On the Launch of the JCM-WDCM Joint Genome Sequencing Project

Takashi Itoh and Moriya Ohkuma

Japan Collection of Microorganisms, RIKEN BioResource Research Center
Outline of RIKEN BRC-JCM

理研バイオリソースセンター 微生物材料開発室
Japan Collection of Microorganisms, RIKEN BioResource Research Center

History

1981 Establishment in Wako campus
2004 Integrated with RIKEN BRC as Microbe Division
2007 Appointed as a core member of National Bioresource Project
2012 Relocation to Tsukuba campus

Mission

Collection, Preservation, Quality Control, Distribution of diverse species of Bacteria, Archaea, Fungi useful for environmental and health sciences to contribute to scientific community

Human Resources

Staff Members

Head
Curators
Gene Analysis
Resource Advancement
Agency-staff and Part-timers
Postdoctoral Fellows
Visiting Scientists
Students

Dr. Ohkuma
7
1
3
17
4
2
2
Microbial Resources in JCM

1) Strains relevant to researches of human health and environmental conservation

2) Type strains and related strains of bacteria, archaea, filamentous fungi and yeasts (confine to BSL 1 and 2 microorganisms)

3) Model microorganisms for academic researches (e.g., genome-sequenced microorganisms)

- Indigenous bacteria in human intestines, skins and oral cavities
- Symbiotic microbes associated with plants and insects/termites
- Oleaginous yeasts utilizing starch
- Bacteria that cause metal corrosion
Microbial Resources in JCM

1) Strains relevant to researches of human health and environmental conservation

2) Type strains and related strains of bacteria, archaea, filamentous fungi and yeasts (confine to BSL 1 and 2 microorganisms)

3) Model microorganisms for academic researches (e.g., genome-sequenced microorganisms)

Strain Holding

- Yeasts: 893 (3,681) strains
- Filamentous Fungi: 460 (4,031) strains
- Archaea: 404 (817) strains
- Bacteria: 7,585 (18,989) strains
  - Proteobacteria: 40%
  - Actinobacteria: 29%
  - Firmicutes: 23%
  - Bacteroides: 7%
  - Others (15 phyla): 1%
- Others: micro-algae: 26 (5) strains

Total: 27,544 strains (As of November, 2018)

type strains or derived from type materials
Microbial Resources in JCM

1) Strains relevant to researches of human health and environmental conservation

2) Type strains and related strains of bacteria, archaea, filamentous fungi and yeasts (confine to BSL 1 and 2 microorganisms)

3) Model microorganisms for academic researches (e.g., genome-sequenced microorganisms)

---

**Pyrococcus furiosus** Fiala and Stetter 1986

| Taxonomy in NCBI database: Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae. |

**84227** — DSM 3638 — K. O. Stetter and G. Fiala Vc 1.


Type strain [371.3007].

Medium: 1511. Temperature: 95°C. Anaerobic.

Source: Hot sediment at the beach of Porto di Levante, Vulcano Island, Italy [3007].

G+C (mol%) 38.3 (Tm 37.1 (HPLC) [3007].

DNA-DNA relatedness: [3007.219].

Phylogeny: 16S rRNA gene (A9603518).

Other taxonomic data: Polyamine [3632].

Genome sequence: A0099902 [3519].

Genomic DNA is available from RIKEN BRC-DNA Bank: JGD 07704.

Publication(s) using this strain [A04195, A08175, A11327, A13147, A13513, A13523, A13527, A14107, A15176].

Patent publication(s) using this strain [HP2009-06921A].

Delivery category: Domestic, B; Overseas, B.

---

**Genome sequence (accession No.)**

As of November, 2018, genomic sequences of more than 2,100 strains are accessible.

