Inter-cluster competition and resource partitioning may govern the ecology of Frankia

I. Sarkar1,3 · G. Sen1 · S. Bhattacharyya2 · M. Gtari4 · A. Sen1,2,3

Received: 18 December 2021 / Revised: 9 April 2022 / Accepted: 10 April 2022 / Published online: 16 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Microbes live in a complex communal ecosystem. The structural complexity of microbial community reflects diversity, functionality, as well as habitat type. Delineation of ecologically important microbial populations along with exploration of their roles in environmental adaptation or host–microbe interaction has a crucial role in modern microbiology. In this scenario, reverse ecology (the use of genomics to study ecology) plays a pivotal role. Since the co-existence of two different genera in one small niche should maintain a strict direct interaction, it will be interesting to utilize the concept of reverse ecology in this scenario. Here, we exploited an ‘R’ package, the RevEcoR, to resolve the issue of co-existing microbes which are proven to be a crucial tool for identifying the nature of their relationship (competition or complementation) persisting among them. Our target organism here is Frankia, a nitrogen-fixing actinobacterium popular for its genetic and host-specific nature. According to their plant host, Frankia has already been sub-divided into four clusters C-I, C-II, C-III, and C-IV. Our results revealed a strong competing nature of CI Frankia. Among the clusters of Frankia studied, the competition index between C-I and C-III was the largest. The other interesting result was the co-occurrence of C-II and C-IV groups. It was revealed that these two groups follow the theory of resource partitioning in their lifestyle. Metabolic analysis along with their differential transporter machinery validated our hypothesis of resource partitioning among C-II and C-IV groups.

Keywords Reverse ecology · R · Resource partitioning · Frankia · Competition index · Complementation index

Introduction
The term adaptation is often considered as the cornerstone of ecology and evolutionary study (Hugenholtz et al. 1998; Friedman and Alm, 2012; Lassalle et al. 2015). Assorted traits allow an organism to survive in a specific habitat as well as shaping the interface between environment and genome helps them to endure natural selection pressure (Hugenholtz et al. 1998). Any living organism has to face two specific types of competition, i.e., intra-species competition and inter-species competition. However, ecological interaction does not only include competition, there is complementation (mutualism, commensalism) too. Complementation or cooperation synergy is also important when assessing the interaction among different organisms inhabiting in a specific niche. Permanent residents of a niche are thus the end product or “entangled bank” of all three selective forces (Gilbert et al. 2009; Duffy et al. 2007; Ings et al. 2009; Faust and Raes 2012; Berry and Widder 2014). Deciphering this relationship among the myriad of co-inhabiting strains is a daunting task. Ecological dynamics and cross-talk in a specific niche is potentially driven by the interactions among inhabiting species (Berlow et al. 2009; Dunne 2006; Poisot et al. 2015; Kéfi et al. 2012). Network analysis comes in handy in this aspect (Zhang 2011; Suzuki et al 2017). An important property of this type of network is its non-static nature,
i.e., these networks constantly change on the arrival of new nodes and links (Barabasi and Albert 1999; Barabasi 2009).

However, a comprehensive study on ecological genomics needs a thorough understanding of genomic machinery regulating the response of a species to an environmental clue (Levy and Borenstein 2012). The traditional approach was to identify an ecologically adaptive phenotype first and detect associated genetic variations (Amann et al. 1995). With the advancement of functional genomics and system biology, a new model popularly known as Reverse Ecology has been derived.

*Frankia* is well known for its nitrogen-fixing ability while remaining associated with actinorhizal plants (Sellstedt and Richau 2013; Berry et al. 2011). Actinorhizal plants are pioneer plants growing in the preliminary stages of succession. Those plants may disappear in seral stages of succession. However, they are responsible for improving the soil quality and increasing the soil micronutrient level which later can be used by other plants and soil microbes in the successive seral stages. *Frankia* contributes to the fitness of their host plants by providing nitrogenous compounds (Sellstedt Richau 2013).

