

# GigaScience

## The Global Catalogue of Microorganisms 10K type strain sequencing project: An international effort of the culture collection community toward closing the genomic gaps for the validly published prokaryotic and fungi species

--Manuscript Draft--

<b>Manuscript Number:</b>	GIGA-D-18-00079	
<b>Full Title:</b>	The Global Catalogue of Microorganisms 10K type strain sequencing project: An international effort of the culture collection community toward closing the genomic gaps for the validly published prokaryotic and fungi species	
<b>Article Type:</b>	Commentary	
<b>Funding Information:</b>	Strategic Priority Research Program of the Chinese Academy of Sciences (XDA19050301)	Dr. Juncai Ma
	Bureau of International Cooperation, Chinese Academy of Sciences (153211KYSB20160029)	Dr. Juncai Ma
	Bureau of International Cooperation, Chinese Academy of Sciences (153211KYSB20150010)	Dr. Juncai Ma
	National key Research Program of China (2017YFC1201202)	Dr. Juncai Ma
	National key Research Program of China (2016YFC1201303)	Dr. Juncai Ma
	National key Research Program of China (2016YFC0901702)	Dr. Juncai Ma
	13th Five-year Informatization Plan of Chinese Academy of Sciences (XXH13506)	Dr. Juncai Ma
	National Science Foundation for Young Scientists of China (31701157)	Dr. Linhuan Wu
<b>Abstract:</b>	<p>Genomic information is essential to underpin taxonomic, phylogenetic and functional studies to comprehensively decipher the characteristics of microorganisms, to explore microbiomes through metagenomics and also answer fundamental questions of nature and human life. However, there is still a big gap of genomic sequencing information for the bacterial and archaeal species with validly published names. For fungal type strains, the remaining gap is even larger. Global Catalogue of Microorganisms (GCM)'s led type strain sequencing project is an international coordinated effort to close current gaps in the genomic maps of microbes and hence to promote research through deep-mining of the genomic data.</p>	
<b>Corresponding Author:</b>	Juncai Ma  CHINA	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>		
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Linhuan Wu	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Linhuan Wu	
	Kevin McCluskey	
	Philippe Desmeth	

	Shuangjiang Liu
	Sugawara Hideaki
	Ye Yin
	Ohkuma Moriya
	Takashi Itoh
	Cha Young Kim
	Jung-Sook Lee
	Yuguang Zhou
	Hiroko Kawasaki
	Manzour Hazbón
	Vincent Robert
	Teun Boekhout
	Nelson Lima
	Lyudmila Evtushenko
	Kyria Boundy-Mills
	Boyke Bunk
	Edward R.B. Moore
	Lily Eurwilaichitr
	Supawadee Ingsriswang
	Heena Shah
	Su Yao
	Tao Jin
	Jinqun Huang
	Wenyu Shi
	Qinglan Sun
	Guomei Fan
	Wei Li
	Xian Li
	Ipek Kurtböke
	Juncai Ma
<b>Order of Authors Secondary Information:</b>	
<b>Opposed Reviewers:</b>	
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
Are you submitting this manuscript to a special series or article collection?	No
<b>Experimental design and statistics</b>	Yes
Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our	

<p><a href="#">Minimum Standards Reporting Checklist.</a> Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	
<p><b>Resources</b></p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	<p>Yes</p>
<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	<p>Yes</p>

1 **The Global Catalogue of Microorganisms 10K type strain sequencing project: An**  
2  
3 **international effort of the culture collection community toward closing the**  
4  
5 **genomic gaps for the validly published prokaryotic and fungi species**  
6  
7  
8  
9

