

GigaScience

The Global Catalogue of Microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species --Manuscript Draft--

Manuscript Number:	GIGA-D-18-00079R1	
Full Title:	The Global Catalogue of Microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species	
Article Type:	Commentary	
Funding Information:	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
Abstract:	Genomic information is essential for taxonomic, phylogenetic and functional studies to comprehensively decipher the characteristics of microorganisms, to explore microbiomes through metagenomics, and to answer fundamental questions of nature and human life. However, large gaps remain in the available genomic sequencing information published for bacterial and archaeal species, and the gaps are even larger for fungal type strains. The Global Catalogue of Microorganisms (GCM) leads an internationally coordinated effort to sequence type strains and close gaps in the genomic maps of microbes. Hence, the GCM aims to promote research by deep-mining genomic data.	
Corresponding Author:	Juncai Ma CHINA	
Corresponding Author Secondary Information:		
Corresponding Author's Institution:		
Corresponding Author's Secondary Institution:		
First Author:	Linhuan Wu	
First Author Secondary Information:		
Order of Authors:	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
Order of Authors Secondary Information:	
Response to Reviewers:	<p>The full post information and ORCID IDs of the Correspondent authors are provided.</p> <p>All the revisions suggested by the editors are accepted expect somewhere I added the explanations(highlighted in coloured text in the manuscript).</p> <p>A version with track changes mode and a clear version were uploaded.</p>
Additional Information:	
Question	Response
Are you submitting this manuscript to a special series or article collection?	No
Experimental design and statistics	Yes
Full details of the experimental design and	

<p>statistical methods used should be given in the Methods section, as detailed in our Minimum Standards Reporting Checklist. 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>Resources</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 Research Resource Identifiers (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our Minimum Standards Reporting Checklist?</p>	<p>Yes</p>
<p>Availability of data and materials</p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in publicly available repositories (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 Minimum Standards Reporting Checklist?</p>	<p>Yes</p>

**The Global Catalogue of Microorganisms 10K type strain sequencing project:
closing the genomic gaps for the validly published prokaryotic and fungi species**

Linhuan Wu^{1,2,3}, Kevin McCluskey^{4,5}, Philippe Desmeth^{4,6}, Shuangjiang Liu^{2,4},
Sugawara Hideaki⁷, Ye Yin⁸, Ohkuma Moriya⁹, Takashi Itoh⁹, Cha Young Kim¹⁰,
Jung-Sook Lee¹⁰, Yuguang Zhou¹¹, Hiroko Kawasaki¹², Manzour Hernando Hazbón¹³,
Vincent Robert¹⁴, Teun Boekhout^{14,15,16}, Nelson Lima¹⁷, Lyudmila Evtushenko¹⁸,
Kyria Boundy-Mills^{4,19}, Boyke Bunk²⁰, Edward R.B. Moore²¹, Lily Eurwilaichitr^{4,22}
Supawadee Ingsriswang²², Heena Shah²³, Su Yao²⁴, Tao Jin⁸, Jinqun Huang⁸, Wenyu
Shi¹, Qinglan Sun¹, Guomei Fan¹, Wei Li¹, Xian Li¹, İpek Kurtböke^{4,25*}, Juncai
Ma^{1,2,3,4*}

¹Microbial Resource and Big Data Center, Institute of Microbiology, Chinese
Academy of Sciences, Beijing 100101, China

²State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese
Academy of Sciences, Beijing 100101, China

³WFCC-MIRCEN World Data Center for Microorganisms, Beijing 100101, China

⁴World Federation of Culture Collections (WFCC)

⁵Fungal Genetics Stock Center, Kansas State University, Manhattan, KS 66506,

USA⁶Belgian Coordinated Collections of Micro-organisms Program, Belgian Science
Policy Office, Brussels 231 1050, Belgium

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

1 ⁸BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China

2
3 ⁹Japan Collection of Microorganisms/Microbe Division, RIKEN BioResource Center,
4
5
6 Koyadai 3-1-1, Tsukuba, Ibaraki 305-0074, Japan

7
8
9 ¹⁰Korean Collection for Type Cultures, Korea Research Institute of Bioscience and
10
11
12 Biotechnology (KRIBB), 181 Ipsin-gil, Jeongeup-si, Jeollabuk-do, 56212, Republic
13
14 of Korea

15
16
17 ¹¹China General Microbiological Culture Collection Center, Institute of Microbiology,
18
19
20 Chinese Academy of Sciences, Beijing10010, China

21
22 ¹²NITE Biological Resource Center, National Institute of Technology and
23
24
25 Evaluation, 2-5-8 Kazusakamatari, Kisarazu, Chiba 292-0818, Japan