This symbiosis has been found worldwide with a broad range of ecological and environmental conditions mainly in poor and trivial-productive soils (i.e., saline sea-soil, landslide area, and soil of volcanic eruption) (Schwencke and Carú, 2001; Benson and Dawson, 2007; Sen et al. 2014; Simonet et al 1986). Diversity also exists among *Frankia* as well as their host plants (Cronquist 1968; Schwintzer 2012) (Supplementary Fig. 1). The phylogeny of *Frankia* allows them to be distributed into four distinct clades chiefly based on their association with host plants. Clade (C)-I, II, and III contained nodulating nitrogen fixers. The C-I infect Myricaceae, Betulaceae, and Casuarinaceae families (except *Gymnostoma*), while Coriariaceae, Dryadoideae (all actinorhizal Rosaceae), Datiscaeaceae, and *Ceanothus* (Rhamnaceae) are infected by C-II *Frankia*. Most promiscuous C-III *Frankia* associates with Colletiaeae (all actinorhizal Rhamnaceae except *Ceanothus*), Elaeagnaceae, Myricaceae, Casuarinaceae (only *Gymnostoma*), and occasionally *Alnus* (Normand et al. 2007); however, the fourth Clade (C-IV) possessed “atypical” strains that are either unable to re-infect actinorhizal host plants or are impotent of fixing N2 in the nodules (Normand et al. 2007). Interestingly, the ineffective C-IV strains have been isolated from *Ceanothus* root nodules along with C-III *Frankia*, and both were present on the outer cover of the nodules. Thus, this phylogeny arises four distinct questions about *Frankia* host specificity and categorically *Frankia* ecology. However, there are still some missing links that exist in the host specificity nature of *Frankia*. For instance, a) How C-I and C-III *Frankia* are sharing the same family of host plants? b) Is it possible that C-II and C-IV share the same root nodule or at least the same location of a root nodule together? c) Why *Ceanothus* can only be infected with C-II and not by C-III members? d) Why Gymnostoma can only be infected with C-III and not by C-I *Frankia*?

To answer these questions, we have taken up an in-silico approach which will virtually throw light on the ecological aspect of *Frankia*.

**Materials and methods**

**Genome selection and sequence retrieval**

A total of 44 *Frankia* genomes whose complete genome sequences are available in the IMG database were selected for this study (Supplementary Table 1). The protein-coding gene sequences and translated amino acid sequences of select strains were downloaded from the IMG database (https://img.jgi.doe.gov/). A whole-genome-based phylogenetic tree was generated for getting a clear view of the clustering pattern of considered *Frankia*. A whole-genome-based phylogen-eny tree was generated through the Type (Strain) Genome Server (TYGS) (https://tygs.dsmz.de) (Meier-Kolthoff and Göker 2019). The TYGS sub-divided the whole genome-based phylogeny generation into a pair-wise comparison of genome sequences using Genome Blast Distance Phylogeny (GBDP) and accurate inter-genomic distances calculation by 'trimming' and distance formula (Meier-Kolthoff et al. 2013). 100 distance replicates were used for that analysis. Digital DNA–DNA hybridization (DDH) scores and confidence intervals were calculated through GGDC2.1 server (Meier-Kolthoff and Göker 2019). Intergenomic distance-based phylogeny was generated through a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 (Lefort et al. 2015). Branch point support was inferred from 100 bootstrap replicates per branch. The phylogeny was generated using the default parameters. Heatmaps with species cluster, sub-species cluster, G+C amount, delta statistics, the genome size (bp), and protein content were generated through the TYGS server using default parameters.

**Metabolic pathway prediction**

All the strains included in this study are still not there in the KEGG database (Kanehisa 2002). Hence, to keep the uniformity, we went for predicting the metabolic pathways for all the studied strains. KEGG Automatic Annotation Server (KAAS)
program (Cao et al. 2016) on R-studio (Allaire 2012). This reverse ecology analysis was done using the RevEcoR program in R. Seed sets were generated with the command ‘getSeedSets’ and ‘seed.set’. Finally, competition and cooperation index were calculated with the “complementarity index(net[[1]],net[[2]], seed.sets[[1]], seed.sets[[2]])” and “competition index(net[[1]],net[[2]], seed.sets[[1]], seed.sets[[2]])” with default parameters. The result was visualized by heatmaps generated with R for better visualization and easy understanding.

Reverse ecology analysis

Reverse ecology analysis was done using the RevEcoR program (Cao et al. 2016) on R-studio (Allaire 2012). This method conjectures the ecology of an organism from its genomic feature (Levy and Borenstein 2012; Borenstein et al. 2008). Several tools like NetSeed (Carr et al. 2012) and NetCorporate (Levy et al. 2015) were developed for understanding the interaction among species and their environment. Nevertheless, they did not support the metabolic reconstruction of organisms and were also limited to small-scale analysis (Cao et al. 2016). Much of these problems have been taken care of by a recently developed R-based program called RevEcoR (Cao et al 2016) which utilized the reverse ecological framework for reconstructing the metabolic network of an organism and predicting species interaction in terms of competition and complementation.

