9-12-2017

Permanent Draft Genome sequence for Frankia sp. strain CcI49, a Nitrogen-Fixing Bacterium Isolated from Casuarina cunninghamiana that Infects Elaeagnaceae

Samira R. Mansour  
*Suez Canal University*

Erik Swanson  
*University of New Hampshire, Durham*

Zakkary McNutt  
*University of New Hampshire, Durham*

Celine Pesce  
*University of New Hampshire, Durham*, Celine.Pesce@unh.edu

Kelsey Harrington  
*University of New Hampshire, Durham*

See next page for additional authors

Follow this and additional works at: https://scholars.unh.edu/mcbs_facpub

**Recommended Citation**


This Article is brought to you for free and open access by the Molecular, Cellular and Biomedical Sciences at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Molecular, Cellular and Biomedical Sciences Scholarship by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.
SOIL DWELLING ACTINOBACTERIA OF THE GENUS FRANKIA FORM AN ENDOPHYTIC SYMBIOSIS WITH ACTINORHIZAL PLANTS, WHICH ARE COMPRISED OF OVER 200 SPECIES FROM 8 ANGIOSPERM FAMILIES [1, 2]. ACTINORHIZAL PLANTS IN SYMBIOSIS WITH FRANKIA PLAY IMPORTANT ECOCLOGICAL ROLES AS PIONEER SPECIES AND ARE USED IN AGROFORESTRY, LAND RECLAMATION, CROP PROTECTION, AND SOIL STABILIZATION PROJECTS [3]. MOLECULAR PHYLOGENETIC APPROACHES HAVE IDENTIFIED FOUR MAJOR CLUSTERS OF FRANKIA THAT ALSO FOLLOW HOST PLANT SPECIFICITY GROUPS [4-7]. MEMBERS OF CLUSTER 1 ARE DIVIDED INTO SUB-CLUSTER 1A THAT ARE INFECTIVE ON ALNUS AND MYRTICACEAE AND SUB-CLUSTER 1B STRAINS WHICH ARE INFECTIVE ON ALLOCASUARINA, CASUARINA AND MYRTICACEAE. CLUSTER 2 REPRESENTS STRAINS INFECTIVE ON CORIARIACEAE, DATISCACEAE, DRYADOIDEAE AND Ceanothus, WHILE CLUSTER 3 COMPRISES STRAINS THAT ARE INFECTIVE ON COLLETIEAE, ELAEAGNACEAE, GYMNSTOMA AND MYRTICACEAE. FINALLY, CLUSTER 4 GROUPS FRANKIA STRAINS ISOLATED FROM ACTINORHIZAL NODULES THAT ARE UNABLE TO UNDERMATE THE NITROGEN-FIXATION PROCESS (Fix-) AND/OR RE-INFECT THEIR HOST PLANT CAUSING NODULATION (Nod-) AND ARE CLASSIFIED AS “ATYPICAL FRANKIA”. GENOMES FOR REPRESENTATIVES FROM EACH CLUSTER HAVE BEEN QUENTENCED [8]. THE AVAILABILITY OF THESE FRANKIA GENOME DATABASES HAS OPENED UP THE USE OF “OMICS” APPROACHES. ANALYSIS OF FRANKIA GENOMES HAS REVEALED NEW POTENTIAL IN RESPECT TO METABOLIC DIVERSITY, NATURAL PRODUCT BIOSYNTHESIS, AND STRESS TOLERANCE, WHICH MAY AID THE COSMOPOLITAN NATURE OF THE ACTINORHIZAL SYMBIOSIS.