26
27
28 ¹³American Type Culture Collection, 10801 University Boulevard, Manassas, VA
29
30
31 20110, USA

32
33
34 ¹⁴Westerdijk Fungal Biodiversity Institute, Utrecht 3534CT, Netherlands

35
36
37 ¹⁵Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Spui
38
39
40 21 1012 WX Amsterdam, Netherlands

41
42 ¹⁶Shanghai Key Laboratory of Molecular Medical Mycology, Shanghai Institute of
43
44
45 Mycology, Shanghai Changzheng Hospital, Shanghai 200003, China

46
47
48 ¹⁷Micoteca da Universidade do Minho, Biological Engineering Centre, 4710-057
49
50
51 Braga, Portugal

52
53 ¹⁸All-Russian Collection of Microorganisms, GK Skryabin Institute of Biochemistry
54
55
56 and Physiology of Microorganisms RAS, Pushchino, Moscow Region 142290, Russia

57
58
59 ¹⁹Phaff Yeast Culture Collection, Food Science and Technology Department,
60
61
62
63
64
65

1 University of California Davis, 1 Shields Avenue, Davis, CA 95616-8598, USA

2
3 ²⁰Leibniz-Institute DSMZ – German Collection of Microorganisms and Cell Cultures,
4
5
6 D-38124 Braunschweig, Germany

7
8
9 ²¹Culture Collection University of Gothenburg (CCUG), Sahlgrenska Academy of the
10
11 University of Gothenburg, SE-41346 Gothenburg, Sweden

12
13
14 ²²Bioresources Technology Unit, Thailand Bioresource Research Center, National
15
16
17 Center for Genetic Engineering and Biotechnology, Bangkok National Science and
18
19
20 Technology Development Agency, 113, Thailand

21
22
23 ²³National Collection of Type Cultures, Public Health England, Porton Down,
24
25
26 Salisbury, Wiltshire SP4 0JG, UK

27
28 ²⁴China Center of Industrial Culture Collection, Beijing 100015, China

29
30
31 ²⁵Faculty of Science, Health, Education and Engineering, University of the Sunshine
32
33
34 Coast, Maroochydore, Queensland 4558, Australia

35
36
37
38
39 *Correspondence address. İpek Kurtböke [Faculty of Science, Health, Education and
40
41
42 Engineering , University of the Sunshine Coast Locked Bag 4, MAROOCHYDORE
43
44
45 DC QLD 4558, AUSTRALIA]. Tel: [+61 07 5430 2819]; IKurtbok@usc.edu.au.

46
47
48 ORCID ID: 0000-0003-2056-2484

49
50
51 Juncai Ma [NO.1 Beichen West Road, Chaoyang District, Beijing 100101, China].

52
53
54 Tel: [+86 10 64807422]; E-mail: ma@im.ac.cn. ORCID ID: [0000-0001-6382-8014](https://orcid.org/0000-0001-6382-8014)

55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1 Linhuan Wu: wulh@im.ac.cn
2
3 Kevin McCluskey: mccluskeyk@ksu.edu
4
5
6 Philippe Desmeth: Philippe.DESMETH@belspo.be
7
8
9 Shuangjiang Liu: liusj@im.ac.cn
10
11
12 Sugawara Hideaki: hsugawar@nig.ac.jp
13
14
15 Ye Yin: yinye@genomics.cn
16
17 Ohkuma Moriya: mohkuma@riken.jp
18
19
20 Takashi Itoh: ito@jcm.riken.jp
21
22
23 Cha Young Kim: kimcy@kribb.re.kr
24
25
26 Jung-Sook Lee: jslee@kribb.re.kr
27
28 Yuguang Zhou: zhouyg@im.ac.cn
29
30
31 Hiroko Kawasaki: kawasaki-hiroko@nite.go.jp
32
33
34 Manzour Hernando Hazbón: mhazbon@atcc.org
35
36
37 Vincent Robert: v.robert@westerdijkinstituut.nl
38
39 Teun Boekhout: t.boekhout@westerdijkinstituut.nl
40
41
42 Nelson Lima: nelson@ie.uminho.pt
43
44
45 Lyudmila Evtushenko: lie99@mail.ru
46
47
48 Kyria Boundy-Mills: klbmills@ucdavis.edu
49
50
51 Boyke Bunk: boyke.bunk@dsmz.de
52
53
54 Edward R.B. Moore: erbmoore@ccug.se
55
56
57 Lily Eurwilaichitr: lily@biotec.or.th
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1 Heena Shah: Heena.Shah@phe.gov.uk