This analysis takes into account all the metabolic network data (in terms of KEGG ontology or KO) for all studied organisms and calculates competition and complementation indices. More shared KO id predicts more competition, whereas less shared KO id predicts more complementation. First, we obtained the KO profiles of all considered Frankia and were fed into RevEcoR program in R. Seed sets were generated with the command ‘getSeedSets’ and ‘seed.set’. Finally, competition and cooperation index were calculated with the “complementarity index(net[[1]],net[[2]], seed.sets[[1]], seed.sets[[2]])” and “competition index(net[[1]],net[[2]], seed.sets[[1]], seed.sets[[2]])” with default parameters. The result was visualized by heatmaps generated with R for better visualization and easy understanding.

Results and discussion

Competing behavior of Clad-I (C-I) Frankia

The whole-genome-based phylogeny divided the Frankia genus into four distinct groups (Supplementary Fig. 2). The major group was C-I with 23 strains and the smallest group was C-II with four strains. C-III and C-IV contained 11 and 6 strains, respectively. Results obtained from the reverse ecology study were analyzed as per the categorization obtained from this whole-genome-based phylogeny. It was found that competition was more among different Frankia clusters than complementation or mutualism. Intra-cluster mutualism was greater than inter-cluster mutualism (Supplementary file 1).

The reverse ecology analysis established C-I Frankia as a general adversary against all other clusters, especially C-III and C-IV, indicating their antagonistic nature (Supplementary file 1). The complementation index between C-I and C-III varied from 0.00 to 0.56, whereas the competition index ranged from 0.55 to 0.99. C-I and C-III share host plants (Myricaceae and Casuarinaceae) and may have tight competition for utilizing the host resources. However, Gymnostoma (another Casuarinaceae) is infected only by C-III (but not C-I) (Normand et al. 2007; Ngom et al. 2016) maybe because of strong competition from C-III against C-I.

Exceptional cases of surface contamination

Some C-III Frankia strains like R43 (Pujic et al. 2015), CcI149 (Mansour et al. 2017), and G2 (Nouioui et al. 2016) were indeed isolated from Casuarina root nodules; however, they were not clustered into the C-I group, because they were incapable of infecting the Casuarina roots, but were capable of infecting the Elaeagnaceae group. Based on that, they were confirmed to be Cluster III Frankia. Vemulapally et al. (2019) worked on R43 and confirmed that this strain was isolated from Casuarina nodules as a surface contaminant. Similarly, CcI149 and G2 are also ineffective toward Casuarina and Alnus, but can infect Elaeagnus and Hippophäe (Mansour et al. 2017; Nouioui et al. 2016). Their spore physiology, pigment patterns are different from C-I Frankia strains (Mansour et al. 2017). Although detailed experimental analysis has not been done on CcI149 and G2, there is a high probability that those two strains were also isolated from Casuarina root nodules as a result of surface contamination, as well. The reverse ecology analysis also showed distinct competition and less complementation between C-I and C-III clusters which supports the experimental results of
Vemulapally et al. (2019) indicating less mutualism and more competition among C-I and C-III Frankia. This result also reiterates host differentiation between these two groups. Similar antagonistic nature was found between C-I and C-II supporting their host differentiation nature. Overall, C-I Frankia was found to be a common competitor against the other three clusters. It is possible that due to this competition, C-I Frankia remains restricted to the Hamamelidaceae subclass of actinorhizal plants and cannot infect other potent subclasses (Normand et al. 2007; Ben et al. 2018), i.e., Rosidae, Dilleniidae, and Magnoliidae, which are the main host plants for C-II, C-III, and C-IV.

C-III members infect C-I-specific host plant families due to shared metabolic pathways

It has been reported in previous studies (Normand et al. 2007; Ngom et al. 2016; Benson et al. 2004) that C-III Frankia can infect Eleagnaceae, Rhamnaceae, Myricaceae, Betulaceae (only Alnus sp.), and Casuarinaceae (only Gymnostoma sp.). Besides, Alnus and members of Myricaceae can also be infected by C-I Frankia (Normand et al. 2007; Ngom et al. 2016; Benson et al. 2004). However, reverse ecology analysis revealed strong competition between them. We found a high competition index value among C-I and C-III Frankia (Supplementary File 1- sheet name ‘competition_CI_CIII’) and a low complementation value among them (Supplementary File 1- sheet name ‘complementation_CI_CIII’). The competition index between C-I and C-III Frankia ranged from 0.55 to 0.99, whereas the complementation was quite low (0.00–0.5 approximately).

