



Article

Independent Component Analysis Reveals the Transcriptional Regulatory Modules in *Bradyrhizobium diazoefficiens* USDA110

Zhi-Peng Gao, Wei-Cheng Gu, Jie Li, Qin-Tian Qiu and Bin-Guang Ma *

Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China; gaozhipen@webmail.hzau.edu.cn (Z.-P.G.); gwc@webmail.hzau.edu.cn (W.-C.G.); lijie15527048825@163.com (J.L.); qiuqintian@webmail.hzau.edu.cn (Q.-T.Q.)

* Correspondence: mbg@mail.hzau.edu.cn; Tel./Fax: +86-027-87280877

Abstract: The dynamic adaptation of bacteria to environmental changes is achieved through the coordinated expression of many genes, which constitutes a transcriptional regulatory network (TRN). *Bradyrhizobium diazoefficiens* USDA110 is an important model strain for the study of symbiotic nitrogen fixation (SNF), and its SNF ability largely depends on the TRN. In this study, independent component analysis was applied to 226 high-quality gene expression profiles of *B. diazoefficiens* USDA110 microarray datasets, from which 64 iModulons were identified. Using these iModulons and their condition-specific activity levels, we (1) provided new insights into the connection between the FixLJ-FixK₂-FixK₁ regulatory cascade and quorum sensing, (2) discovered the independence of the FixLJ-FixK₂-FixK₁ and NifA/RpoN regulatory cascades in response to oxygen, (3) identified the FixLJ-FixK₂ cascade as a mediator connecting the FixK₂-2 iModulon and the Phenylalanine iModulon, (4) described the differential activation of iModulons in *B. diazoefficiens* USDA110 under different environmental conditions, and (5) proposed a notion of active-TRN based on the changes in iModulon activity to better illustrate the relationship between gene regulation and environmental condition. In sum, this research offered an iModulon-based TRN for *B. diazoefficiens* USDA110, which formed a foundation for comprehensively understanding the intricate transcriptional regulation during SNF.

Keywords: independent component analysis; transcriptional regulatory network; functional modules; transcriptome; rhizobia; *Bradyrhizobium diazoefficiens*



Citation: Gao, Z.-P.; Gu, W.-C.; Li, J.; Qiu, Q.-T.; Ma, B.-G. Independent Component Analysis Reveals the Transcriptional Regulatory Modules in *Bradyrhizobium diazoefficiens* USDA110. *Int. J. Mol. Sci.* **2023**, *24*, 12544. <https://doi.org/10.3390/ijms241612544>

Academic Editor: Weifeng Xu

Received: 30 June 2023

Revised: 30 July 2023

Accepted: 5 August 2023

Published: 8 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bradyrhizobium diazoefficiens USDA110 (formerly known as *Bradyrhizobium japonicum* USDA110) [1] is an important model strain for the study of symbiosis. It can establish symbiotic relationship with *Glycine max*, *Vigna unguiculata*, *Vigna radiata*, *Macroptilium atropurpureum* [2], and has a strong ability of symbiotic nitrogen fixation (SNF). It is widely used in agriculture and ecological fields [3]. *B. diazoefficiens* USDA110, as a member of rhizobia, establish symbiosis with legumes, mainly including the following processes: (1) *B. diazoefficiens* USDA110 receives the signal molecules from plant roots and releases nodulation factors; (2) *B. diazoefficiens* USDA110 enters the soybean roots and differentiates into bacteroids, which are then wrapped by functional symbiotic membranes to form a low-oxygen environment; (3) *B. diazoefficiens* USDA110 expresses the related genes of nitrogenase complex and initiates the nitrogen fixation process [4–6]. Moreover, *B. diazoefficiens* USDA110 also has denitrification ability [7]. The correct implementation of this series of complex biological processes depends on the orderly expression among genes. A comprehensive understanding of the transcriptional regulatory network (TRN) of *B. diazoefficiens* USDA110 can help us uncover the molecular mechanisms related to SNF.

In recent years, there has been great progress in the omics research of *B. diazoefficiens* USDA110 [8–10]. The number of databases related to rhizobia omics is increasing, and the relevant gene annotation information and multi-omics data are gradually being completed.

Based on these data, Yang et al. established the genome-scale metabolic network of the *B. diazoefficiens* USDA110 [11] and Ma et al. constructed its protein-protein interaction network [12]. However, there are rare studies on the TRN of *B. diazoefficiens* USDA110, and related databases, such as RhizoRegNet [13], PRODORIC [14], RegPrecise [15], RegTransBase [16], and RhizoBindingSites [17], have incomplete information and lack relevant connections between gene regulation and environmental conditions.

Independent component analysis (ICA) [18] is an unsupervised machine learning algorithm, which is particularly suitable for blind source separation of mixed signals. ICA has been used in bacterial transcriptomic datasets to identify independent sets of regulated genes and their regulatory factors [19–22]. Briefly, by decomposing the gene expression matrix through ICA, we can obtain groups of genes (called iModulons) that represent independent regulatory signals. These genes may be regulated by the same or related regulators [22]. Each iModulon has a corresponding activity in different conditions, which reflects the number of iModulon genes present under that condition. Unlike regulons [23], which are obtained from experimental data by analyzing transcription factor (TF) binding sites (TFBS) in a bottom-up fashion, iModulons are calculated based on gene expression via a top-down fashion [24].

Researchers have generated several meaningful hypotheses by applying the iModulon analysis process to different species. Chauhan et al. applied the iModulon analysis process to *Sulfolobus acidocaldarius* and predicted a DNA export system composed of five uncharacterized genes [25]. Lim et al. compiled the gene expression profiles of *Pseudomonas putida* KT2440 during growth and stationary phases, used iModulons to describe the transition of transcriptional levels from the former to the latter, and revealed multiple evolutionary strategies for achieving rapid growth rate in this species [19]. Rajput et al. compiled the gene expression profiles of *Pseudomonas aeruginosa* under different nutrient conditions and antibiotic treatments, and then used ICA to analyze the iModulon structure, resulting in new roles of transcriptional regulators in antibiotic efflux regulation and the association between nutrient substrates and expression of antibiotic efflux pump genes [20]. By now, the iModulon structures of 11 species have been analyzed, and related datasets and results are present in the iModulonDB database [26].

Currently, most studies on the transcriptional regulatory relationships of *B. diazoefficiens* USDA110 were carried out by analyzing transcriptome differences in gene-deficient strains, combined with motif search for TFBS in promoter region [27–31]. Socorro Mesa et al. studied the FixLJ-FixK₂-FixK₁ cascade of *B. diazoefficiens* USDA110 and found that FixJ positively regulates FixK₂ at 0.5% oxygen concentration [32]. Analysis of the differences in bacterial transcription levels by microarray showed that 204 genes were affected by FixK₂ in *B. diazoefficiens* USDA110, and this number further increased 21 days after bacteroid formation; some of the affected genes are even not regulated by FixJ. These results indicate that the transcriptional regulation of *B. diazoefficiens* USDA110 is very complex, and that the target genes regulated by the same transcription factor may vary with environmental changes. Although some transcriptional regulatory relationships can be accurately obtained using experimental methods, they cannot provide a full picture for the transcriptional regulation of *B. diazoefficiens* USDA110 under different environmental conditions.

To better understand the TRN of *B. diazoefficiens* USDA110, we compiled a microarray dataset based on the GPL3401 microarray platform, which contains 226 *B. diazoefficiens* USDA110 transcriptome samples. Using the ICA algorithm, 64 iModulons were extracted and they were used to extend the existing regulons. We provided new insights into the relationship between the FixLJ-FixK₂-FixK₁ regulatory cascade and quorum sensing. We described the independence of the FixLJ-FixK₂-FixK₁ and NifA/RpoN cascades in response to oxygen environment. We identified the differential activation states of iModulons under different environments in *B. diazoefficiens* USDA110, and, based on this, proposed the notion of active-TRN, demonstrating the relevant connection between gene regulation and environmental condition.

2. Results

2.1. Independent Component Analysis Revealed the Independent Signal Modules of the *B. diazoefficiens* USDA110 Transcriptome

We used the ICA algorithm to decompose the gene expression matrix and obtained 64 independent components (iModulons, see Supplementary notes and Supplementary table for details). These iModulons contain a total of 2269 genes, accounting for 27.6% of the annotated genes, and explain 81.1% of the gene expression variation. To identify the features of iModulons, we first constructed the draft TRN of *B. diazoefficiens* USDA110 based on published literature and the RegPrecise database [15]. The draft TRN includes 633 regulatory relationships involving 43 regulators. Next, we compared each iModulon with known regulons and found that 13 iModulons significantly overlapped with existing regulons. We examined the degree of overlap between iModulons and regulons by calculating regulon recall (RR) and iModulon recall (MR); the degree of overlap was divided into four categories with a threshold value of 0.6 for both RR and MR. These categories were well-matched, regulon subset, regulon discovery, and poorly-matched [19]. The results are shown in Figure 1A, where the HutC iModulon shows a high degree of overlap with the known HutC regulon, indicating that it accurately captures the previously reported regulation of nine target genes by the histidine utilization repressor HutC (*blr6245*). Among these target genes, *blr6249* and *blr6250* are not in the currently known HutC regulon (Figure 1B). Through annotation information, it was found that these two genes (*blr6249*, *blr6250*) and the two genes (*blr6247*, *blr6248*) in the HutC regulon belong to the same transcriptional unit. It can be assumed that all the above four genes are regulated by HutC. Therefore, the HutC regulon has been successfully expanded by the HutC iModulon. After analyzing the upstream promoter regions of the HutC iModulon using MEME [33], three motifs were obtained (Figure S3), among which motif-2 (Figure S3B) and motif-3 (Figure S3C) were enriched in all the four promoter regions, suggesting that these motifs are potential binding sites for the HutC TF. In addition, the activity of the HutC iModulon decreased under anaerobic conditions (Figure 1C), which is consistent with previous studies [34]. Therefore, the iModulon effectively reflects gene regulatory relationships and can supplement the existing TRN.