Several Frankia strains isolated from Casuarina nodules are unable to re-infect it, but are able to infect other actinorhizal plant genera like Elaeagnus [9, 10]. Although isolated from Casuarina nodules, these Frankia strains are classified as members of cluster 3 based on molecular phylogeny and genomes for two members of this group have been sequenced [11, 12]. Frankia sp. strain CcI49 was isolated from root nodules

**Abstract**

Frankia sp. strain CcI49 was isolated from Casuaria cunninghamiana nodules. However the strain was unable to re-infect Casuaria, but was able to infect other actinorhizal plants including Elaeagnaceae. Here, we report the 9.8-Mbp draft genome sequence of Frankia sp. strain CcI49 with a G+C content of 70.5 % and 7,441 candidate protein-encoding genes. Analysis of the genome revealed the presence of a bph operon involved in the degradation of biphenyls and polychlorinated biphenyls.

Key words: Actinorhizal symbiosis, bioremediation, nitrogen fixation, natural products, host microbe interactions, genomes.
of Casuarina cunninghamiana grown on the edge of a cultivated field on side of the highway in Ismailia-Port Said, Egypt. The fresh nodules were washed, dissected into individual lobes, and surface-sterilized as described previously [13]. Each lobe was checked for sterility in sterile nutrient-rich medium. Nodules that were free from contamination were selected, dissected and homogenized, the homogenates were transferred to 100-ml screw caped bottle containing modified BAP medium for outgrowth. Hyphal outgrowth was homogenized and plated onto solid medium. After 3-4 weeks, colonies picked from the plates were homogenized and incubated in liquid medium. Surprisingly, Frankia sp. strain CcI49 produced reddish colonies, while other Frankia isolates from Casuarina do not. Frankia sp. strain CcI49 produced sporangia and spores that were smaller and narrower than normal Frankia sporangia and spores (Figure 1). Spores from Frankia sp. strain CcI49 had a high germination rate similar to Frankia strain CeI5 [14, 15]. We tested the ability of Frankia sp. strain CcI49 to re-infect actinorhizal plants. Four different actinorhizal plant species were tested to assay the plant host range and ten plants of each species were inoculated. Frankia sp. strain CcI49 was unable to infect C. cunninghamiana and Alnus glutinosa, but formed nodules on Elaeagnus angustifolia and Hippophae rhamnoides. All ten of the E. angustifolia and H. rhamnoides plants tested formed nodules. Thus, Frankia sp. strain CcI49 had a plant-host-specificity pattern similar to Frankia sp. strains G2 and R43 [9, 10] from cluster 3 also isolated from Casuarina root nodules. Frankia sp. strain CcI49 genome was chosen to be sequenced for several reasons including an interesting physiology including the production of a reddish pigment, development of smaller sporangia and spores than are typical found with Frankia, and providing more information on this Frankia subcluster.

Sequencing of the draft genome of Frankia sp. strain CcI49 was performed at the Hubbard Center for Genome Studies (University of New Hampshire, Durham, NH) using Illumina technology techniques [16]. A standard Illumina shotgun library was constructed and sequenced using the Illumina HiSeq2500 platform, which generated 7,939,466 reads (260-bp insert size) totaling 1,921 MBp. The Illumina sequence data were trimmed by Trimmonatic version 0.32 [17], assembled using Spades version 3.5 [18], and ALLPaths-LG version r52488 [19]. The final draft assembly for Frankia sp. strain CcI49 consisted of 78 contigs with an N50 contig size of 282.1 kb and 167X coverage of the genome. The final assembled genome contained a total sequence length of 9,758,130 bp with a G+C content of 70.5%.

Figure 1. Photomicrograph of Frankia sp. strain CcI49 grown in liquid culture. The elongated arrow shows the presence of a long, narrow sessile sporangium containing differentiated mature spores at distal end (short arrow). Size bar represents 32 µm.