A similar metabolic strategy may be one of the reasons for a similar host association. A comparative study on metabolic pathways revealed 213 shared (common) pathways between these two aforementioned groups (Supplementary file 2). This indicates that both C-I and C-III Frankia can use the same nutrients for their overall metabolic activities. Moreover, a study on the nodulation signaling of Frankia proposed the presence of C-I-specific NIN factor (nodulation inducing factor) among C-III members (38) which supports our hypothesis that both C-I and C-III use similar resources as well as a similar signaling cascade (at least to some extent) which resulted in the fact that both C-III and C-I can infect Myrica and Alnus.

Resource partitioning among C-II and C-IV strains

Previous studies have reported the isolation of C-II and C-IV strains from a single nodule (Amann et al. 1995). However, our present analysis on reverse ecology revealed more competition and less complementation between these two groups. From the competition index (Supplementary Fig. 3a, 3b) among C-II and C-IV stains, we see that the intra-cluster competition among C-II members ranged from 0.35 to 0.48, and for C-IV, the values ranged from 0.31 to 0.49. However, the inter-cluster competition between C-II and C-IV was higher (0.69–0.87) than the intra-cluster competition. Hence, a distinct competition between C-II and C-IV was evident. Moreover, when we compared the complementation index of C-II and C-IV strains, intra-cluster complementation was more than inter-cluster complementation. These raise the question that how two excessively competing groups can share the same nodule at a time and grow successfully in them?

To understand the ecological aspect lying behind this highly fascinating result, we looked at comprehensive metabolic pathways of those clusters (C-II and C-IV) (Table 1; comparative metabolic pathways among four Frankia clusters provided in Supplementary file 2). Most of the metabolic pathways showed a difference between C-II and C-IV strains; however, a major disparity was observed in aromatic amino acids metabolism and carbohydrate metabolism (Table 1). It is a well-established fact that both aromatic amino acids and carbohydrates, being the major source of carbon and nitrogen, are crucial aspects of nodulation and nitrogen fixation (Borenstein et al. 2008, Carr et al. 2012). A discrepancy in metabolism among C-II and C-IV strains could be a possible reason for the fact that these strains share the same host and same nodules irrespective of their high contest, since they utilize different sources of energy. This supports an ecological paradigm termed niche partitioning or more specifically resource partitioning (Levy et al. 2015) which states that competing species can co-exist using the same environment differently. Moreover, when we compared the transporter families among C-II and C-IV Frankia (Fig. 1), it revealed prominent differences. Only 1.5% of genes were shared among C-II and C-IV Frankia strains. This further validates our hypothesis of resource partitioning between C-II and C-IV. More interestingly, there were very few similarities among the four clusters of Frankia. Only 8.2% of transporters (Fig. 1) are shared among all four Frankia clusters which support their host-specific nature largely.

Extraordinary cases of Gymnostoma and Ceanothus

Although, there were more than 200 shared metabolic pathways between C-I and C-III Frankia (discussed earlier), a total of 67 metabolic pathways were detected that were present in the C-III group but not in C-I. Furthermore, the nif cluster of C-III also showed a distinct distinction between them and C-I, i.e., nifV gene was located at distance from other nif genes in C-III, whereas in C-I, all the nif genes were clustered together. A similar finding was also reported earlier (Normand et al. 2007). All these factors may cumulatively affect the infection cycle of C-III Frankia in Gymnostoma.
Table 1 Comparative account of metabolic pathways between C-II and C-IV *Frankia* strains. X indicates the presence of the specific metabolic pathway.