The remaining 51 iModulons were characterized by information from public databases and literature. Adopting the scripts provided by Lim [19], we classified these iModulons into 10 functional categories based on the biological processes and functions they were involved in (Figure 1D). The FixK₂-2 iModulon, which is related to the oxygen adaptation mechanism, made the largest contribution to gene expression variances, indicating that the sense of oxygen environment is the most important factor affecting the biological functions of *B. diazoefficiens* USDA110. Other iModulons with significant contributions to the expression variances were mainly involved in SNF (NifA/RpoN iModulon and NnrR iModulon), cell mobility (Flagella iModulon), and the secretion system (T3SS iModulon) (Figure 1E), which reflects the main biological characteristics of *B. diazoefficiens* USDA110 and its physiological states under specific environmental conditions.

2.2. The FixK₂ iModulon Illustrates the Regulatory Cascade of FixK₂ and Implies Its Relationship with Quorum Sensing

As the FixK₂-2 iModulon had the greatest contribution to gene expression variances, we performed an extensive analysis to identify and characterize it. FixK₂ is a CRP/FNR-type TF that plays a central role in the adaptation of *B. diazoefficiens* USDA110 to low-oxygen environments, SNF, and denitrification processes [30,35]. FixK₂ is regulated by a two-component system FixLJ, in which FixL senses the oxygen concentration in the environment and phosphorylates FixJ under a condition of less than 5% oxygen concentration, and the phosphorylated FixJ activates FixK₂. Then, FixK₂ activates the microaerobic adaptive operon *fixNOQP* and *fixGHIS*, several heme synthesis genes (*hemA*, *hemB*, *hemN₁*, *hemN₂*), denitrification genes (*nnrR*, *napEDABC*, *nirK*, *norCBQD*, *nosRZDYFLX*), some hydrogen oxidation genes (*hup* genes) [32], and the *fixK₁* gene that is homologous to *fixK₂*. The

FixK₂-2 iModulon includes 174 genes, most of which are related to low-oxygen adaptation, and 70 genes overlap with the FixK₂ regulon, accounting for 58.3% of the FixK₂ regulon genes (Figure 2A). The activity of FixK₂-2 iModulon significantly increases in microaerobic environments and symbiotic nodulation state (Figure 2B), indicating that the FixK₂-2 iModulon of *B. diazoefficiens* USDA110 plays a key role in the adaptation of low-oxygen environment, which is consistent with the main function of FixK₂ regulon.

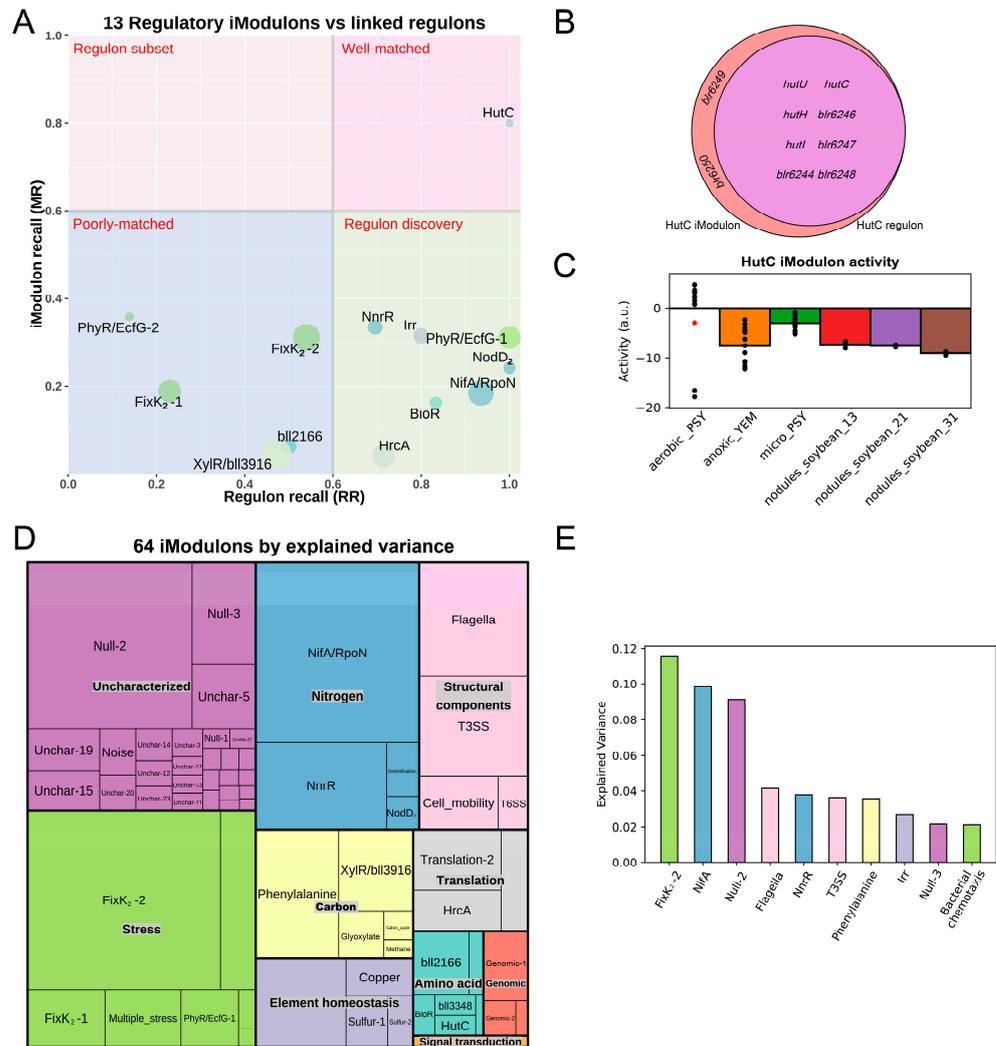


Figure 1. The compendium of Independent Component Analysis (ICA) for the *B. diazoefficiens* USDA110 transcriptome. (A) The scatter plot shows Regulon Recalls (RR) and iModulon Recalls (MR) for the 13 regulatory iModulons. (B) Comparison of genes in the reported HutC regulon and the HutC iModulon. (C) HutC iModulon activities under anoxic conditions. Asterisks indicate the reference condition (aerobic with PSY media). The x-axis represents different environmental conditions: “aerobic_PSY” represents *B. diazoefficiens* cultured in PSY medium under aerobic conditions; “anoxic_YEM” represents its cultivation in YEM medium under anoxic conditions; “micro_PSY” represents the cultivation environment in PSY medium under microaerobic conditions; “nodules_soybean_13”, “nodules_soybean_21”, and “nodules_soybean_31” represent 13 days, 21 days, and 31 days after the establishment of symbiosis, respectively. (D) A treemap of *B. diazoefficiens* USDA110’s 64 iModulons. The size of each rectangle indicates the explained variances in the data sets attributable to each iModulon. (E) Ten iModulons with the highest explained variances.

