Broscheit, J.-A. In silico investigation of propofol binding sites in human serum albumin using explicit and implicit solvation models. *Comput. Biol. Chem.* **70**, 191–197 (2017).

Acknowledgements

This work was supported by the Japan Society for the Promotion of Science under Grants-in-Aid for Scientific Research (KAKENHI) (Grant numbers 18K19674, 18KK0436, and 20H00562) and the Japan Agency for Medical Research and Development (Grant numbers 20wm0225012h0001 and 21fk0108129h0502) to F.M.. The authors gratefully acknowledge the compute resources and support provided by Prof Thomas Dandekar and the Universität Würzburg IT Center with funding from the German Research Foundation (DFG) through Grant No. INST 93/878-1 FUGG. This work was financially supported by the Government of the Russian Federation through the ITMO Fellowship and Professorship Program. The authors acknowledge the FSER-2021-0013 project and ITMO fellowship program for the infrastructure support. We thank Margaret Biswas, PhD, from Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

Author contributions

K.D., S.S. and F.M. wrote the main manuscript text and K.D. prepared tables and figures. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35487-9>.

Correspondence and requests for materials should be addressed to K.D., S.S. or F.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023