The assembled Frankia sp. strain CcI49 genome was annotated via the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), and resulted in 7,411 candidate protein-encoding genes, 46 tRNA and 2 rRNA regions. The genome features of Frankia sp. strain CcI49 fall outside the realm of the other cluster 1b genomes, but similar to other cluster 3 isolates from Casuarina (Table 1). Phylogenetic analysis of the 23S rDNA shows that Frankia sp. strain CcI49 groups with the cluster 3 strains (Figure S1). The genome size and corresponding number of CDSs were larger than the typical cluster 1b, but fit within those values reported for cluster 3 genomes [8]. The genome also contained a nif, 2 hup, and 1 shc operons encoding the nitrogenase and uptake hydrogenase enzymes and the hopanoid biosynthetic pathway, respectively. The operons were organized similar to those reported for Frankia cluster 3 genomes [8].
Table 1. Genome features of \textit{Frankia} sp. strain CcI49 and other \textit{Frankia} strains isolated from \textit{Casuarina} root nodules.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Source</th>
<th>Location$^1$</th>
<th>Size (Mb)</th>
<th>No. of Contigs</th>
<th>\textit{Frankia} cluster</th>
<th>No. of CDS</th>
<th>Host Plants$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcI49</td>
<td>This study</td>
<td>Egypt</td>
<td>9.76</td>
<td>78</td>
<td>3</td>
<td>7,441</td>
<td>Elaeagnaceae</td>
</tr>
<tr>
<td>R43</td>
<td>[12]</td>
<td>USA</td>
<td>10.45</td>
<td>46</td>
<td>3</td>
<td>7,644</td>
<td>Elaeagnaceae</td>
</tr>
<tr>
<td>KB5</td>
<td>[23]</td>
<td>Australia</td>
<td>5.46</td>
<td>420</td>
<td>1b</td>
<td>4,958</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>CcI3</td>
<td>[24]</td>
<td>USA</td>
<td>5.43</td>
<td>1</td>
<td>1b</td>
<td>4,598</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>CeD</td>
<td>[25]</td>
<td>Senegal</td>
<td>5.00</td>
<td>120</td>
<td>1b</td>
<td>4,403</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>Allo2</td>
<td>[26]</td>
<td>USA</td>
<td>5.33</td>
<td>110</td>
<td>1b</td>
<td>4,838</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>Thr</td>
<td>[27]</td>
<td>Egypt</td>
<td>5.31</td>
<td>171</td>
<td>1b</td>
<td>4,805</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>BMG5.23</td>
<td>[28]</td>
<td>Tunisia</td>
<td>5.27</td>
<td>167</td>
<td>1b</td>
<td>4,747</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>CcI6</td>
<td>[29]</td>
<td>Egypt</td>
<td>5.39</td>
<td>138</td>
<td>1b</td>
<td>4,902</td>
<td>Casuarinaceae</td>
</tr>
<tr>
<td>BR</td>
<td>[30]</td>
<td>Brazil</td>
<td>5.23</td>
<td>180</td>
<td>1b</td>
<td>4,777</td>
<td>Casuarinaceae</td>
</tr>
</tbody>
</table>

$^1$The source of the isolate
$^2$Re-infection plant host range

Analysis of the \textit{Frankia} sp. strain CcI49 revealed the presence of the \textit{bph} operon coding for a potential metabolic pathway involved in the degradation of biphenyl and polychlorinated biphenyls (Figure 2). The \textit{bph} operon is also present in \textit{Frankia} sp. strains EuI1c and EUN1f genomes [20] and was also found in the genomes of \textit{Frankia} sp. strains G2 and R43 [11, 12], cluster 3 strains isolated from \textit{Casuarina} root nodules. Both \textit{Frankia} sp. strains CcI49 and EUN1f contained the entire \textit{bph} operon, while two genes (\textit{bphA3} and \textit{bphH}) are missing in \textit{Frankia} sp. EuI1c (Figure 2). The presence of the complete \textit{bph} operon suggests that \textit{Frankia} sp. strain CcI49 may be capable of degrading these recalcitrant xenobiotics.