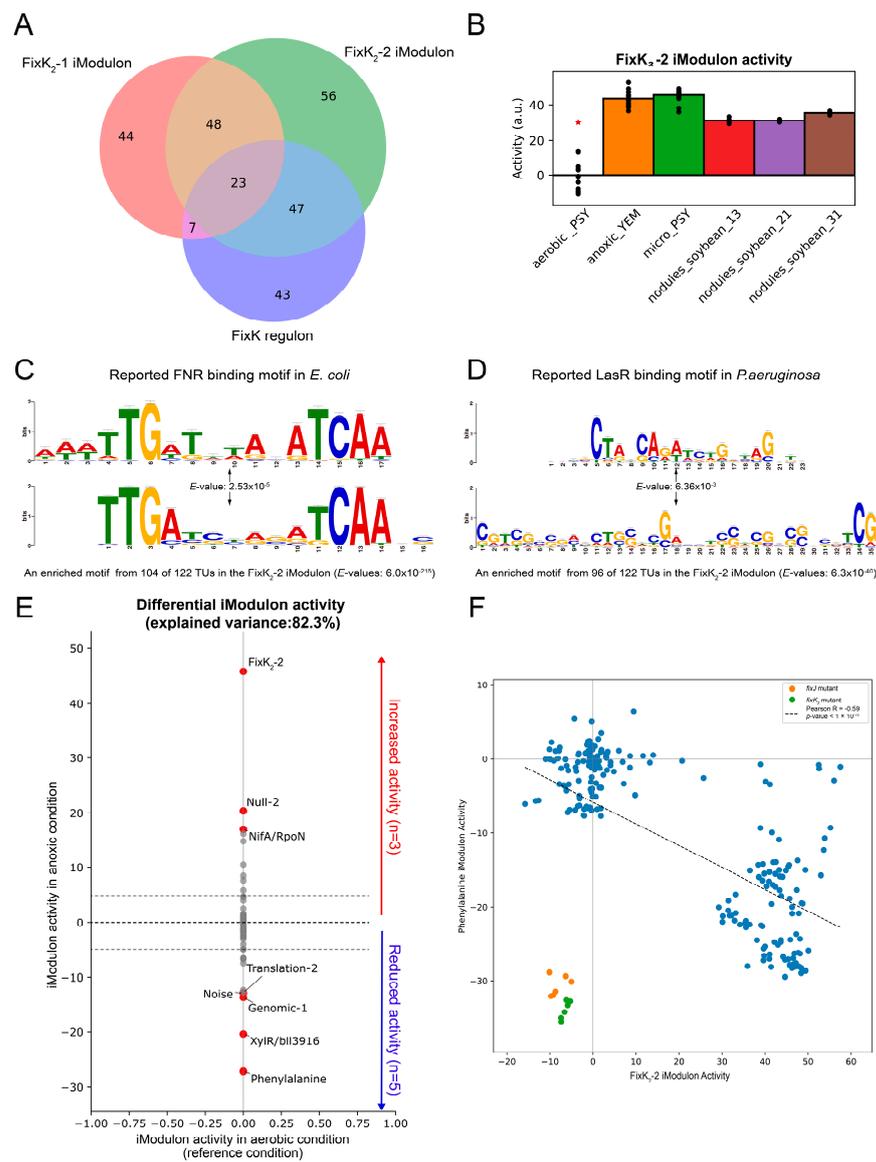


Figure 2. Overview of the FixK₂ iModulon. **(A)** Comparison of genes in the reported FixK regulon, FixK₂-1 iModulon and FixK₂-2 iModulon. **(B)** FixK₂-2 iModulon activities under anoxic condition. The x-axis represents different environmental conditions: “aerobic_PSY” represents *B. diazoefficiens* cultured in PSY medium under aerobic conditions; “anoxic_YEM” represents its cultivation in YEM medium under anoxic conditions; “micro_PSY” represents the cultivation environment in PSY medium under microaerobic conditions; “nodules_soybean_13”, “nodules_soybean_21”, and “nodules_soybean_31” represent 13 days, 21 days, and 31 days after the establishment of symbiosis, respectively. Asterisks indicate the reference condition (aerobic with PSY media). **(C)** Comparison of a motif found in the upstream regions of the FixK₂-2 iModulon genes and the previously known binding motif of FNR in *Escherichia coli*. **(D)** Comparison of a motif found in the upstream regions of the FixK₂-2 iModulon genes and the previously known binding motif of LasR in *P. aeruginosa*. **(E)** Differentially changed iModulon activities between anoxic condition and aerobic condition (threshold = 5 and *FDR* = 0.05). The red dots represent the iModulons with significant differences in activity; the grey dots represent the iModulons with no significant difference in activity. **(F)** Correlation of activities between the FixK₂-2 iModulon and the Phenylalanine iModulon. The orange dots represent the Phenylalanine iModulon and FixK₂-2 iModulon activities of the *fixJ* mutant strain. The green dots represent the Phenylalanine iModulon and FixK₂-2 iModulon activities of the *fixK₂* mutant strain. The blue dots represent the Phenylalanine iModulon and FixK₂-2 iModulon activities of the rest samples in different conditions.

Through MEME enrichment analysis for the upstream sequence of all transcriptional units of the FixK₂-1 and FixK₂-2 iModulons member genes, we identified a motif that is highly consistent with the known FixK₂ binding site (Figure 2C). Since FixK₂ belongs to the Fnr family, this binding site is highly similar to the Fnr binding site of *Escherichia coli*. The FixK₂-1 iModulon involves 84 operons, 47 of which contain this motif; the FixK₂-2 iModulon involves 122 operons, 104 of which contain this motif. Though comparison with the FixK₂ regulon, we found 148 genes that do not belong to the FixK₂ regulon (Figure 2A), and these genes may be potential targets of FixK₂, which requires further validation. It is worth noting that the activities of the FixK₂-1 and FixK₂-2 iModulons significantly decreased in *fixJ* and *fixK₂* mutant strains (Figure S4), indicating that *fixJ* and *fixK₂* are the key TFs regulating the gene expression in FixK₂-1 and FixK₂-2 iModulons.

By comparing the activities of the FixK₂-1 and FixK₂-2 iModulons, we found that the activity of the FixK₂-2 iModulon was significantly higher than that of the FixK₂-1 iModulon in the *blr1880* project. After the deletion of the *blr1880* gene, the FixK₂-1 iModulon activity was significantly decreased (Figure S5). Blr1880 is a LuxR family TF and this family of proteins is usually associated with quorum sensing [36]. We hypothesized that the difference in activity between the FixK₂-1 and FixK₂-2 iModulons is due to the quorum sensing function mediated by *blr1880*. To verify this hypothesis, we annotated the enriched motifs using TOMTOM and found that a common motif, “CGTCCGCSGHSTSSWCGABVKSSTCGMSSHSVTTCG”, in the FixK₂-1 and FixK₂-2 iModulons is similar to the LasR TFBS motif in *Pseudomonas aeruginosa* (*p*-values of 1.08×10^{-5} and 1.75×10^{-5} , respectively) (Figure 2D). The LasR TF belongs to the LuxR protein family and regulates quorum sensing in *Pseudomonas aeruginosa* [37]. This suggests that these sites are potential binding sites for *blr1880*, indicating that FixK₂ and *blr1880* have interrelated regulatory functions. Given that FixK₂ is regulated by oxygen and *blr1880* primarily participates in quorum sensing, it is worthwhile investigating whether the oxygen concentration also impacts quorum sensing in *B. diazoefficiens* USDA110. By examining the gene composition of the FixK₂-1 and FixK₂-2 iModulons, we found that the FixK₂-2 iModulon contains more genes related to denitrification, including *napABCDE* and *nnrR*, while the FixK₂-1 iModulon does not. This indicated that the FixK₂-2 iModulon better represents the FixLJ/FixK₂/NnnR cascade, while the FixK₂-1 iModulon better represents the *blr1880*-involved quorum sensing. Overall, these two iModulons clearly demonstrate that FixK₂ is involved in anaerobic adaptation and denitrification processes in *B. diazoefficiens* USDA110 and imply a potential link between the genes involved in microaerobic adaptation and those involved in quorum sensing. This provides a new clue for further understanding the relationship between these biological processes.

Next, by comparing the activities of iModulons under aerobic and anaerobic conditions, 8 iModulons with significant differences were obtained, among which 3 iModulons were significantly up-regulated and 5 iModulons were significantly down-regulated. The largest activity changes are shown in FixK₂-2 iModulon and Phenylalanine iModulon (Figure 2E). The Phenylalanine iModulon contains 25 genes, and the genes with high weights mainly include hydroxypropanoate isomerase (*hlyi*), aldehyde dehydrogenase (*bll4784*), and cytochrome c-type (*blr7488*), which play crucial roles in carbohydrate metabolism and energy production. Therefore, the decreased activity of Phenylalanine iModulon indicates that carbohydrate metabolism is suppressed, resulting in reduced energy production and inhibition of basic cellular metabolism under anaerobic conditions. In most cases, the activities of FixK₂-2 iModulon and Phenylalanine iModulon are negatively correlated significantly (Figure 2F), while this negative correlation disappeared in the *fixJ* mutant and the *fixK₂* mutant, indicating that FixLJ/FixK₂ may act as a mediator connecting the FixK₂-2 and Phenylalanine iModulons.

2.3. The *NifA/RpoN* iModulon Successfully Captures Nitrogen-Fixing Gene Cluster

NifA is one of the most important TFs in *B. diazoefficiens* USDA110, and its target genes include the *nif* gene cluster directly involved in nitrogenase synthesis, as well as

the *fixRnifA* operon, which is subject to NifA-dependent autoregulation under low-oxygen conditions. It also regulates genes involved in SNF such as *fixA*, *fixBCX*, *nif*, *fix*, *hup* operons, *groESL*, *fdxN*, and *rpoN*, as well as genes with unknown functions such as *nrgA* and *nrgBC* [38,39]. Studies have shown that NifA is also a necessary gene for the maximal expression of denitrification genes in *B. diazoefficiens* USDA110 [7]. NifA cooperates with RNA polymerase containing sigma 54 to activate gene expression. In *B. diazoefficiens* USDA110, sigma 54 is encoded by two highly similar and functionally equivalent genes (*rpoN₁* and *rpoN₂*) [40]. The NifA/RpoN iModulon includes 158 genes, most of which are located in a region approximately 2 Mb from the start site of DNA replication, and this region mainly contains genes directly related to SNF (Figure 3A). By enriching the upstream promoter regions of the NifA/RpoN iModulon using MEME, a motif highly similar to the *Vibrio cholerae* RpoN TFBS motif was found (Figure 3B). Therefore, the above results are consistent with the functional characteristics of NifA.