10  
11 Linhuan Wu<sup>1,2,3</sup>, Kevin McCluskey<sup>4,23</sup>, Philippe Desmeth<sup>5,23</sup>, Shuangjiang Liu<sup>2,23</sup>,  
12  
13 Sugawara Hideaki<sup>6</sup>, Ye Yin<sup>7</sup>, Ohkuma Moriya<sup>8</sup>, Takashi Itoh<sup>8</sup>, Cha Young Kim<sup>9</sup>, Jung-  
14  
15 Sook Lee<sup>9</sup>, Yuguang Zhou<sup>10</sup>, Hiroko Kawasaki<sup>11</sup>, Manzour Hernando Hazbón<sup>12</sup>,  
16  
17 Vincent Robert<sup>13</sup>, Teun Boekhout<sup>13,24,25</sup>, Nelson Lima<sup>14</sup>, Lyudmila Evtushenko<sup>15</sup>, Kyria  
18  
19 Boundy-Mills<sup>16,23</sup>, Boyke Bunk<sup>17</sup>, Edward R.B. Moore<sup>18</sup>, Lily Eurwilaichitr<sup>19,23</sup>  
20  
21 Supawadee Ingsriswang<sup>19</sup>, Heena Shah<sup>20</sup>, Su Yao<sup>21</sup>, Tao Jin<sup>7</sup>, Jinqun Huang<sup>7</sup>, Wenyu  
22  
23 Shi<sup>1</sup>, Qinglan Sun<sup>1</sup>, Guomei Fan<sup>1</sup>, Wei Li<sup>1</sup>, Xian Li<sup>1</sup>, İpek Kurtböke<sup>22,23\*</sup> & Juncai Ma<sup>1,</sup>  
24  
25  
26  
27  
28  
29  
30  
31 2,3,23\*

32  
33  
34 #To whom correspondence should be addressed: İpek Kurtböke  
35  
36 [IKurtbok@usc.edu.au](mailto:IKurtbok@usc.edu.au) & Juncai Ma [ma@im.ac.cn](mailto:ma@im.ac.cn)  
37  
38  
39  
40  
41

42 1 Microbial Resource and Big Data Center, Institute of Microbiology, Chinese  
43  
44 Academy of Sciences, Beijing 100101, China  
45  
46

47 2 State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese  
48  
49 Academy of Sciences, Beijing 100101, China  
50  
51  
52

53 3 WFCC-MIRCEN World Data Center for microorganisms, WDCM, Beijing 100101,  
54  
55 China  
56  
57

58 4 Fungal Genetics Stock Center, Kansas State University, Manhattan KS 66506, USA  
59  
60  
61  
62  
63  
64  
65

1 5 Belgian Coordinated Collections of Micro-organisms Programme, Belgian, Science  
2  
3 Policy Office, Brussels 231 1050, Belgium  
4

5  
6 6 National Institute of Genetics, Yata, Mishima 411-8540, Japan  
7

8  
9 7 BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China  
10

11 8 Japan Collection of Microorganisms/ Microbe Division, RIKEN BioResource Center,  
12  
13 Koyadai 3-1-1, Tsukuba, Ibaraki 305-0074, Japan  
14

15  
16 9 Korean Collection for Type Cultures (KCTC), Korea Research Institute of Bioscience  
17  
18 and 30 Biotechnology (KRIBB), 181 Ipsin-gil, Jeongeup-si, Jeollabuk-do 56212,  
19  
20 Republic of Korea  
21  
22

23  
24 10 China General Microbiological Culture Collection Center, Institute of Microbiology,  
25  
26 Chinese Academy of Sciences, Beijing10010, China  
27

28  
29 11 NITE Biological Resource Center (NBRC), National Institute of Technology and 24  
30  
31 Evaluation, 2-5-8 Kazusakamatari, Kisarazu, Chiba 292-0818, Japan  
32  
33

34  
35 12 American Type Culture Collection, 10801 University Boulevard, Manassas, VA  
36  
37 20110, USA  
38  
39

40  
41 13 Westerdijk Fungal Biodiversity Institute, Utrecht 3534CT, Netherlands  
42

43  
44 14 Micoteca da Universidade do Minho, Biological Engineering Centre, 4710-057  
45  
46 Braga, Portugal  
47  
48

49  
50 15 All-Russian Collection of Microorganisms, G.K. Skryabin Institute of Biochemistry  
51  
52 and Physiology of Microorganisms RAS, Pushchino, Moscow region 142290, Russia  
53  
54

55  
56 16 Phaff Yeast Culture Collection Food Science & Technology Department University  
57  
58 of California Davis One Shields Avenue, Davis, CA 95616-8598, USA  
59  
60

61  
62  
63  
64  
65

1 17 Leibniz-Institute DSMZ – German Collection of Microorganisms and Cell Cultures,  
2  
3 D-38124 Braunschweig, Germany  
4  
5

6 18 Culture Collection University of Gothenburg (CCUG), Sahlgrenska Academy of the  
7  
8 University of Gothenburg, SE-41346 Gothenburg, Sweden  
9  
10