| Pathway class                                                                 | C-IV | C-II |
|------------------------------------------------------------------------------|------|------|
| Biosynthesis—amino acid biosynthesis                                         |      |      |
| β-alanine biosynthesis III                                                   | X    |      |
| l-asparagine biosynthesis I                                                  | X    |      |
| l-asparagine biosynthesis II                                                 | X    |      |
| l-cysteine biosynthesis I                                                    | X    |      |
| l-cysteine biosynthesis III (from L-homocysteine)                            |      | X    |
| l-ornithine biosynthesis I                                                   | X    |      |
| Biosynthesis—carbohydrate biosynthesis                                       |      |      |
| dTDP-4-O-demethyl-β-L-noviose biosynthesis                                   | X    |      |
| Heptadecane biosynthesis                                                    | X    |      |
| Biosynthesis—cofactor, prosthetic group, electron carrier, and vitamin biosynthesis |      |      |
| 1,4-Dihydroxy-2-naphthoate biosynthesis                                      | X    |      |
| N10-Formyl-tetrahydrofolate biosynthesis                                     | X    |      |
| Biotin biosynthesis from 8-amino-7-oxononanoate I                            |      | X    |
| Factor 420 polyglutamylation                                                 | X    |      |
| Folate transformations I                                                     | X    |      |
| Heptaprenyl diphosphate biosynthesis                                         | X    |      |
| NAD biosynthesis from 2-amino-3-carboxymuconate semialdehyde                 | X    |      |
| Tetrahydrofolate salvage from 5,10-methenyltetrahydrofolate                  | X    |      |
| Thiazole biosynthesis II (aerobic bacteria)                                  | X    |      |
| Ubiquinol-8 biosynthesis (prokaryotic)                                       | X    |      |
| Biosynthesis—fatty acid and lipid biosynthesis                              |      |      |
| Cis-vaccenate biosynthesis                                                   | X    |      |
| CDP-diacylglycerol biosynthesis II                                           | X    |      |
| Phosphatidylethanolamine biosynthesis I                                      | X    |      |
| Biosynthesis—metabolic regulator biosynthesis                               |      |      |
| Glucosylglycerate biosynthesis I                                            | X    |      |
| Mannosylglucosylglycerate biosynthesis I                                     | X    |      |
| Biosynthesis—other biosynthesis                                              |      |      |
| l-dopachrome biosynthesis                                                   |      | X    |
| Degradation/utilization/assimilation—amine and polyamine degradation        |      |      |
| 4-Aminobutanoate degradation II                                             | X    |      |
| 4-Aminobutanoate degradation III                                            | X    |      |
| N-Acetylglucosamine degradation I                                           | X    |      |
| Creatinine degradation I                                                     | X    |      |
| Urea degradation I                                                           | X    |      |
| Degradation/utilization/assimilation—aromatic compound degradation          |      |      |
| 2,2'-Dihydroxybiphenyl degradation                                          | X    |      |
| 2-Chlorobenzene degradation                                                  | X    |      |
| 4,5-Dichlorocatechol degradation                                             | X    |      |
| 4-Chloronitrobenzene degradation                                             | X    |      |
| 4-Nitrotoluene degradation II                                               | X    |      |
| Benzene degradation                                                         | X    |      |
| Biphenyl degradation                                                        | X    |      |
| Carbazole degradation                                                       | X    |      |
| Chlorinated phenols degradation                                              | X    |      |
| Cinnamate and 3-hydroxycinnamate degradation to 2-oxopent-4-enoate           | X    |      |
| Diphenyl ethers degradation                                                 | X    |      |
| Gentisate degradation I                                                      | X    |      |
| Phenol degradation I (aerobic)                                               | X    |      |
| Protocatechuic acid degradation                                              | X    |      |
| Protocatechuate degradation II (ortho-cleavage pathway)                     | X    |      |
Table 1 (continued)