Two oxygen-responsive regulatory cascades (FixLJ-FixK₂ and RegSR-NifA) exist in *B. diazoefficiens* USDA110. The activation status of the NifA/RpoN iModulon under different oxygen conditions is displayed in Figure 3C. It can be seen that the NifA/RpoN iModulon is activated under anaerobic conditions, and it is no longer activated when *regR*, *nifA*, or *rpoN* is knocked out. This indicates that the NifA/RpoN iModulon can effectively represent the regulatory function of the RegSR-NifA cascade.

By comparing the iModulon activity differences between free-living and symbiotic conditions, 13 significantly different iModulons were identified, including 7 significantly up-regulated and 6 significantly down-regulated iModulons (threshold = 5, FDR = 0.01). It can be seen that the NifA/RpoN and Flagella iModulons exhibited the largest activity changes (Figure 3D). The Flagella iModulon contains 46 genes, mainly including functional genes involved in flagella synthesis and assembly, which play a crucial role in the motility of *B. diazoefficiens* USDA110. Therefore, the decreased activity of this module suggests that under symbiotic conditions, the expression of flagella genes is suppressed and the motility is reduced, which is consistent with the function of *B. diazoefficiens* USDA110 in developing into a bacteroid.

2.4. Differential iModulon Activation of *B. diazoefficiens* USDA110 under Different Environmental Conditions

Over the past 20 years, transcriptome studies of rhizobia have been grouped according to specific stages of the symbiotic process analyzed: bacteroids vs. free-living bacteria, anaerobic vs. aerobic conditions, rhizosphere or flavonoid compounds vs. non-induced cultures, and denitrifying vs. anaerobic conditions [9]. We first compared the iModulon activities of *B. diazoefficiens* USDA110 under bacteroids and free-living aerobic conditions (threshold = 5, FDR = 0.05) and identified 38 significantly different iModulons. Then we investigated the activity changes of these iModulons during different stages of SNF. As shown in Figure 4, there were significant differences in the activation status of these iModulons at different stages of SNF. Under free-living aerobic conditions, when *B. diazoefficiens* USDA110 was induced by genistein for 8 h, the Flagella iModulon responsible for flagella formation and assembly and the NodD₂ iModulon related to nodulation factor were significantly activated, indicating that *B. diazoefficiens* USDA110 activated the motility for root recruitment and the ability to transmit nodulation signals to the plant. Under free-living microaerobic conditions, the NifA/RpoN and FixK₂-2 iModulons were significantly activated, indicating that the anaerobic environment activated nitrogen fixation structural genes and related regulatory genes, while the Flagella iModulon was in a suppressed state, indicating that the rhizobia no longer needed the motility provided by flagella when nitrogen fixation genes were activated. Twenty-one days after nodulation, more iModulons, including T3SS iModulon and T6SS iModulon, were activated when compared with those under free-living anaerobic conditions, indicating that, at this stage, the bacteroids had the capability of extracellular secretion and transport to adapt to the substance exchange demands between the bacterial cells and the plant host. These three different environmental

conditions correspond to different stages of SNF of *B. diazoefficiens* USDA110. The activity changes of these different iModulons revealed the modular activation of gene functions in response to environmental changes during SNF.

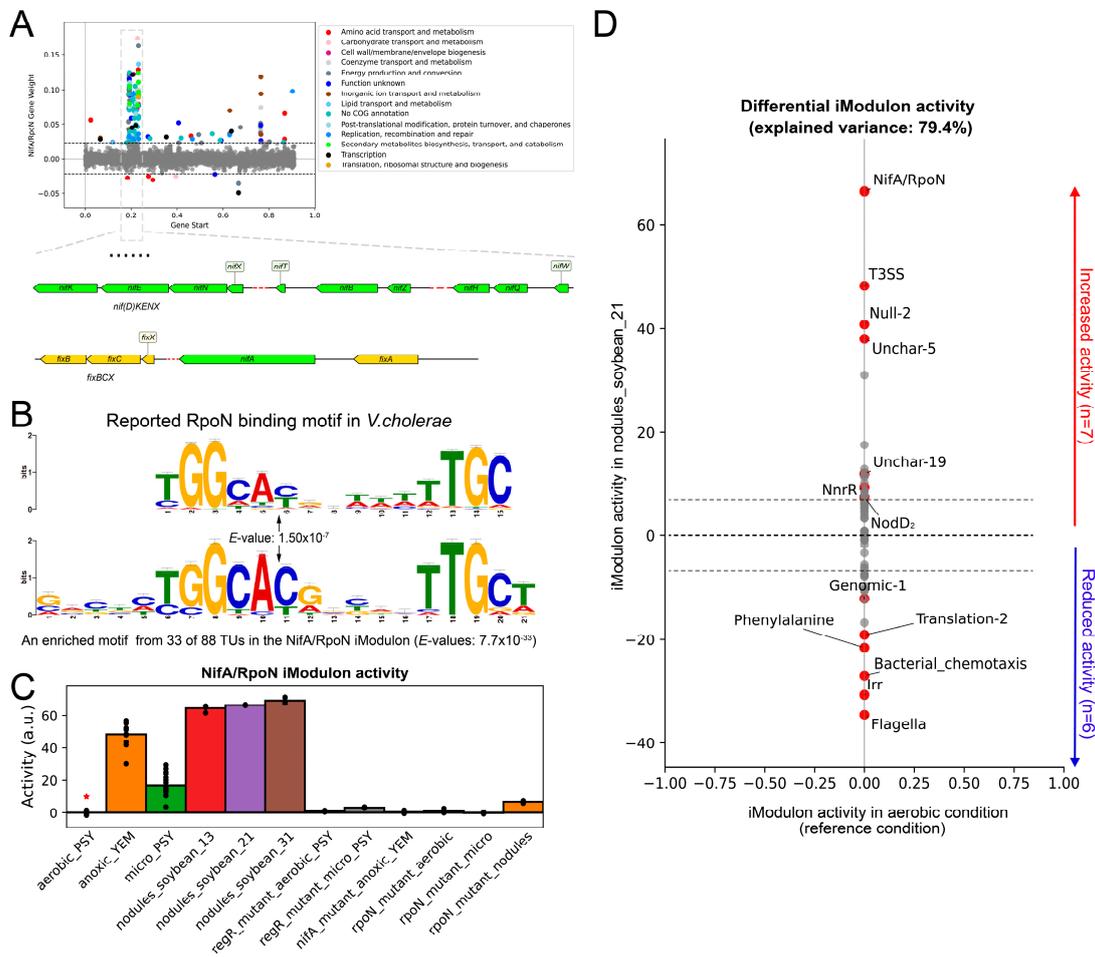


Figure 3. The NifA/RpoN iModulon successfully captures the nitrogen-fixing gene cluster. **(A)** Scatter plot of the NifA/RpoN iModulon gene weights. Genes are colored by COG categories. **(B)** Comparison of a motif found in the upstream regions of the NifA/RpoN iModulon genes with the previously known binding motif of RopN in *V. cholerae*. **(C)** NifA/RpoN iModulon activities under anoxic condition. The x-axis represents different environmental conditions: “aerobic_PSY” represents *B. diazoefficiens* cultured in PSY medium under aerobic conditions; “anoxic_YEM” represents its cultivation in YEM medium under anoxic conditions; “micro_PSY” represents the cultivation environment in PSY medium under microaerobic conditions; “nodules_soybean_13”, “nodules_soybean_21”, and “nodules_soybean_31” represent 13 days, 21 days, and 31 days after the establishment of symbiosis, respectively; “regR_mutant_aerobic_PSY” and “regR_mutant_micro_PSY” represent *regR* mutant strain cultured in PSY medium under aerobic conditions and microaerobic conditions, respectively; “nifA_mutant_anoxic_YEM” represents *nifA* mutant strain cultured in YEM medium under anoxic conditions; “rpoN_mutant_aerobic” and “rpoN_mutant_micro” represent *rpoN* mutant strain cultured in PSY medium under aerobic conditions and microaerobic conditions, respectively; “rpoN_mutant_nodules” represents the condition of established symbiosis between *rpoN* mutant strain and legumes. Asterisks indicate the reference condition (aerobic with PSY media). **(D)** Differentially changed iModulon activities between symbiosis condition and aerobic condition (threshold = 5 and *FDR* = 0.01). The red dots represent the iModulons with significant differences in activity; the grey dots represent the iModulons with no significant difference in activity.