11 19 Bioresources Technology Unit Thailand Bioresource Research Center (TBRC),  
12  
13 National Center for Genetic Engineering and Biotechnology (BIOTEC), Bangkok  
14  
15 National Science and Technology Development Agency (NSTDA), 113, Thailand  
16  
17  
18

19 20 National Collection of Type Cultures (NCTC), Public Health England (PHE),  
20  
21 Porton Down,Salisbury,Wiltshire,SP4 0JG, UK  
22  
23  
24

25 21 China Center of Industrial Culture Collection (CICC), Beijing 100015, China  
26  
27

28 22 Faculty of Science, Health, Education and Engineering University of the Sunshine  
29  
30 Coast Maroochydore DC Queensland 4558, Australia  
31  
32

33 23 World Federation of Culture Collections (WFCC)  
34  
35

36 24 Institute of Biodiversity and Ecosystem Dynamics (IBED), University of  
37  
38 Amsterdam, Spui 21 1012 WX Amsterdam, The Netherlands  
39  
40

41 25 Shanghai Key Laboratory of Molecular Medical Mycology, Shanghai Institute of  
42  
43 Mycology, Shanghai Changzheng Hospital, Shanghai 200003, China  
44  
45  
46

47  
48  
49  
50 Email address of the authors:

51  
52  
53 Linhuan Wu: [wulh@im.ac.cn](mailto:wulh@im.ac.cn)  
54

55  
56 Kevin McCluskey: [mcccluskeyk@ksu.edu](mailto:mcccluskeyk@ksu.edu)  
57

58  
59 Philippe Desmeth: [Philippe.DESMETH@belspo.be](mailto:Philippe.DESMETH@belspo.be)  
60  
61  
62  
63  
64  
65

1 Shuangjiang Liu: [liusj@im.ac.cn](mailto:liusj@im.ac.cn)  
2  
3 Sugawara Hideaki: [hsugawar@nig.ac.jp](mailto:hsugawar@nig.ac.jp)  
4  
5  
6 Ye Yin: [yinye@genomics.cn](mailto:yinye@genomics.cn)  
7  
8  
9 Ohkuma Moriya: [mohkuma@riken.jp](mailto:mohkuma@riken.jp)  
10  
11  
12 Takashi Itoh: [ito@jcm.riken.jp](mailto:ito@jcm.riken.jp)  
13  
14  
15 Cha Young Kim: [kimcy@kribb.re.kr](mailto:kimcy@kribb.re.kr)  
16  
17 Jung-Sook Lee: [jslee@kribb.re.kr](mailto:jslee@kribb.re.kr)  
18  
19  
20 Yuguang Zhou: [zhouyg@im.ac.cn](mailto:zhouyg@im.ac.cn)  
21  
22  
23 Hiroko Kawasaki: [kawasaki-hiroko@nite.go.jp](mailto:kawasaki-hiroko@nite.go.jp)  
24  
25  
26 Manzour Hernando Hazbón: [mhazbon@atcc.org](mailto:mhazbon@atcc.org)  
27  
28  
29 Vincent Robert: [v.robert@westerdijkinstituut.nl](mailto:v.robert@westerdijkinstituut.nl)  
30  
31 Teun Boekhout: [t.boekhout@westerdijkinstituut.nl](mailto:t.boekhout@westerdijkinstituut.nl)  
32  
33  
34 Nelson Lima: [nelson@ie.uminho.pt](mailto:nelson@ie.uminho.pt)  
35  
36  
37 Lyudmila Evtushenko: [lie99@mail.ru](mailto:lie99@mail.ru)  
38  
39  
40 Kyria Boundy-Mills: [klbmills@ucdavis.edu](mailto:klbmills@ucdavis.edu)  
41  
42  
43 Boyke Bunk: [boyke.bunk@dsmz.de](mailto:boyke.bunk@dsmz.de)  
44  
45  
46 Edward R.B. Moore: [erbmoore@ccug.se](mailto:erbmoore@ccug.se)  
47  
48  
49 Lily Eurwilaichitr: [lily@biotec.or.th](mailto:lily@biotec.or.th)  
50  
51  
52 Supawadee Ingsriswang: [supawadee@biotec.or.th](mailto:supawadee@biotec.or.th)  
53  
54  
55 Heena Shah: [Heena.Shah@phe.gov.uk](mailto:Heena.Shah@phe.gov.uk)  
56  
57  
58 Su Yao: [milly@china-cicc.org](mailto:milly@china-cicc.org)  
59  
60  
61  
62  
63  
64  
65 Tao Jin: [jintao@genomics.cn](mailto:jintao@genomics.cn)