**Genomic DNA**

As of October, 2018, genomic DNA samples of more than 760 strains are available from the RIKEN DNA BANK.
NBRP (National BioResource Program) Genome Information Upgrading Program. Under the program, JCM has determined genome sequences of various microbial strains useful for researchers, in collaboration with the other genome-sequencing centers/institutes.
FY 2014 Eukaryotic strains, 217 strains, Data publication in 2016 (http://www.jcm.riken.jp/cgi-bin/nbrp/nbrp_list.cgi)

The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species

Culture collection partners

ATCC, BCCM/LMG, BCRC, CAIM, CBS, CCUG, CECT, CICC, CIP, CGMCC, ICMP, JCM, KCTC, KMM, MUM, NCTC, NBRC, NCAIM, PCU, TBRC, TISTR, UCD, VKM
Timeline of the JCM-WDCM Joint Genome-Sequencing Project

Contract of the research collaboration
(JCM-WDCM Joint Genome Sequencing Project)  
December 2017

Selection of microbial strains to be genome-sequenced
FY2018 51 strains
FY2019- 50-100 strains/year  
May 2018

Conclusion of MTA (category II)* for 2018 resource delivery
*relevant to for-profit organization(s)  
September 2018

Preparation of DNA samples

Delivery of the dried DNA samples to BGI@HongKong
15 strains (1st batch): delivered in September 2018
31 strains (2nd batch): will be delivered by January 2019?
## List of First Batch Strains for the Genome-Sequencing

<table>
<thead>
<tr>
<th>JCM No.</th>
<th>Species name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. JCM 14781</td>
<td><em>Amphritea balenae</em></td>
</tr>
<tr>
<td>2. JCM 17466</td>
<td><em>Anaerostipes butyricicus</em></td>
</tr>
<tr>
<td>3. JCM 31056</td>
<td><em>Butyricoccus faecihominis</em></td>
</tr>
<tr>
<td>4. JCM 18676</td>
<td><em>Butyricimonas faecihominis</em></td>
</tr>
<tr>
<td>5. JCM 18677</td>
<td><em>Butyricimonas paravirosa</em></td>
</tr>
<tr>
<td>6. JCM 14719</td>
<td><em>Calditerricola satsumensis</em></td>
</tr>
<tr>
<td>7. JCM 12580</td>
<td><em>Lentibacillus kapialis</em></td>
</tr>
<tr>
<td>8. JCM 18682</td>
<td><em>Parabacteroides faecis</em></td>
</tr>
<tr>
<td>9. JCM 23203</td>
<td><em>Pelagicoccus mobilis</em></td>
</tr>
<tr>
<td>10. JCM 30531</td>
<td><em>Porphyromonas pasteri</em></td>
</tr>
<tr>
<td>11. JCM 31740</td>
<td><em>Sulfodiicoccus acidiphilus</em></td>
</tr>
<tr>
<td>12. JCM 13583</td>
<td><em>Thermogymnononas acidicola</em></td>
</tr>
<tr>
<td>13. JCM 19902</td>
<td><em>Thermus composti</em></td>
</tr>
<tr>
<td>14. JCM 30071</td>
<td><em>Virgibacillus kapii</em></td>
</tr>
<tr>
<td>15. JCM 11219</td>
<td><em>Vulcanisaeta souniana</em></td>
</tr>
</tbody>
</table>

including archaea, anaerobic bacteria, thermophilic bacteria, halophilic bacteria
Output of the Genome-Sequencing