| Pathway class                                                                 |   |
|-------------------------------------------------------------------------------|---|
| Uurate conversion to allantoin I                                              | X |
| Degradation/utilization/assimilation—carbohydrate degradation                 |   |
| Alginate degradation                                                          | X |
| D-galactose degradation I (Leloir pathway)                                    | X |
| D-galactose degradation V (Leloir pathway)                                    | X |
| Fructose degradation                                                          | X |
| Kojibiose degradation                                                         | X |
| L-arabinose degradation I                                                      | X |
| L-rhamnose degradation I                                                       | X |
| Melibiose degradation                                                         | X |
| Xylose degradation I                                                           | X |
| Degradation/utilization/assimilation—carboxylate degradation                  |   |
| β-D-glucuronide and d-glucuronate degradation                                  | X |
| 2-Amino-3-carboxymuconate semi-aldehyde degradation to 2-oxopentenoate        | X |
| 2-Methylcitrate cycle I                                                       | X |
| Degradation/utilization/assimilation—chlorinated compound degradation         |   |
| 2-Chlorobenzoate degradation                                                  | X |
| 4,5-Dichlorocatechol degradation                                              | X |
| Chlorinated phenols degradation                                               | X |
| Pentachlorophenol degradation                                                  | X |
| Degradation/utilization/assimilation—degradation/utilization/assimilation—other |   |
| Acrylonitrile degradation I                                                   | X |
| Octane oxidation                                                             | X |
| Degradation/utilization/assimilation—fatty acid and lipid degradation         |   |
| Cholesterol degradation to androstenedione I (cholesterol oxidase)           | X |
| Phospholipases                                                               | X |
| Triacylglycerol degradation                                                   | X |
| Degradation/utilization/assimilation—inorganic nutrient metabolism           |   |
| Aminomethylphosphonate degradation                                            | X |
| Cyanate degradation                                                           | X |
| Dimethyl sulfide degradation II (oxidation)                                   | X |
| Gly phosphate degradation II                                                   | X |
| Nitrate reduction V (assimilatory)                                            | X |
| Nitrogen fixation I (ferredoxin)                                              | X |
| Thiosulfate disproportionation IV (rhodanese)                                 | X |
| Degradation/utilization/assimilation—nucleoside and nucleotide degradation   |   |
| Guanosine nucleotides degradation III                                         | X |
| Pyrimidine deoxyribonucleosides degradation                                   | X |
| UTP and CTP dephosphorylation I                                              | X |
| Degradation/utilization/assimilation—polymeric compound degradation          |   |
| Alginate degradation                                                          | X |
| Degradation/utilization/assimilation—secondary metabolite degradation        |   |
| β-D-glucuronide and d-glucuronate degradation                                  | X |
| N-acetylglucosamine degradation I                                             | X |
| D-sorbitol degradation I                                                       | X |
| Limonene degradation I (α-limonene)                                          | X |
| Limonene degradation II (l-limonene)                                         | X |
| Generation of precursor metabolite and energy                                |   |
| Aerobic respiration I (cytochrome c)                                         | X |
| Glycerol-3-phosphate to cytochrome bo oxidase electron transfer              | X |
| Glycerol-3-phosphate to fumarate electron transfer                           | X |
(the single Casuarinaceae family that cannot be infected by C-I but by C-III). Since symbiosis also depends on the signaling machinery between the host plant and *Frankia*, we compared the signal peptide proteins present in C-I and C-III groups (Fig. 2). A total of 59 different types of signal proteins were found in the C-III group that was absent in C-I, which contains 115 types of signal peptide proteins that are absent in C-II or C-III. This demonstrated the fact that the signaling mechanisms adopted by these different groups of *Frankia* are quite distinct. Similarly, we have compared the signal peptides of C-II and C-III groups of *Frankia* and got a set of 80 unique signalP proteins for the C-II group. Those were absent in the C-III and C-I groups indicating a discrepancy in the signaling cascade between these two groups. We have not considered the atypical C-IV *Frankia*, since they always remain ineffective irrespective of their host plant.

Root nodule formation does not only depend upon the relative abundance of *Frankia* in rhizospheric soil but also commodiously hinges on the acceptability of the host plant toward a specific type of *Frankia* (Tekaya et al. 2018). Since our knowledge regarding the Gymnostoma as well as Ceanothus genomes and their metabolism along with the signaling system is very elusive, further contemplation especially investigation on root extract of such plants is required to reveal this aforementioned unusual behavior of *Frankia* nodulation.

### Intra-cluster mutualism among *Frankia*

The complementation index among different clusters of *Frankia* revealed higher intra-cluster mutualism than inter-cluster mutualism. The value ranged from 0.58 to 0.96 (for C-I), 0.70 to 0.89 (for C-II), 0.67 to 0.99 (for C-III), 0.62 to 0.96 (C-IV). Although Ben et al. (2018) proposed that the diversity of *Frankia* in their host plant root nodules is independent of the *Frankia* diversity and abundance, but is dependent on the choice of the host plant, the intra-cluster complementation may support the host-specific nature of
Frankia in the majority of cases. However, further future studies will strengthen this proposal.

Conclusions

Previous studies have proposed in places where nitrogen is still unavailable or present in very low concentration (in case of landslide and soil erosion), a mutualistic association between actinorhizal plants and Frankia remains unchanged as a result of selective advantage (Normand and Bouquet 1989). Frankia being nitrogen-fixing actinobacteria is vital from both ecological and agricultural aspects. However, the ecological aspect of Frankia was not much explored. In this study, we have specifically focused on the ecological principles acting on this genus which provide them the ability of host specificity or host differentiation. We have used the reverse ecology analysis among 44 Frankia strains distributed in 4 clusters. The analysis revealed the following:

(a) The C-I Frankia is a general competitor against other clusters and validated their host specificity toward the Hamamelidae subclass of actinorhizal plants.
(b) Another very exciting finding was the presence of niche partitioning canon between C-II and C-IV Frankia strains that allow them to share the same nodule via utilizing different nutrition sources, especially for carbohydrate and aromatic amino acids.