1 Jinqun Huang: [huangjinqun@genomics.cn](mailto:huangjinqun@genomics.cn)

2  
3 Wenyu Shi: [shiwu@im.ac.cn](mailto:shiwu@im.ac.cn)

4  
5  
6 Qinglan Sun: [sunql@im.ac.cn](mailto:sunql@im.ac.cn)

7  
8  
9 Guomei Fan: [gmfan@im.ac.cn](mailto:gmfan@im.ac.cn)

10  
11 Wei Li: [liweil2227@im.ac.cn](mailto:liweil2227@im.ac.cn)

12  
13 Xian Li: [lixian@im.ac.cn](mailto:lixian@im.ac.cn)

14  
15 İpek Kurtböke: [IKurtbok@usc.edu.au](mailto:IKurtbok@usc.edu.au)

16  
17 Juncai Ma: [ma@im.ac.cn](mailto:ma@im.ac.cn)

## 22 23 24 25 **Abstract**

26  
27  
28 Genomic information is essential to underpin taxonomic, phylogenetic and functional  
29  
30 studies to comprehensively decipher the characteristics of microorganisms, to explore  
31  
32 microbiomes through metagenomics and also answer fundamental questions of nature  
33  
34 and human life. However, there is still a big gap of genomic sequencing information  
35  
36 for the bacterial and archaeal species with validly published names. For fungal type  
37  
38 strains, the remaining gap is even larger. Global Catalogue of Microorganisms (GCM)'s  
39  
40 led type strain sequencing project is an international coordinated effort to close current  
41  
42 gaps in the genomic maps of microbes and hence to promote research through deep-  
43  
44 mining of the genomic data.  
45  
46  
47  
48  
49  
50  
51

52  
53 **Keywords:** Phylogenomics, Taxonomy, Biodiversity, Whole-genome sequencing,  
54  
55 Type strains, Bacteria, Archaea, Fungi  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



## Background

Microorganisms are the most abundant organisms on the earth. The total prokaryotic diversity may comprise up to  $10^9$  species<sup>[1]</sup>. For prokaryotic species, only new names published in the International Journal of Systematic and Evolutionary Microbiology (IJSEM) as an original article or in the 'Validation Lists' are considered to be valid. At the end of 2017, the number of validly published species and subspecies of prokaryotes was 15,081 while that was 12,981 at the end of December 2015<sup>[2]</sup>, increasing by 1088 and 1012 respectively in the year of 2016 and 2017. The most commonly accepted estimation on the number of existing fungal species is the 2.27 million species hypothesized by Hawksworth<sup>[3]</sup> while the number of species reported in the Dictionary of Fungi is only about 100,000.

The taxonomy of microorganisms, including classification, identification and nomenclature, has developed from morphological and metabolic characterization, numerical taxonomy based on phenetic analyses, then incorporated chemotaxonomy, and finally to polyphasic approaches combining phenotypic, chemotaxonomic, genotypic and genomic information. The IJSEM recently announced that starting January 2018, authors of new taxa descriptions are asked to provide genome sequence data with descriptions of novel taxa for manuscript submission<sup>[4]</sup>. As such, taxonomy of microorganisms thus entered into the genomics era. A genomic 'gold standard' for setting a consistent microbial species definition are urgently needed. To meet this end, sequencing the Type strains of the validly published prokaryotic and fungal species is

1 a foundational step.

2  
3 Besides the microorganisms that can be isolated and maintained *in situ*, the vast  
4 majority of microorganisms are not yet cultivated, and thus are relatively poorly studied.

5  
6 Culture-independent approaches have been developed to investigate the compositions  
7 and functions of the environmental and human microbiomes. However, accurate  
8 taxonomic and functional predictions based upon metagenomic data are dependent  
9 upon the availability of high-quality reference genomic data<sup>[5]</sup>. Therefore, genome  
10 sequencing of the Type strains of recognized microbial species will provide a  
11 taxonomic context for metagenomic data analysis, which is commonly comprised by  
12 short and incomplete sequences from complex environmental communities.