2
3 Su Yao: milly@china-cicc.org

4
5
6 Tao Jin: jintao@genomics.cn

7
8
9 Jinqun Huang: huangjinqun@genomics.cn

10
11
12 Wenyu Shi: shiwy@im.ac.cn

13
14
15 Qinglan Sun: sunql@im.ac.cn

16
17
18 Guomei Fan: gmfan@im.ac.cn

19
20
21 Wei Li: liwei2227@im.ac.cn

22
23
24 Xian Li: lixian@im.ac.cn

25
26
27 İpek Kurtböke: IKurtbok@usc.edu.au

28
29
30 Juncai Ma: ma@im.ac.cn

31 32 33 **Abstract**

34
35
36 **Genomic information is essential for taxonomic, phylogenetic and functional**
37
38
39 **studies to comprehensively decipher the characteristics of microorganisms, to**
40
41
42 **explore microbiomes through metagenomics, and to answer fundamental**
43
44
45 **questions of nature and human life. However, large gaps remain in the available**
46
47
48 **genomic sequencing information published for bacterial and archaeal species,**
49
50
51 **and the gaps are even larger for fungal type strains. The Global Catalogue of**
52
53
54 **Microorganisms (GCM) leads an internationally coordinated effort to sequence**
55
56
57 **type strains and close gaps in the genomic maps of microbes. Hence, the GCM**
58
59
60 **aims to promote research by deep-mining genomic data.**
61
62
63
64
65

1 *Keywords:* Phylogenomics; Taxonomy; Biodiversity; Whole-genome sequencing;

2
3 Type strains; Bacteria; Archaea;Fungi
4
5
6
7

8 **Introduction**

9
10
11 Microorganisms are the most abundant organisms on Earth. The total diversity of
12
13 prokaryotes may comprise up to 10^9 species [1]. For prokaryotic species, only new
14
15 names published in the International Journal of Systematic and Evolutionary
16
17 Microbiology (IJSEM) as an original article or in the ‘Validation Lists’ are considered
18
19 valid. As of the end of 2017, a total of 15,081 valid prokaryotic species and
20
21 subspecies were published, compared to 12,981 at the end of December 2015 [2]. The
22
23 number of publications increased by 1088 in 2016, and by 1012 in 2017. The most
24
25 commonly accepted estimate of the number of existing fungal species is 2.27 million
26
27 species, as hypothesized by Hawksworth [3], while the number of species reported in
28
29 the Dictionary of Fungi is only about 100,000.
30
31

32
33
34 The taxonomy of microorganisms, including their classification, identification and
35
36 nomenclature, has developed from morphological and metabolic characterization, to
37
38 incorporate numerical taxonomy based on phenetic analyses, chemotaxonomy, and
39
40 finally polyphasic approaches combining phenotypic, chemotaxonomic, genotypic
41
42 and genomic information. The IJSEM recently announced that since January 2018,
43
44 authors of new taxa descriptions were asked to provide genome sequence data with
45
46 descriptions of novel taxa with their manuscript submissions [4]. As such, the
47
48 taxonomy of microorganisms has entered the genomics era. A genomic ‘gold standard’
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 for consistent microbial species definitions is urgently needed. To meet this end, a
2
3 fundamental step is to sequence the type strains of validly published prokaryotic and
4
5 fungal species.
6

7
8
9 Besides the microorganisms that can be isolated and maintained in situ, the vast
10
11 majority of microorganisms cannot yet be cultivated, and thus are relatively poorly
12
13 studied. Culture-independent approaches have been developed to investigate the
14
15 compositions and functions of environmental and human microbiomes. However,
16
17 accurate taxonomic and functional predictions based upon metagenomic data are
18
19 dependent upon the availability of high quality reference genomic data [5]. Therefore,
20
21 sequencing the genomes of type strains of recognized microbial species will provide a
22
23 taxonomic context for metagenomic data analysis, which is commonly comprised of
24
25 short and incomplete sequences from complex environmental communities.
26
27

28
29
30
31
32
33 Microorganisms possess extensive genomic and metabolic diversity, which makes
34
35 them ideal biotechnological tools. Decoding the full genomes of the type strains of
36
37 various species to provide reference genomes will thus enable genes to be associated
38
39 with functions, such as metabolic activity, virulence, antibiotic production or
40
41 resistance, biomass deconstruction, cellulose agricultural nitrogen fixation and the
42
43 liberation of environmental phosphorus, etc. Access to microbial genomic sequences
44
45 will significantly contribute to future studies in microbial biology, ecology, and
46
47 biochemistry and these will, in turn, accelerate the discovery of new natural products
48
49 and drugs [5].
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