Illumina HiSEQ Platform

Scaffold: 31 to 168
N50: 31K to 932K

<table>
<thead>
<tr>
<th>Sample Name (#)</th>
<th>Illumina Data (Mb)</th>
<th>Scaffold Number (#)</th>
<th>Genome Size (bp)</th>
<th>N50 (bp)</th>
<th>Gene Number (#)</th>
<th>ncRNA Number (#)</th>
<th>Repeat Number (#)</th>
<th>Annotation Number (#)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCM.11219</td>
<td>1.119(458X)</td>
<td>43</td>
<td>2,439,499</td>
<td>109,967</td>
<td>2,689</td>
<td>38</td>
<td>59</td>
<td>2,435 (90.55%)</td>
</tr>
<tr>
<td>JCM.12580</td>
<td>1.123(309X)</td>
<td>168</td>
<td>3,632,762</td>
<td>46,708</td>
<td>3,696</td>
<td>97</td>
<td>462</td>
<td>3,438 (93.01%)</td>
</tr>
<tr>
<td>JCM.13553</td>
<td>1.115(635X)</td>
<td>34</td>
<td>1,753,463</td>
<td>932,198</td>
<td>1,678</td>
<td>47</td>
<td>83</td>
<td>1,528 (91.06%)</td>
</tr>
<tr>
<td>JCM.14719</td>
<td>1.120(465X)</td>
<td>135</td>
<td>2,406,988</td>
<td>31,316</td>
<td>2,473</td>
<td>104</td>
<td>107</td>
<td>2,304 (93.16%)</td>
</tr>
<tr>
<td>JCM.17466</td>
<td>1.115(351X)</td>
<td>63</td>
<td>3,168,906</td>
<td>152,633</td>
<td>3,313</td>
<td>62</td>
<td>220</td>
<td>3,013 (90.94%)</td>
</tr>
<tr>
<td>JCM.18676</td>
<td>1.127(234X)</td>
<td>62</td>
<td>4,801,610</td>
<td>218,321</td>
<td>4,028</td>
<td>87</td>
<td>596</td>
<td>3,808 (94.53%)</td>
</tr>
<tr>
<td>JCM.18677</td>
<td>1.122(202X)</td>
<td>78</td>
<td>5,547,395</td>
<td>209,124</td>
<td>4,607</td>
<td>84</td>
<td>607</td>
<td>4,345 (94.31%)</td>
</tr>
<tr>
<td>JCM.18682</td>
<td>1.124(168X)</td>
<td>51</td>
<td>6,667,077</td>
<td>234,597</td>
<td>5,105</td>
<td>92</td>
<td>362</td>
<td>4,860 (95.2%)</td>
</tr>
<tr>
<td>JCM.19902</td>
<td>1.131(542X)</td>
<td>73</td>
<td>2,084,806</td>
<td>63,575</td>
<td>2,216</td>
<td>67</td>
<td>177</td>
<td>2,166 (97.74%)</td>
</tr>
<tr>
<td>JCM.23203</td>
<td>1.128(149X)</td>
<td>53</td>
<td>7,526,366</td>
<td>483,364</td>
<td>5,920</td>
<td>64</td>
<td>230</td>
<td>5,305 (89.61%)</td>
</tr>
<tr>
<td>JCM.30071</td>
<td>1.123(256X)</td>
<td>31</td>
<td>4,377,410</td>
<td>442,210</td>
<td>4,480</td>
<td>95</td>
<td>171</td>
<td>4,325 (96.54%)</td>
</tr>
<tr>
<td>JCM.30531</td>
<td>1.116(513X)</td>
<td>20</td>
<td>2,173,637</td>
<td>422,138</td>
<td>1,694</td>
<td>68</td>
<td>369</td>
<td>1,656 (97.75%)</td>
</tr>
<tr>
<td>JCM.31056</td>
<td>1.122(374X)</td>
<td>65</td>
<td>2,998,932</td>
<td>114,052</td>
<td>2,999</td>
<td>88</td>
<td>176</td>
<td>2,672 (89.09%)</td>
</tr>
<tr>
<td>JCM.31740</td>
<td>1.124(471X)</td>
<td>75</td>
<td>2,383,621</td>
<td>56,674</td>
<td>2,634</td>
<td>48</td>
<td>116</td>
<td>2,118 (80.41%)</td>
</tr>
</tbody>
</table>

Note: X represents the sequencing multiplier, the amount of sequencing reads divided by the size of the genome;
Publication process of the genome sequences

1. Registration in the DDBJ/ENA/GenBank
   - Further analyses (including re-sequencing, completion of the whole genome)
   - Review the output

Publication
- Public Databases
- WDCM GCM website
- JCM Online Catalogue
Acknowledgements

All of the partners involved in the GCM 2.0 projects