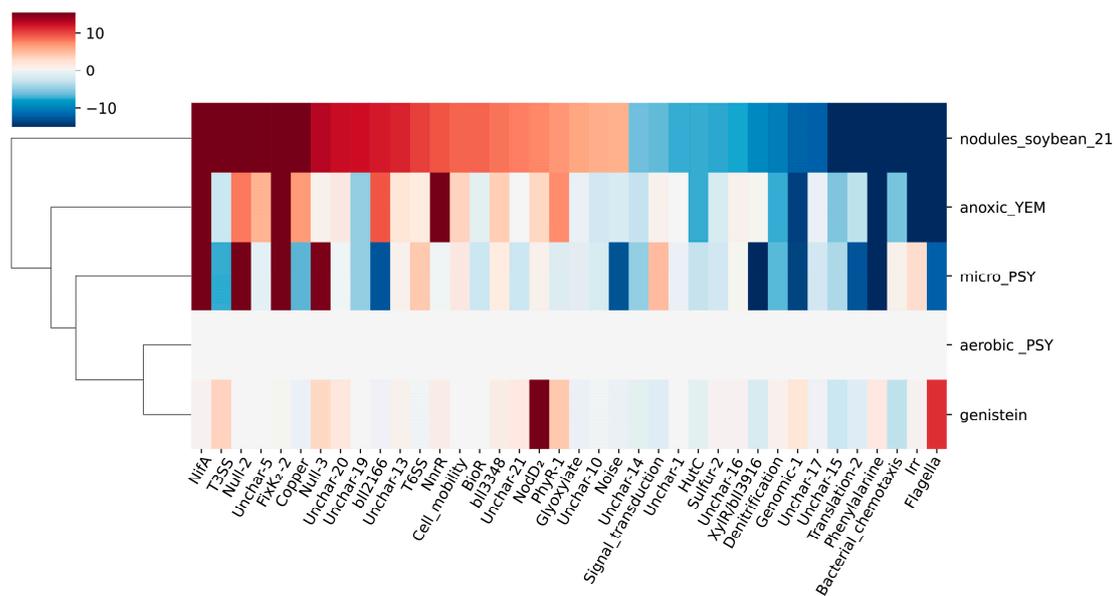


Figure 4. iModulon activity during different stages of symbiosis. Thirty-eight key iModulons were selected by comparing the iModulon activities of *B. diazoefficiens* USDA110 under bacteroid and free-living aerobic conditions (threshold = 5, *FDR* = 0.05). “aerobic_PSY” represents the free-living condition; “genistein” represents the state of *B. diazoefficiens* USDA110 being recruited by the plant; “micro_PSY” represents the middle stage of symbiosis; “nodules_soybean_21” represents the late stage of symbiosis; “anoxic_YEM” represents denitrification.

2.5. Active-TRN Based on iModulon Activity

Life activities rely on the orderly regulation of gene expression, and the regulatory relationships among genes constitute the TRN. To determine the regulatory relationship between genes required methods such as electrophoretic mobility shift assay, DNase I footprinting assay, and in vitro run-off transcription assay to verify whether TFs directly bind to the upstream promoter regions of target genes. Although these methods produced accurate results, they were time-consuming and labor-intensive, and only a small number of regulatory relationships could be captured in one experiment. High-throughput technologies such as transcriptomics, proteomics, and chromatin immunoprecipitation sequencing, combined with computational analysis methods, can more easily obtain a global transcriptional regulatory landscape of a species. As is known to all, most regulatory relationships are condition-dependent, which means that the regulatory effects of TFs vary with environmental conditions [41]. However, the current research on TRN often depicts static TRNs that do not distinguish between regulatory activation states, and thus cannot describe the transcriptional regulatory relationships under different environmental conditions. To overcome this drawback, this study analyzed the regulatory effects of TFs in *B. diazoefficiens* USDA110 under four different conditions: genistein induction, microaerobiosis, bacteroid, and denitrification, based on the activity characteristics of iModulons.

As shown in Figure 5, the active transcriptional regulatory network (active-TRN) induced by genistein contains 59 edges and 11 nodes (Figure 5A), accounting for 9.1% of the total regulatory relationships, indicating that only a small number of key pathways are activated in *B. diazoefficiens* USDA110 after receiving the genistein signal. The active-TRN under microaerobic condition contains 312 edges and 23 nodes (Figure 5B), mainly including the FixLJ-FixK₂-FixK regulatory cascade and the NifA regulon related to low-oxygen adaptation, which is consistent with previous studies and with the related iModulon results analyzed above. The active-TRN under symbiotic conditions contains 364 edges and 30 nodes (Figure 5C), with 22 nodes appearing in both microaerobic and symbiotic conditions, and the remaining differentially expressed nodes reflect the condition-specific activation of regulatory relationships under symbiotic conditions. The active-TRN under

denitrification conditions contains 358 edges and 24 nodes (Figure 5D), with 21 nodes overlapping with the nodes under microaerobic condition. These results demonstrated the regulatory role of TFs in *B. diazoefficiens* USDA110 under different environmental conditions.

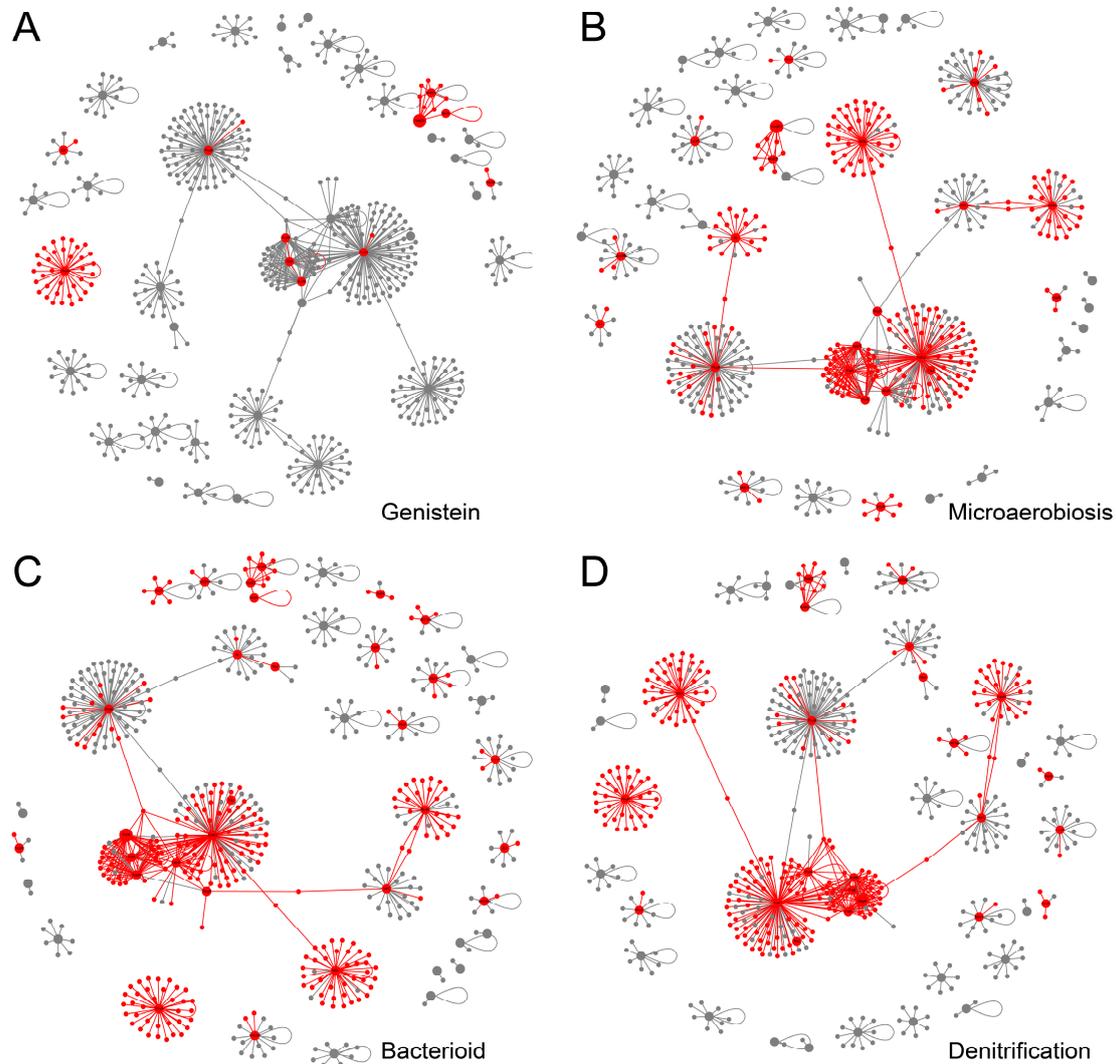


Figure 5. The active-TRNs based on iModulon activities. Red nodes represent active genes; grey nodes represent inactive genes. (A) TRN activation after genistein induction. (B) TRN activation in microaerobiosis conditions. (C) TRN activation 21 days after symbiotic nodulation. (D) TRN activation under denitrification conditions.

3. Discussion

In this study, we used ICA to decompose 226 different transcriptomic datasets of *B. diazoefficiens* USDA110 into 64 iModulons. Many of these iModulons correspond to important TFs in *B. diazoefficiens* USDA110. Using these iModulons, we provided new insights into the relationship between the FixLJ-FixK₂-FixK₁ regulatory cascade and quorum sensing. We described the independence of FixLJ-FixK₂-FixK₁ and NifA/RpoN regulatory cascades in response to oxygen environments. We calculated the differential activation status of iModulons in *B. diazoefficiens* USDA110 under different environmental conditions, and based on this, we proposed the active-TRNs, which demonstrated the relationship between gene regulation and environmental condition.

Up to now, there are only fifteen samples from three studies of *B. diazoefficiens* USDA110 RNA-seq data in the GEO database, which is much smaller than the amount of microarray data and is not sufficient for ICA analysis. Previous studies have found that ICA performs

poorly on microarray data [42]. However, we obtained good results this time because we used data from the same microarray platform, and most of the data came from the same institution, thus avoiding noise and errors introduced by differences in microarray data processing methods.