13  
14 Microorganisms, possessing extensive genomic and metabolic diversity are ideal tools  
15 for biotechnology. Decoding the full genomes of reference Type strains of species will  
16 thus enable association of gene content with function, such as metabolic activity,  
17 virulence, antibiotic production or resistance, biomass deconstruction, cellulose  
18 agricultural nitrogen fixation and the liberation of environmental phosphorus, etc.

19  
20 Access to microbial genomic sequences will significantly contribute to future studies  
21 of microbial biology, ecology, and biochemistry and these will, in turn, accelerate the  
22 discovery of new natural products and drugs<sup>[5]</sup>.

## 23 **Current status of strain sequencing project**

24  
25 The descriptions of prokaryotic species are based on living cultures and one  
26 representative strain is designated as the nomenclatural “type”. It is a requirement of  
27 IJSEM and the International Committee on Systematics of Prokaryotes (ICSP) that the

1 Type strains of new species should be deposited in, at least, two recognized collections  
2  
3 in two different countries. The Type strains of 15,081 prokaryotic species are widely  
4  
5 preserved in more than 130 culture collections. In mycology the trend is similar,  
6  
7 although hitherto a fungal type specimen has to be metabolically inactive.  
8  
9

10  
11 At present, the selection of strains for whole-genome sequencing has been based  
12  
13 predominantly on medical, ecological or industrial importance, which, often leads to  
14  
15 bias in assessing phylogenetic relationships. There are several ongoing phylogenetic  
16  
17 based microbial sequencing projects. The Genomic Encyclopedia of Bacteria and  
18  
19 Archaea (GEBA) led by the US Department of Energy Joint Genome Institute (US DOE  
20  
21 JGI) has pioneered the partnership between culture collections and sequencing projects.  
22  
23

24  
25 The GEBA project has published the outcomes of its first stage 1,003<sup>[7]</sup> Type strain  
26  
27 whole-genome sequences at year 2017. The newly started third-stage of GEBA has a  
28  
29 focus on the genomes of soil, plant-associated, and newly described Type strains<sup>[8]</sup>.  
30  
31

32  
33 Similarly, the US DOE JGI collaborated with international research teams conducted a  
34  
35 five-year project to sequence 1000 fungal genomes from across the Fungal Tree of  
36  
37 Life<sup>[9]</sup>. The overall plan is to fill in gaps in the Fungal Tree of Life by sequencing at  
38  
39 least two reference genomes from the more than 500 recognized families of fungi.  
40  
41

42  
43 A large number of microbial species Type strains remain to un-sequenced. Hence, the  
44  
45 World Federation of Culture Collections (WFCC) and the World Data Centre for  
46  
47 Microorganisms (WDCM) have initiated an international community-led project  
48  
49 towards full genome sequencing of microbial Type strains to support the continued  
50  
51 scientific discovery and biotechnology utilization. Considering the wide distribution of  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Type strains, international cooperation of the global culture collection community is  
2  
3 essential for success.  
4  
5

## 6 **Emerging enhancements of culture collection in the genomic era**

7

8  
9 Current efforts from culture collections to explore the diversity of microorganisms and  
10  
11 to harness their genes, properties and products are still insufficient. While type  
12  
13 collections are not always large and diverse, the genome sequence efforts of GCM will  
14  
15 obviously make the resources in smaller collections more accessible.  
16  
17

18  
19 The World Data Center for Microorganisms (WDCM) is the data center of the World  
20  
21 Federation for Culture Collections (WFCC) and the Microbial Resources Centers  
22  
23 Network (MIRCEN). The WDCM is working on facilitating the application of cutting-  
24  
25 edge information technology to improve the interoperability of microbial data, to  
26  
27 promote access and use of data, and coordinate international cooperation between  
28  
29 culture collections, scientists and other user communities. The first stage of the Global  
30  
31 Catalogue of Microorganisms (GCM) project started in 2012 which focused on the data  
32  
33 sharing of strain catalogue from culture collections<sup>[10]</sup>. The proposed Type strain  
34  
35 sequencing project is the continuation of the GCM project as its second stage, GCM  
36  
37  
38  
39  
40  
41  
42  
43  
44 2.0.