Bioinformatic analysis of this genomes by the use of the AntiSMASH program [21] revealed the presence of high numbers of secondary metabolic biosynthetic gene clusters, which is consistent with previous results with other \textit{Frankia} genomes including cluster 3 [8, 22]. Table 2 shows a comparison of the various profiles of different \textit{Frankia} strains isolated from \textit{Casuarina} for these secondary metabolic biosynthetic gene clusters. The profile of \textit{Frankia} sp. strain CcI49 differed from those shown by \textit{Frankia} strains that are able re-infect \textit{Casuarina} and was similar to the pattern exhibited by the other two cluster 3 strains (R43 and G2) isolated from \textit{Casuarina} nodules. These cluster 3 genomes contained more polyketide synthase (PKS) biosynthetic clusters than the cluster 1b genomes. The \textit{Frankia} sp. strain CcI49 genome contained several unique clusters that had homologues in other bacteria or were completely novel.

In summary, the \textit{Frankia} sp. strain CcI49 genome has revealed an interesting potential metabolic pathways and natural product profile, and serves as a representative of \textit{Frankia} cluster 3. Further analysis of this genome and experimental evidence will be needed to support the predicted natural product profile and metabolic potential of \textit{Frankia} sp. strain CcI49.

![Figure 2. The bph operon is present in Frankia strains CcI49, EuI1c, G2, R43, and EUN1f. bph genes encode the following: BphA, A2 and A3, Biphenyl 2,3-dioxygenase; BphB, cis-2,3-dihydro-2,3-dihydroxybiphenyl dehydrogenase; BphC, 2,3-dihydroxybiphenyl 1,2-dioxygenase; BphD, 2-hydroxy-6-phenyl-6-oxohexa-2,4-dieneoate (HOPDA) hydrolase; BphE, 2-hydroxypenta-2,4-dienoate hydratase; BphF, acylating acetaldehyde dehydrogenase; and BphG, 4-hydroxy-2-oxovalerate aldolase.](http://www.jgenomics.com)
Table 2. Biosynthetic gene clusters for natural products found in the genomes from *Casuarina Frankia* strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th><em>Frankia</em> Cluster</th>
<th>No. of Biosynthetic gene clusters 1</th>
<th>NRPS 2</th>
<th>PKS 3</th>
<th>Terpene</th>
<th>Siderophore</th>
<th>Bacteriocin</th>
<th>Lantipeptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cc49</td>
<td>3</td>
<td>42</td>
<td>6</td>
<td>17</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>R43</td>
<td>3</td>
<td>38</td>
<td>4</td>
<td>14</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>G2</td>
<td>3</td>
<td>35</td>
<td>8</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>K85</td>
<td>lb</td>
<td>34</td>
<td>4</td>
<td>9</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Cc93</td>
<td>lb</td>
<td>29</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>CoD</td>
<td>lb</td>
<td>30</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>All02</td>
<td>lb</td>
<td>32</td>
<td>7</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Thr</td>
<td>lb</td>
<td>33</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>BMGS23</td>
<td>lb</td>
<td>31</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Cc66</td>
<td>lb</td>
<td>33</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>BR</td>
<td>lb</td>
<td>29</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

1 Biosynthetic gene clusters were identified by the use of the AntiSMASH software [21]
2 NRPS: Nonribosomal peptide synthase
3 PKS: polyketide synthase including Type I, II, III, Trans-AT, and other types

**Nucleotide sequence accession numbers**

This whole-genome shotgun sequence has been deposited at DDBJ/EMBL/GenBank under the accession number MOWP00000000.1. The version described in this paper is the first version, MOWP01000000.

**Supplementary Material**

Figure S1. http://www.jgenomics.com/v05p0119s1.pdf

**Acknowledgement**

Partial funding was provided by the New Hampshire Agricultural Experiment Station. This is Scientific Contribution Number 2471. This work was supported by the USDA National Institute of Food and Agriculture Hatch 022821 (LST), Agriculture and Food Research Initiative Grant 2015-67014-22849 from the USDA National Institute of Food and Agriculture (LST), the College of Life Science and Agriculture at the University of New Hampshire-Durham and the Science and Technology Development Fund (STDF), grant No. 900, as International Cooperation Project (SM and LST). Sequencing was performed on an Illumina HiSeq2500 purchased with an NSF MRI Grant: DBI-1229361 to WK Thomas.

**Competing Interests**

The authors have declared that no competing interest exists.

**References**