Using the ICA method, we obtained 64 iModulons from the transcriptome expression matrix of *B. diazoefficiens* USDA110, of which 13 iModulons significantly overlapped with regulons, reflecting the known regulatory relationships. Although the remaining iModulons cannot be directly associated with known regulatory relationships, most of their features can be supplemented through gene functional annotation. Among these 13 regulatory iModulons, only the HutC iModulon had a high degree of overlap with the HutC regulon, possibly due to incomplete information in the TRN of *B. diazoefficiens* USDA110. The sizes of the iModulons we obtained were almost all larger than those of the corresponding regulons, indicating that an iModulon may contain multiple regulons with similar expression patterns. Through motif analysis (Supplementary data file), we discovered potential TFBSs, and the relationships between these binding sites and TFs need further verification. Although the Null-2 iModulon in our results does not contain any gene (namely, no gene has a weight above the threshold), its contribution to expression variance ranks third, which means we cannot ignore this “strange” independent signal. We found that the gene weight distribution in the Null-2 iModulon was relatively average, which means that no gene significantly contributed to this independent signal. We found that the Null-2 iModulon activity significantly increased in anaerobic environments, alkaline environments, high temperature environments, and starvation environments, indicating that the Null-2 iModulon responded to extreme environments. In particular, the Null-2 iModulon activity was significantly higher in starvation than in other environmental conditions. This indicates that many genes are involved in the adaptation of the organism to extreme environments and these genes have diverse expression patterns.

By comparing the differences between the FixK₂-1 iModulon and FixK₂-2 iModulon, we found that the FixK₂-1 iModulon is closely related to quorum sensing, while the FixK₂-2 iModulon is more representative of the FixLJ/FixK₂/NnnR cascade. Under stress conditions, the activity of the FixK₂-1 iModulon increased, suggesting a potential connection between quorum sensing genes and the stress resistance mechanism of *B. diazoefficiens* USDA110. Studies have shown that the LuxR homologous genes in *Sinorhizobium meliloti* contribute to its adaptation to environmental challenges, such as high osmotic pressure, nutrient starvation, and nodule competition, thereby increasing its chances of survival in the stressful rhizosphere environment [43]. Whether the LuxR homologous gene *blr1880* in *B. diazoefficiens* has similar functions requires further investigation. By checking the activity correlations between the FixK₂-2 iModulon and the Phenylalanine iModulon, we found that the FixLJ/FixK₂ TFs serve as a bridge connecting FixK₂-2 iModulon and Phenylalanine iModulon. It was found that the *bll0330* gene in Phenylalanine iModulon is directly regulated by FixK₂, and the protein product of *bll0330* gene was a two-component response regulator. This suggests that the FixLJ/FixK₂ TFs may negatively regulate the activity of the Phenylalanine iModulon by directly regulating *bll0330*, which may involve new regulatory cascades and requires further experimental verification.

The FixLJ-FixK₂ cascade and the RegSR-NifA cascade are relatively independent but interrelated regulatory cascades in *B. diazoefficiens* USDA110 in response to oxygen concentration [38], which are reflected in iModulon activity. Under low oxygen conditions, both the FixLJ-FixK₂ and NifA/RpoN iModulons show high activity. However, when *fixJ* or *fixK₂* is absent, FixK₂ iModulon activity decreases, but NifA/RpoN iModulon activity is not affected. Similarly, when *regR* or *nifA* or *rpoN* is absent, NifA/RpoN iModulon activity decreases, but FixLJ-FixK₂ iModulon activity is not affected, reflecting the independence of the two regulatory cascades in response to oxygen environments. We found that the LuxR homologous gene *blr1880* seems to play a special role in regulating FixLJ-FixK₂ iModulon activity. When *blr1880* is absent, the FixLJ-FixK₂ iModulon remains activated under aerobic conditions, while the NifA/RpoN iModulon is inhibited. This indicates that

blr1880 responds to the inhibitory effect of aerobic environments on the FixK₂ iModulon activity by participating in the FixLJ-FixK₂ cascade, but the specific mechanism still needs further investigation.

We compared the activation status of iModulons under different environmental conditions and found significant differences in activities, which were consistent with the functional requirements of adaptation to the corresponding environments. In *B. diazoefficiens* USDA110, genistein induced the activation of NodD₂ and Flagella iModulons, but did not affect other nitrogen fixation structural genes. Although nitrogen fixation-related modules were significantly activated under anaerobic conditions, additional iModulons were required for complete SNF function. Under anaerobic conditions with YEM medium, denitrification function of *B. diazoefficiens* USDA110 was induced. We found that the NifA/RpoN iModulon and FixK₂₋₂ iModulon still exhibited high activities, indicating that oxygen concentration is a direct factor regulating the activities of these two iModulons. These results demonstrate the perfect coordination between gene expression and environmental adaptation in *B. diazoefficiens* USDA110, as well as the relative independence of its gene functional modules, which may find its application in Synthetic Biology: when the biological functions from *B. diazoefficiens* USDA110 are needed, the independence of gene functional modules benefits their implementation in other chassis organisms.

In the past decades, a series of methods have been developed to construct gene regulatory networks. The existing inference methods are based on different principles [44], and the rise of machine learning has opened new possibilities for inferring TRNs. It should be noted that living organisms have stimulus-response coupling mechanisms in order to perceive and respond to changes in different environmental conditions, such as the bacterial one-component and two-component regulatory systems. These systems activate downstream regulatory cascades only after receiving specific environmental signals. However, most of the existing TRNs cannot describe the corresponding relationship between environmental condition and TF regulation. Therefore, in this study, we proposed the active-TRNs under different environmental conditions based on iModulon activity data. We assumed that when the iModulon activity significantly increased, the regulatory relationships involved in it were activated. We first identified the iModulons with significantly increased activity under different environmental conditions, and then extracted the corresponding genes and regulatory relationships. It should be noted that we did not make further screening based on changes in TF expression, because the TFs of two-component regulatory systems do not directly depend on their gene expression changes but on the proteins that sense environment signals in the system. The PyModulon software package provides an interface for inferring iModulon activity in a new dataset using a precalculated IcaData object, so that, with our modified scripts, people can easily calculate the significantly activated iModulons in the new dataset and obtain the active-TRNs under different environmental conditions. Our results demonstrated the TF regulatory relationships activated under different environmental conditions and extended the application of iModulon activity data.

4. Materials and Methods

4.1. Data Collection and Preprocessing

We downloaded all the microarray data of Affymetrix custom *Bradyrhizobium japonicum* strain USDA110 19K array (Platform GPL3401) from the GEO database, which covers 23 series. Samples with poor replicability (sample correlation coefficient < 0.95) were removed, and the final dataset contained 226 samples involving 54 unique experimental conditions. The median Pearson correlation coefficient between sample replicates is 0.96 (Figure S1A). We performed log₂ transformation on the data. In order to minimize the noise caused by batch effects among samples, only the samples from two institutions (ETH Zurich and TU Dresden) were used in the dataset (Figure S1B). The data have been verified by Principal Components Analysis (PCA) and have high self-consistency between the samples.

At the same time, we also selected a reference condition for each project, which was usually the control group in the experiment design. Following a previous study [45], we normalized the data by subtracting the mean value of the corresponding reference condition from all expression data in the project. Sastry et al. released the procedure for analyzing iModulons (<https://github.com/avsastry/modulome-workflow> (accessed on 27 March 2023) [46], and we referred to their public code and made some modifications for our research. Our scripts and related datasets are available at <https://github.com/mbglab/BdTRNiMod>.

4.2. Independent Component Analysis

Following the PyModulon workflow, we used the well-organized dataset as input and the scripts provided by Sastry [46] (with modifications) to decompose the expression matrix into two matrices, A and M [47]. Briefly, the scikit-learn (v0.23.2) [48] implementation of FastICA [49] was performed 100 times with a convergence tolerance of 10^{-7} in order to obtain robust independent components (ICs). The resulting independent components (ICs) were clustered using DBSCAN [50] to identify robust ICs, with an epsilon of 0.1 and minimum cluster seed size of 50. To account for identical components with opposite signs, the following distance metric was used for computing the distance matrix: $d_{x,y} = 1 - \|\rho_{x,y}\|$, where $\rho_{x,y}$ is the Pearson correlation coefficient between components x and y . The final robust ICs were defined as the centroids of the cluster. We performed clustering of the gene expression profile multiple times for dimensions between 20 and 215 with a step size of 5. The optimal dimension of 210 was finally chosen (Figure S2).

4.3. Characterization of iModulons

We obtained 64 iModulons and determined the iModulon gene sets using functions provided by PyModulon. The RefSeq ID of *B. diazoefficiens* USDA110 genome was BA000040.2, and the gene annotation came from the NCBI Prokaryotic Genome Annotation Pipeline. In addition, KEGG [51] and Cluster of Orthologous Groups (COG) [52] information was obtained using EggNOG mapper [53]. Operon information was obtained from BioCyc [54]. The gene ontology (GO) annotation came from AmiGO2 [55]. The draft TRN of *B. diazoefficiens* USDA110 was constructed based on the RegPrecise database [15] and our manually collected data from the literature. To determine the overlap between iModulons and regulons, we performed TF enrichment analysis, and the relevant method was described as in previous studies [45]. We used “iModulon recall” and “regulon recall” values to evaluate the degree of overlap between predicted iModulons and regulons. “iModulon recall” represents the proportion of shared genes and genes in an iModulon, while “regulon recall” is the proportion of shared genes and genes in a regulon [22]. Common motifs in the upstreams of iModulons were enriched by using the find_motifs function that uses MEME [33] with a maximum window of 30 bp. When motifs were compared, the web-based TOMTOM [56] was utilized. As the statistical significance, E -values were provided as adjusted p -values when more than one common sequence was identified or compared.