## 45 **Project development and current progress**

46

47  
48 The project was firstly announced during the 14<sup>th</sup> International Culture Collections  
49  
50 Conference held in July 2017 in Singapore in conjunction with the IUMS conferences.  
51  
52

53  
54 After that, a launching ceremony in Beijing was held in October 2017. Culture  
55  
56 collections including ATCC, BCCM, CBS, CCUG, CGMCC, DSMZ, FGSC, JCM,  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 KCTC, MUM NBRC, and TBRC sent representatives to join the launching ceremony  
2  
3 and had thorough discussions on the organization and operational procedures of the  
4  
5 GCM 2.0 project. Some culture collections, such as CICC, NCTC, TISTR, UCD-FST,  
6  
7 and VKM, although not present in the launching ceremony, expressed their willingness  
8  
9 to join the cooperation.  
10  
11  
12

13  
14 The entire project is expected to be completed in 5 years, including a pilot stage in the  
15  
16 first year. The GCM 2.0 includes two core subprojects, the 10K bacterial and archaeal  
17  
18 Type strain sequencing and the fungal Type strain sequencing. It will also embrace  
19  
20 several satellite projects based on specific scientific targets. GCM 2.0 coordinates its  
21  
22 projects through a steering committee and five interlinked working groups: 1) Bacteria  
23  
24 Selection, 2) Fungi Selection, 3) Standard Operational Procedures (SOPs), 4) Database  
25  
26 and, 5) Intellectual Property Right and Legal Issue.  
27  
28  
29  
30  
31

32  
33 The project has established SOPs for DNA extraction, sample submission, sequencing  
34  
35 and data processing to ensure that all the genetic resources, data and associated  
36  
37 metadata of Type strains are appropriately obtained, recorded and stored. A project of  
38  
39 ISO standard, proposed by the WDCM, AWI 20170: Specification on data integration  
40  
41 and publication in microbial resource centres is under development. Data generated  
42  
43 from this project will be published on the GCM portal and also deposited in public  
44  
45 databases, NCBI/EMBL/DDBJ and others.  
46  
47  
48  
49  
50  
51

52  
53 All validly published bacterial and archaeal type strains, as well as selected reference  
54  
55 fungal type strains frequently used for functional or phylogenetic studies will be on the  
56  
57 candidate sequencing list. Each of these strains has the documentation issued by the  
58  
59  
60  
61  
62  
63  
64  
65

1 providing culture collection, which assures purity and identity as the Type strain.  
2  
3 Beijing Genomics Institute (BGI) will support the microbial genome sequencing and  
4  
5 part of data analysis for this project. Sampling works for the pilot-stage have been  
6  
7 initiated. The project has established a global network to collect approximately 800  
8  
9 candidate Type strain samples of collections from Belgium, China, Japan, Korea, The  
10  
11 Netherlands, Portugal, Russia, Sweden, Thailand, USA, and UK. Although extracted  
12  
13 DNA samples are highly preferred, sending the cultured cell samples of the strains is  
14  
15 also acceptable. Importantly, GCM 2.0 will respect all countries regulations regarding  
16  
17 Access and Benefit-Sharing (ABS) under Nagoya Protocol.  
18  
19  
20  
21  
22  
23  
24

25 The GCM Type strain sequencing project encourages all culture collections to  
26  
27 participate in this international cooperation project. Interested parties should be willing  
28  
29 to provide DNA for Type strains in their collections. All microbiologists and institutions  
30  
31 from related fields are welcome to submit subprojects for genomic data related research  
32  
33 questions. Brief proposals, including questions to be addressed and Type strains to be  
34  
35 sequenced and analyzed could be e-mailed to Dr. Juncai Ma at ma@im.ac.cn. Once a  
36  
37 proposal is granted as a subproject, the scientist(s) will be asked to lead the full genome  
38  
39 analysis and jointly publish the generated outcomes.  
40  
41  
42  
43  
44  
45  
46

## 47 **Conclusion:**

48  
49 As a collaborative network of international culture collections, GCM 2.0 will contribute  
50  
51 to a genome-based microbial taxonomic framework, establishing high-quality complete  
52  
53 genome sequences as the new gold standard. The resulting knowledge and tools  
54  
55 generated through this project will not only directly facilitate identifications of  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 microorganisms, but also improve our ability to predict new genes complexes as well  
 2  
 3 as functions from microbial communities, thus expanding our knowledge of the hitherto  
 4  
 5  
 6 undiscovered majority of microbial diversity that may finally lead to sustainable  
 7  
 8  
 9 utilization of microbial resources for human benefit.