We hope with advanced NGS techniques that we will be able to get all the necessary information for the host plants and that will reveal a piece of more complex and interesting information regarding the co-existence of Frankia and host plants. However, we should keep in mind that the mere presence of a gene does not mean that they are expressed properly. Only when a gene is expressed, we can properly say that the genes and their corresponding proteins are being used in related biological functionality. Thus, an RNA-seq or transcriptomic analysis of these Frankia strains under a specific condition (i.e., co-inhabitation of C-II and C-IV, etc.) is further required for knowing the exact mechanism of Frankia interaction.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00203-022-02910-0.

Acknowledgements We acknowledge S Das and C Sarkar for initially standardizing the software. We also acknowledge the Biswa Bangla Genome Center and the University of North Bengal for supporting this work.

Author contributions AS and MG: conceived and developed the idea. IS, GS, and AS: designed the experiment and did the bioinformatics work. SB, IS, and AS: contributed to analyzing the data and wrote the final draft of the manuscript. All authors have read and approved the manuscript.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors declare that the research paper was written in the absence of any commercial or financial relationships that could be construed as a real or potential conflict of interest. Thus authors declare no conflict of interest.

Animal issue No animal or human was treated as the sample in this work.

References

Allaire J (2012) RStudio: integrated development environment for R. Boston, MA, vol 770, p 394
Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169
Barabasi AL (2009) Scale-free networks: a decade and beyond. Science 325:412–413
Barabasi A, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512
Ben Tekaya S, Guerra T, Rodriguez D, Dawson JO, Hahn D (2018) Frankia diversity in host plant root nodules is independent of abundance or relative diversity of Frankia populations in corresponding rhizosphere soils. Appl Environ Microbiol 84(5):e02248-e2317
Benson DR, Dawson JO (2007) Recent advances in the biogeography and genieology of symbiotic Frankia and its host plants. Physiol Plant 130(3):318–330
Benson DR, Vandenh H, Potter D (2004) Actinorhizal symbioses: diversity and biogeography. In: Gillings M (ed) Plant microbiology. BIOS Scientific Publishers Ltd, Oxford, pp 97–127
Berlow EL, Dunne JA, Martinez ND, Stark PB, Williams RJ, Brose U (2009) Simple prediction of interaction strengths in complex food webs. Proc Natl Acad Sci 106(1):187–191
Berry AM, Mendoza-Herrera A, Guo YY, Hayashi J, Persson T, Barabote R, Dem-chenko K, Zhang S, Pawlowski K (2011) New perspectives on nodule nitrogen assimilation in actinorhizal symbioses. Funct Plant Biol 38(9):645–652
Berry D, Widder S (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol 5:219
Borenstein E, Kupiec M, Feldman MW (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc Natl Acad Sci 105:14482–14487
Bousquet J, Girouard E, Strobeck C, Dancik BP, Lalonde M (1989) Restriction fragment polymorphisms in the rDNA region among seven species of Alnus and Betulapapyrifera. Plant Soil 118(1–2):231–240
Cao Y, Wang Y, Zheng X, Li F, Bo X (2016) RevEcoR: an R package for the reverse ecology analysis of microbiomes. BMC Bioinform 17(1):1–6
Carr R, Borenstein E (2012) NetSeed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. Bioinform Oxf Engl 28:734–735
Caspi R, Foerster H, Fulcher CA, Kaipa P, Krumsenacker M, Latendresse M, Paley S, Rhee SY, Sherrar AG, Tissier C, Walk TC (2007) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucl Acids Res 36:D623–D631

Cisoko M, Hocher V, Gherbi H, Gully D, Carré-Mlouka A, Sane S, Fournier P (2018) Actinorhizal signaling molecules: Frankia root hair deforming factor shares properties with NIN inducing factor. Front Plant Sci 9:1494

Cronquist A, California Botanical Society, Berkeley (1968) The evolution and classification of flowering plants. Madroño; West Am J Bot. 20:77–78

Duffy JE, Cardinale BJ, France KE, McIntyre PB, Thébault E, Loreau M (2007) The functional role of biodiversity in ecosystems: incorporating trophic complexity. Ecol Lett 10(6):522–538

Dunne JA (2006) The network structure of food webs. Ecological networks: linking structure to dynamics in food webs. Oxford University Press, Oxford, pp 27–86

Faust K, Raes J (2012) Microbial interactions: from networks to models. Nat Rev Microbiol 10:538–550

Friedman J, Alm EJ (2012) Inferring correlation networks from genomic survey data. PLoS Comp Biol 8:e1002687