4.4. Active-TRN Construction

We constructed active-TRN based on the differences in iModulon activity. First, we extracted the activity for each iModulon from matrix A under the aerobic condition of free-living *B. diazoefficiens* USDA110 as the reference state, and then identified the activity difference between the reference state and the target environmental condition by comparing the distributions of corresponding iModulon activities with a two-sample Kolmogorov–Smirnov test. Then, we selected the iModulons that showed a significant increase in activity under the target environmental condition. Finally, we extracted the regulatory relationships involved in these iModulons to construct the active-TRN.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241612544/s1>. Reference [47] is cited in the supplementary materials.

Author Contributions: Conceptualization, B.-G.M. and Z.-P.G.; methodology, Z.-P.G. and J.L.; software, Z.-P.G. and J.L.; validation, W.-C.G., Q.-T.Q. and J.L.; formal analysis, Z.-P.G. and W.-C.G.; investigation, Z.-P.G. and W.-C.G.; resources, B.-G.M., Z.-P.G. and W.-C.G.; data curation, Z.-P.G., W.-C.G., Q.-T.Q. and J.L.; writing—original draft preparation, Z.-P.G.; writing—review and editing, B.-G.M. and Q.-T.Q.; visualization, Z.-P.G. and W.-C.G.; supervision, B.-G.M.; project administration, B.-G.M.; funding acquisition, B.-G.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant 31971184).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and codes involved in this paper are available at the following website: <https://github.com/mbglab/BdTRNiMod>.

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

References

1. Delamuta, J.R.M.; Ribeiro, R.A.; Ormeño-Orrillo, E.; Melo, I.S.; Martínez-Romero, E.; Hungria, M. Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int. J. Syst. Evol. Microbiol.* **2013**, *63 Pt 9*, 3342–3351. [[CrossRef](#)] [[PubMed](#)]
2. Gottfert, M.; Grob, P.; Hennecke, H. Proposed regulatory pathway encoded by the *nodv* and *nodw* genes, determinants of host specificity in *Bradyrhizobium japonicum*. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2680–2684. [[CrossRef](#)] [[PubMed](#)]
3. Peoples, M.B.; Craswell, E.T. Biological nitrogen fixation: Investments, expectations and actual contributions to agriculture. *Plant Soil* **1992**, *141*, 13–39. [[CrossRef](#)]
4. Dixon, R.; Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2004**, *2*, 621–631. [[CrossRef](#)]
5. Roy, S.; Liu, W.; Nandety, R.S.; Crook, A.; Mysore, K.S.; Pislariu, C.I.; Frugoli, J.; Dickstein, R.; Udvardi, M.K. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. *Plant Cell* **2020**, *32*, 15–41. [[CrossRef](#)]
6. Montiel, J.; Downie, J.A.; Farkas, A.; Bihari, P.; Herczeg, R.; Balint, B.; Mergaert, P.; Kereszt, A.; Kondorosi, E. Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5041–5046. [[CrossRef](#)]
7. Bueno, E.; Mesa, S.; Sanchez, C.; Bedmar, E.J.; Delgado, M.J. NifA is required for maximal expression of denitrification genes in *Bradyrhizobium japonicum*. *Environ. Microbiol.* **2010**, *12*, 393–400. [[CrossRef](#)]
8. Gomes, D.F.; Batista, J.S.D.; Rolla, A.A.P.; da Silva, L.P.; Bloch, C.; Galli-Terasawa, L.V.; Hungria, M. Proteomic analysis of free-living *Bradyrhizobium diazoefficiens*: Highlighting potential determinants of a successful symbiosis. *BMC Genom.* **2014**, *15*, 643. [[CrossRef](#)]
9. Jiménez-Guerrero, I.; Acosta-Jurado, S.; Del Cerro, P.; Navarro-Gómez, P.; López-Baena, F.J.; Ollero, F.J.; Vinardell, J.M.; Pérez-Montaño, F. Transcriptomic Studies of the Effect of nod Gene-Inducing Molecules in Rhizobia: Different Weapons, One Purpose. *Genes* **2018**, *9*, 1. [[CrossRef](#)]
10. Lardi, M.; Murset, V.; Fischer, H.M.; Mesa, S.; Ahrens, C.H.; Zamboni, N.; Pessi, G. Metabolomic Profiling of *Bradyrhizobium diazoefficiens*-Induced Root Nodules Reveals Both Host Plant-Specific and Developmental Signatures. *Int. J. Mol. Sci.* **2016**, *17*, 815. [[CrossRef](#)] [[PubMed](#)]
11. Yang, Y.; Hu, X.P.; Ma, B.G. Construction and simulation of the *Bradyrhizobium diazoefficiens* USDA110 metabolic network: A comparison between free-living and symbiotic states. *Mol. Biosyst.* **2017**, *13*, 607–620. [[CrossRef](#)] [[PubMed](#)]
12. Ma, J.X.; Yang, Y.; Li, G.; Ma, B.G. Computationally Reconstructed Interactome of *Bradyrhizobium diazoefficiens* USDA110 Reveals Novel Functional Modules and Protein Hubs for Symbiotic Nitrogen Fixation. *Int. J. Mol. Sci.* **2021**, *22*, 11907. [[CrossRef](#)] [[PubMed](#)]
13. Krol, E.; Blom, J.; Winnebald, J.; Berhorster, A.; Barnett, M.J.; Goesmann, A.; Baumbach, J.; Becker, A. RhizoRegNet-A database of rhizobial transcription factors and regulatory networks. *J. Biotechnol.* **2011**, *155*, 127–134. [[CrossRef](#)] [[PubMed](#)]
14. Dudek, C.A.; Jahn, D. PRODORIC: State-of-the-art database of prokaryotic gene regulation. *Nucleic Acids Res.* **2022**, *50*, D295–D302. [[CrossRef](#)]
15. Novichkov, P.S.; Kazakov, A.E.; Ravcheev, D.A.; Leyn, S.A.; Kovaleva, G.Y.; Sutormin, R.A.; Kazanov, M.D.; Riehl, W.; Arkin, A.P.; Dubchak, I.; et al. RegPrecise 3.0-A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genom.* **2013**, *14*, 745. [[CrossRef](#)]
16. Cipriano, M.J.; Novichkov, P.N.; Kazakov, A.E.; Rodionov, D.A.; Arkin, A.P.; Gelfand, M.S.; Dubchak, I. RegTransBase—A database of regulatory sequences and interactions based on literature: A resource for investigating transcriptional regulation in prokaryotes. *BMC Genom.* **2013**, *14*, 213. [[CrossRef](#)]