## 14 **List of Abbreviations**

17 <b>Acronym</b>	18 <b>Full Name/Institution</b>	19 <b>Country</b>	20 <b>Homepage</b>
21 ATCC	American Type Culture Collection	USA	<a href="http://www.atcc.org">http://www.atcc.org</a>
22 BCCM	Belgian Coordinated Collections of Microorganisms	Belgium	<a href="http://bccm.belspo.be/">http://bccm.belspo.be/</a>
23	Westerdijk Fungal Biodiversity Institute, formerly CBS-		
24 CBS	KNAW Fungal Biodiversity Centre	Netherlands	<a href="http://www.westerdijkinstituut.nl/">http://www.westerdijkinstituut.nl/</a>
25	Culture Collection University of Gothenburg, Sahlgrenska		
26 CCUG	Academy of the University of Gothenburg	Sweden	<a href="http://www.ccug.se/">http://www.ccug.se/</a>
27			
28 CGMCC	China General Microbiological Culture Collection Center	China	<a href="http://www.cgmcc.net">http://www.cgmcc.net</a>
29			
30 CICC	China Center of Industrial Culture Collection	China	<a href="http://www.china-cicc.org">http://www.china-cicc.org</a>
31			
32	Leibniz-Institut DSMZ-Deutsche Sammlung von		
33 DSMZ	Mikroorganismen und Zellkulturen	Germany	<a href="http://www.dsmz.de">http://www.dsmz.de</a>
34			
35	Fungal Genetics Stock Center, Kansas State University,		
36 FGSC	Department of Plant Pathology	USA	<a href="http://www.fgsc.net/">http://www.fgsc.net/</a>
37			
38	Japan Collection of Microorganisms, RIKEN BioResource		
39 JCM	Center	Japan	<a href="http://jcm.brc.riken.jp/">http://jcm.brc.riken.jp/</a>
40			
41			
42 KCTC	Korean Collection for Type Cultures, Korea Research	Korea	<a href="http://kctc.kribb.re.kr/English/inde">http://kctc.kribb.re.kr/English/inde</a>

1		Institute of Bioscience and Biotechnology		x.aspx
2				
3				
4				
5				
6				<a href="http://www.micoteca.deb.uminho">http://www.micoteca.deb.uminho.</a>
7	MUM	Micoteca da Universidade do Minho, Universidade do Minho	Portugal	
8				pt
9				
10				
11				
12				
13				
14				
15				
16				
17		Biological Resource Center, Department of Biotechnology,		<a href="http://www.nite.go.jp/en/nbrc/inde">http://www.nite.go.jp/en/nbrc/inde</a>
18	NBRC		Japan	
19		National Institute of Technology and Evaluation		x.html
20				
21				
22				
23				
24				
25		National Collection of Type Cultures, Culture Collections,		www.phe-
26	NCTC		UK	
27		Public Health England		culturecollections.org.uk
28				
29				
30				
31		Thailand Bioresource Research Center, National Center for		
32	TBRC		Thailand	<a href="http://www.tbrcnetwork.org">www.tbrcnetwork.org</a>
33		Genetic Engineering and Biotechnology		
34				
35				
36		TISTR Culture Collection (Bangkok MIRCEN), Thailand		
37	TISTR		Thailand	<a href="http://www.tistr.or.th/tistr_culture">http://www.tistr.or.th/tistr_culture</a>
38		Institute of Scientific and Technological Research		
39				
40				
41				
42		Phaff Yeast Culture Collection, Department of Food Science		
43	UCD-FST		USA	<a href="http://phaffcollection.ucdavis.edu/">http://phaffcollection.ucdavis.edu/</a>
44		and Technology, University of California Davis		
45				
46				
47		All-Russian Collection of Microorganisms, Institute of		
48				
49				
50	VKM	Biochemistry and Physiology of Microorganisms, Russian	Russia	<a href="http://www.vkm.ru/">http://www.vkm.ru/</a>
51				
52				
53		Academy of Sciences		