Gilbert N, Russo G, Marris E, Lane N, Sukhdev P, Turner WR, Wilcove DS (2009) The entangled bank unravels. Nature 462:251–252

Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180(18):4765–4774

Ings TC, Montoya JM, Bascombe J, Blüthgen N, Brown L, Dormann P, Barden R, Moiroud A (1986) More than a meal integrating non-feeding interactions into food webs. Ecol Lett 15(4):291–300

Lassalle F, Muller D, Nesme X (2015) Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. Microbiol Res 166(10):729–741

Lefort V, Desper R, Gascuel O (2009) FastME 2.0: a comprehensive, accurate and fast distance-based phylogeny inference program. Mol Biol Evol 32(10):2798–2800

Levy R, Borenstein E (2012) Reverse ecology: from systems to environments and back. Adv Exp Med Biol 751:329–345

Levy R et al (2015) Netcooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation. BMC Bioinf 16:164

Mansour S, Swanson E, McNutt Z, Pesce C, Harrington K, Abebe-Alele F, Simpson S, Morris K, Thomas WK, Tisa LS (2017) Permanent draft genome sequence for Frankia sp strain CeD, a nitrogen-fixing actinobacterium isolated from the root nodules of Casuarina equisetifolia grown in Senegal. Genome 4(2):00265–00316

Normand P, Bouquet J (1989) Phylogeny of nitrogenase sequences in Frankia and other nitrogen-fixing microorganisms. J Mol Evol 29(5):436–447

Normand P, Lapiere P, Tisa LS, Gogarten AJP, Bagnarello E, Benson DR (2007) Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. Genome Res 17(1):7–15

Nouioua I, Gtari M, Góker M, Ghodhbane-Gtari F, Tisa LS, Fernandez MP, Normand P, Huntemann M, Clum A, Pillay M, Varghese N (2016) Draft genome sequence of Frankia strain G2, a nitrogen-fixing actinobacterium isolated from Casuarina equisetifolia and able to nodulate actinorhizal plants of the order Rhamnales. Genome Announc 4(3):e00437-e516

Poisot T, Stouffer DB, Gravel D (2015) Beyond species: why ecological interaction networks vary through space and time. Oikos 124(3):243–251

Pujic P, Bolotin A, Fournier P, Sorokin A, Lapido A, Richau KH, Briolay J, Mebariki F, Normand P, Sellstedt A (2015) Genome sequence of the atypical symbiotic Frankia R43 strain, a nitrogen-fixing and hydrogen-producing actinobacterium. Genome Announc 3(6):e01387-e1415

Schwencke J, Carú M (2001) Advances in actinorhizal symbiosis: host-plant-Frankia interactions, biology, and applications in arid land reclamation. Arid Land Res Manag 15(4):285–327

Schwintzer CR (2012) The biology of Frankia and actinorhizal plants. Academic Press Inc, Cambridge

Sellstedt A, Richau KH (2013) Aspects of nitrogen-fixing actinobacteria, in particular free-living and symbiotic Frankia. FEMS Microbiol Lett 342(2):179–186

Sen A, Daubin V, Abrouk D, Gifford I, Berry AM, Normand P (2014) Phylogeny of the class ACTINOBACTERIA revisited in the light of complete genomes. The orders ‘Frankia’ les and Micrococcaceae should be split into coherent entities: proposal of Frankiales ord. nov., Geodermatophilales ord. nov., Acidothermales ord. nov. and Nakamuraellales ord. nov. Int J Syst Evol Microbiol 64(11):3821–3832

Simonet P, Haurat J, Normand P, Bardin R, Moiroud A (1986) Localization of nif genes on a large plasmid in Frankia sp. strain UQ10321305.009. Mol Gen Genet 204:492–495

Suzuki K, Yoshida K, Nakashima Y, Fukuda S (2017) An equation-free method reveals the ecological interaction networks within complex microbial ecosystems. Methods Ecol Evol 8(12):1774–1785

Tekaya SB, Guerra T, Pacheco cable, Tisa LS, Hahn D (2018) Frankia diversity in host plant root nodules is independent of abundance or relative diversity of Frankia populations in corresponding rhizosphere soils. Appl Environ Microbiol 84(5):e02248–17

Velmurpally S, Guerra T, Hahn D (2019) Localization of typical and atypical Frankia isolates from Casuarina sp. in nodules formed on Casuarina equisetifolia. Plant Soil 435(1):385–393

Zhang W (2011) Constructing ecological interaction networks by correlation analysis: hints from community sampling. New Biol 1(2):81

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