17. Taboada-Castro, H.; Castro-Mondragon, J.A.; Aguilar-Vera, A.; Hernandez-alvarez, A.J.; van Helden, J.; Encarnacion-Guevara, S. RhizoBindingSites, a Database of DNA-Binding Motifs in Nitrogen-Fixing Bacteria Inferred Using a Footprint Discovery Approach. *Front. Microbiol.* **2020**, *11*, 567471. [[CrossRef](#)]
18. Jutten, C.; Héroult, J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Process.* **1991**, *24*, 1–10. [[CrossRef](#)]
19. Lim, H.G.; Rychel, K.; Sastry, A.V.; Bentley, G.J.; Mueller, J.; Schindel, H.S.; Larsen, P.E.; Laible, P.D.; Guss, A.M.; Niu, W.; et al. Machine-learning from *Pseudomonas putida* KT2440 transcriptomes reveals its transcriptional regulatory network. *Metab. Eng.* **2022**, *72*, 297–310. [[CrossRef](#)] [[PubMed](#)]
20. Rajput, A.; Tsunemoto, H.; Sastry, A.V.; Szubin, R.; Rychel, K.; Chauhan, S.M.; Pogliano, J.; Palsson, B.O. Advanced transcriptomic analysis reveals the role of efflux pumps and media composition in antibiotic responses of *Pseudomonas aeruginosa*. *Nucleic Acids Res.* **2022**, *50*, 9675–9688. [[CrossRef](#)]
21. Rychel, K.; Sastry, A.V.; Palsson, B.O. Machine learning uncovers independently regulated modules in the *Bacillus subtilis* transcriptome. *Nat. Commun.* **2020**, *11*, 6338. [[CrossRef](#)] [[PubMed](#)]
22. Sastry, A.V.; Gao, Y.; Szubin, R.; Hefner, Y.; Xu, S.B.; Kim, D.; Choudhary, K.S.; Yang, L.; King, Z.A.; Palsson, B.O. The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **2019**, *10*, 5536. [[CrossRef](#)] [[PubMed](#)]
23. Mejia-Almonte, C.; Busby, S.J.W.; Wade, J.T.; van Helden, J.; Arkin, A.P.; Stormo, G.D.; Eilbeck, K.; Palsson, B.O.; Galagan, J.E.; Collado-Vides, J. Redefining fundamental concepts of transcription initiation in bacteria. *Nat. Rev. Genet.* **2020**, *21*, 699–714. [[CrossRef](#)] [[PubMed](#)]
24. Yuan, Y.; Seif, Y.; Rychel, K.; Yoo, R.; Chauhan, S.; Poudel, S.; Al-bulushi, T.; Palsson, B.O.; Sastry, A.V. Pan-Genome Analysis of Transcriptional Regulation in Six *Salmonella enterica* Serovar Typhimurium Strains Reveals Their Different Regulatory Structures. *Msystems* **2022**, *7*, e0046722. [[CrossRef](#)]
25. Chauhan, S.M.; Poudel, S.; Rychel, K.; Lamoureux, C.; Yoo, R.; Al Bulushi, T.; Yuan, Y.; Palsson, B.O.; Sastry, A.V. Machine Learning Uncovers a Data-Driven Transcriptional Regulatory Network for the Crenarchaeal Thermoacidophile *Sulfolobus acidocaldarius*. *Front. Microbiol.* **2021**, *12*, 753521. [[CrossRef](#)]
26. Rychel, K.; Decker, K.; Sastry, A.V.; Phaneuf, P.V.; Poudel, S.; Palsson, B.O. iModulonDB: A knowledgebase of microbial transcriptional regulation derived from machine learning. *Nucleic Acids Res.* **2021**, *49*, D112–D120. [[CrossRef](#)]
27. Jiménez-Leiva, A.; Cabrera, J.J.; Bueno, E.; Torres, M.J.; Salazar, S.; Bedmar, E.J.; Delgado, M.J.; Mesa, S. Expanding the Regulon of the *Bradyrhizobium diazoefficiens* NnrR Transcription Factor: New Insights Into the Denitrification Pathway. *Front. Microbiol.* **2019**, *10*, 1926. [[CrossRef](#)]
28. Lindemann, A.; Moser, A.; Pessi, G.; Hauser, F.; Friberg, M.; Hennecke, H.; Fischer, H.M. New target genes controlled by the *Bradyrhizobium japonicum* two-component regulatory system RegSR. *J. Bacteriol.* **2007**, *189*, 8928–8943. [[CrossRef](#)]
29. Nishihata, S.; Kondo, T.; Tanaka, K.; Ishikawa, S.; Takenaka, S.; Kang, C.M.; Yoshida, K. *Bradyrhizobium diazoefficiens* USDA110 PhaR functions for pleiotropic regulation of cellular processes besides PHB accumulation. *BMC Microbiol.* **2018**, *18*, 156. [[CrossRef](#)]
30. Parejo, S.; Cabrera, J.J.; Jiménez-Leiva, A.; Tomás-Gallardo, L.; Bedmar, E.J.; Gates, A.J.; Mesa, S. Fine-tuning modulation of oxidation-mediated posttranslational control of *Bradyrhizobium diazoefficiens* FixK₂ transcription factor. *Int. J. Mol. Sci.* **2022**, *23*, 5117. [[CrossRef](#)] [[PubMed](#)]
31. Torres, M.J.; Argandona, M.; Vargas, C.; Bedmar, E.J.; Fischer, H.M.; Mesa, S.; Delgado, M.J. The Global Response Regulator RegR Controls Expression of Denitrification Genes in *Bradyrhizobium japonicum*. *PLoS ONE* **2014**, *9*, e99011. [[CrossRef](#)]
32. Mesa, S.; Hauser, F.; Friberg, M.; Malaguti, E.; Fischer, H.M.; Hennecke, H. Comprehensive assessment of the Regulons controlled by the FixLJ-FixK₂-FixK₁ cascade in *Bradyrhizobium japonicum*. *J. Bacteriol.* **2008**, *190*, 6568–6579. [[CrossRef](#)]
33. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME Suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)] [[PubMed](#)]
34. Goldberg, R.B.; Hanau, R. Regulation of *Klebsiella pneumoniae* hut operons by oxygen. *J. Bacteriol.* **1980**, *141*, 745–750. [[CrossRef](#)]
35. Cabrera, J.J.; Jiménez-Leiva, A.; Tomás-Gallardo, L.; Parejo, S.; Casado, S.; Torres, M.J.; Bedmar, E.J.; Delgado, M.J.; Mesa, S. Dissection of FixK₂ protein-DNA interaction unveils new insights into *Bradyrhizobium diazoefficiens* lifestyles control. *Environ. Microbiol.* **2021**, *23*, 6194–6209. [[CrossRef](#)]
36. Bogino, P.C.; Nievas, F.L.; Giordano, W. A review: Quorum sensing in *Bradyrhizobium*. *Agric. Ecosyst. Environ. Appl. Soil Ecol.* **2015**, *94*, 49–58. [[CrossRef](#)]
37. Kiratisin, P.; Tucker, K.D.; Passador, L. LasR, a transcriptional activator of *Pseudomonas aeruginosa* virulence genes, functions as a multimer. *J. Bacteriol.* **2002**, *184*, 4912–4919. [[CrossRef](#)] [[PubMed](#)]
38. Nienaber, A.; Huber, A.; Gottfert, M.; Hennecke, H.; Fischer, H.M. Three new NifA-regulated genes in the *Bradyrhizobium japonicum* symbiotic gene region discovered by competitive DNA-RNA hybridization. *J. Bacteriol.* **2000**, *182*, 1472–1480. [[CrossRef](#)] [[PubMed](#)]
39. Tsoy, O.V.; Ravcheev, D.A.; Čuklina, J.; Gelfand, M.S. Nitrogen Fixation and Molecular Oxygen: Comparative Genomic Reconstruction of Transcription Regulation in Alphaproteobacteria. *Front. Microbiol.* **2016**, *7*, 1343. [[CrossRef](#)]
40. Kullik, I.; Fritsche, S.; Knobel, H.; Sanjuan, J.; Hennecke, H.; Fischer, H.-M. *Bradyrhizobium japonicum* has two differentially regulated, functional homologs of the σ_{54} gene (rpoN). *J. Bacteriol.* **1991**, *173*, 1125–1138. [[CrossRef](#)]
41. Chen, Y.H.; Yang, C.D.; Tseng, C.P.; Huang, H.D.; Ho, S.Y. GeNOSA: Inferring and experimentally supporting quantitative gene regulatory networks in prokaryotes. *Bioinformatics* **2015**, *31*, 2151–2158. [[CrossRef](#)] [[PubMed](#)]
42. Sastry, A.V.; Hu, A.; Heckmann, D.; Poudel, S.; Kavvas, E.; Palsson, B.O. Independent component analysis recovers consistent regulatory signals from disparate datasets. *PLoS Comp. Biol.* **2021**, *17*, e1008647. [[CrossRef](#)] [[PubMed](#)]

43. Patankar, A.V.; Gonzalez, J.E. An Orphan LuxR Homolog of *Sinorhizobium meliloti* Affects Stress Adaptation and Competition for Nodulation. *Appl. Environ. Microbiol.* **2009**, *75*, 946–955. [[CrossRef](#)]
44. Marbach, D.; Costello, J.C.; Kuffner, R.; Vega, N.M.; Prill, R.J.; Camacho, D.M.; Allison, K.R.; Kellis, M.; Collins, J.J.; Stolovitzky, G.; et al. Wisdom of crowds for robust gene network inference. *Nat. Methods* **2012**, *9*, 796–804. [[CrossRef](#)] [[PubMed](#)]
45. Yoo, R.; Rychel, K.; Poudel, S.; Al-bulushi, T.; Yuan, Y.; Chauhan, S.; Lamoureux, C.; Palsson, B.O.; Sastry, A. Machine Learning of All *Mycobacterium tuberculosis* H37Rv RNA-seq Data Reveals a Structured Interplay between Metabolism, Stress Response, and Infection. *Mosphere* **2022**, *7*, e0003322. [[CrossRef](#)] [[PubMed](#)]
46. Sastry, A.V.; Poudel, S.; Rychel, K.; Yoo, R.; Lamoureux, C.R.; Chauhan, S.; Haiman, Z.B.; Al Bulushi, T.; Seif, Y.; Palsson, B.O. Mining all publicly available expression data to compute dynamic microbial transcriptional regulatory networks. *bioRxiv* **2021**, Preprint.
47. McConn, J.L.; Lamoureux, C.R.; Poudel, S.; Palsson, B.O.; Sastry, A.V. Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC Bioinform.* **2021**, *22*, 584. [[CrossRef](#)]
48. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
49. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* **1999**, *10*, 626–634. [[CrossRef](#)]
50. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; pp. 226–231.
51. Kanehisa, M.; Furumichi, M.; Sato, Y.; Ishiguro-Watanabe, M.; Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **2021**, *49*, D545–D551. [[CrossRef](#)]
52. Galperin, M.Y.; Wolf, Y.I.; Makarova, K.S.; Alvarez, R.V.; Landsman, D.; Koonin, E.V. COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **2021**, *49*, D274–D281. [[CrossRef](#)] [[PubMed](#)]
53. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [[CrossRef](#)] [[PubMed](#)]
54. Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. [[CrossRef](#)]
55. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [[CrossRef](#)] [[PubMed](#)]
56. Gupta, S.; Stamatoyannopoulos, J.A.; Bailey, T.L.; Noble, W.S. Quantifying similarity between motifs. *Genome Biol.* **2007**, *8*, R24. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.