---

## Ethics approval and consent to participate

60  
61  
62  
63  
64  
65



1 Not applicable  
2  
3  
4  
5

## 6 **Consent for publication** 7

8 Not applicable  
9  
10  
11  
12

## 13 **Availability of data and material** 14

15 Not applicable  
16  
17  
18  
19  
20  
21

## 22 **Acknowledgements** 23

24  
25 Authors here thank the WFCC Executive Board for supports, and also acknowledge the  
26  
27 GCM type strain sequencing project members, partners, and advisors, who have made  
28  
29 the cooperation possible. We thank Takashi Gojobori, Guoping Zhao, Xuewei Xu,  
30  
31 Xiaoyang Zhi, Hua Xiang, Lei Cai for their insightful suggestions.  
32  
33  
34  
35  
36  
37  
38

## 39 **Funding** 40

41  
42 This work is supported by the Strategic Priority Research Program of the Chinese  
43  
44 Academy of Sciences (Grant No. XDA19050301), the Bureau of International  
45  
46 Cooperation, Chinese Academy of Sciences (Grant No. 153211KYSB20160029,  
47  
48 153211KYSB20150010), the National key Research Program of China (Grant No  
49  
50 2017YFC1201202, 2016YFC1201303, 2016YFC0901702), the 13th Five-year  
51  
52 Informatization Plan of Chinese Academy of Sciences(Grant No. XXH13506), the  
53  
54 National Science Foundation for Young Scientists of China(Grant No. 31701157).  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3 **Competing interests**  
4  
5

6 The authors declare that they have no competing interests.  
7  
8  
9

10  
11 **Authors' contributions**  
12  
13

14 L.H.W drafted the original manuscript text with detailed input from other authors. All  
15 authors participated in the GCM type strain sequencing project and have read and  
16  
17 approved the final manuscript  
18  
19  
20  
21  
22  
23  
24

25 **References**  
26  
27

- 28 1. Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. Proc Natl  
29 Acad Sci USA. 2002; 99:10494–9. doi:10.1073/pnas.142680199.  
30  
31 2. Garrity GM. A New Genomics-Driven Taxonomy of Bacteria and Archaea: Are We There Yet?  
32 J Clin Microbiol. 2016; 54: 1956-63. doi: 10.1128/JCM.00200-16.  
33  
34 3.Hawksworth DL. The magnitude of fungal diversity: the 1.5 million species estimate revisited.  
35 Mycol Res. 2001; 105:1422–1432.  
36  
37 4. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal  
38 standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol.  
39 2018, 68:461–466. doi: 10.1099/ijsem.0.002516.  
40  
41 5. Kim Y, Koh I, Rho M. Deciphering the human microbiome using next-generation sequencing  
42 data and bioinformatics approaches. Methods. 2015; 79-80: 52-9.  
43  
44 doi: 10.1016/j.ymeth.2014.10.022.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 6. Kang HS. Phylogeny-guided (meta)genome mining approach for the targeted discovery of new  
2  
3 microbial natural products. *J Ind Microbiol Biotechnol.* 2017; 44:285-293. doi: 10.1007/s10295-  
4  
5 016-1874-z.  
6

7  
8  
9 7. Mukherjee S, Seshadri R, Varghese NJ, Eloje-Fadrosch EA, Meier-Kolthoff JP, Göker M, et al.  
10  
11 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat*  
12  
13 *Biotechnol.* 2017; 35:676–683. doi: 10.1038/nbt.3886.  
14  
15

16  
17 8. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, et al. *Genomic*  
18  
19 *Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil and plant-*  
20  
21 *associated and newly described type strains.* *Stand Genomic Sci.* 2015; 10:26. doi:  
22  
23 10.1186/s40793-015-0017-x.  
24  
25

26  
27  
28 9. 1000 fungal genomes Project. [https://jgi.doe.gov/our-science/science-programs/fungal-](https://jgi.doe.gov/our-science/science-programs/fungal-genomics/1000-fungal-genomes/)  
29  
30 [genomics/1000-fungal-genomes/](https://jgi.doe.gov/our-science/science-programs/fungal-genomics/1000-fungal-genomes/). Accessed March 1, 2017 .  
31  
32

33  
34 10. Wu L, Sun Q, Desmeth P, Sugawara H, Xu Z, McCluskey K, et al. World data centre for  
35  
36 *microorganisms: an information infrastructure to explore and utilize preserved microbial strains*  
37  
38 *worldwide.* *Nucleic Acids Res.* 2017; 45: D611-D618. doi: 10.1093/nar/gkw903.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65