The current status of the strain sequencing project

1
2
3
4 Descriptions of prokaryotic species are based on living cultures and one
5
6 representative strain is designated as the nomenclatural ‘type’. The IJSEM and
7
8 International Committee on Systematics of Prokaryotes requires that the type strains
9
10 of new species should be deposited in at least two recognized collections in two
11
12 different countries. The type strains of 15,081 prokaryotic species are widely
13
14 preserved in more than 130 culture collections. In mycology the trend is similar,
15
16 although hitherto a fungal type specimen must be metabolically inactive.
17
18
19 Presently, the selection of strains for whole-genome sequencing is based
20
21
22 predominantly on medical, ecological or industrial importance, which often leads to
23
24
25 bias in assessing phylogenetic relationships. There are several ongoing
26
27
28 phylogenetic-based microbial sequencing projects. The Genomic Encyclopedia of
29
30
31 Bacteria and Archaea (GEBA), led by the US Department of Energy Joint Genome
32
33
34 Institute (US DOE JGI), has pioneered the partnership between culture collections and
35
36
37 sequencing projects. The GEBA project published 1003 whole-genome sequences of
38
39
40 type strains in 2017 as the outcome of its first stage [7]. GEBA started the new stage
41
42
43 of the project in 2015, which has a focus on the genomes of soil, plant-associated, and
44
45
46 newly described type strains [8].
47
48
49 Similarly, the US DOE JGI, in collaboration with international research teams,
50
51
52 conducted a five-year project to sequence 1000 fungal genomes from across the
53
54
55 Fungal Tree of Life [9]. The overall plan is to fill in gaps in the Fungal Tree of Life by
56
57
58 sequencing at least two reference genomes from the more than 500 recognized
59
60
61
62
63
64
65

1 families of fungi.
2

3 Many type strains of microbial species remain un-sequenced. Hence, the World
4 Federation of Culture Collections (WFCC) and the World Data Centre for
5
6 Microorganisms (WDCM) have initiated an international community-led project to
7
8
9 sequence the full genomes of microbial type strains to support continued scientific
10
11
12 discovery and biotechnological utilization. Considering the wide distribution of type
13
14
15 strains, cooperation across the global culture collection community is essential for
16
17
18
19 success.
20
21
22
23
24

25 **Emerging enhancements of culture collection in the genomic era**

26
27 Efforts made by culture collection curators to explore the diversity of microorganisms
28
29 and to harness their genes, properties and products remain insufficient. While type
30
31
32 collections are not always large or diverse, the genome sequencing efforts of the
33
34
35 GCM will increase access to resources in smaller collections.
36
37

38
39 The WDCM is the data center of the WFCC and the Microbial Resources Centers
40
41 Network. The WDCM is working on facilitating the application of cutting-edge
42
43
44 information technology to improve the interoperability of microbial data, to promote
45
46
47 access and use of data, and coordinate international cooperation between culture
48
49
50 collections, scientists and other user communities. The first stage of the GCM project
51
52
53 started in 2012, which focused on sharing strain catalogue data from culture
54
55
56 collections [10]. The proposed type strain sequencing project is the continuation of
57
58
59 the GCM project as its second stage, GCM 2.0.
60
61
62
63
64
65

Project development and current progress

The project was first announced during the 14th International Culture Collections Conference, held in Singapore in July 2017 in conjunction with the International Union of Microbiological Societies conferences. Following that, in October 2017, a launch ceremony was held in Beijing, China, at which representatives from the following culture collections were present: the American Type Culture Collection (ATCC), Belgian Coordinated Collections of Microorganisms(BCCM), Biological Resource Center, Japan National Institute of Technology and Evaluation(NBRC), Culture Collection of the University of Gothenburg (CCUG), China General Microbiological Culture Collection Center(CGMCC), Fungal Genetics Stock Center(Kansas State University, USA)(FGSC), Japan Collection of Microorganisms(JCM), Korean Collection for Type Cultures(KCTC), Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen(DSMZ), Micoteca da Universidade do Minho (Portugal)(MUM), Thailand Bioresource Research Center(TBRC) and Westerdijk Fungal Biodiversity Institute (CBS). The meeting representatives held detailed discussions on the organization and operational procedures of the GCM 2.0 project. Some culture collections (e.g., All-Russian Collection of Microorganisms(VKM), China Center of Industrial Culture Collection(CICC), Phaff Yeast Culture Collection (University of California Davis, USA)(UCD-FST), Thailand Institute of Scientific and Technological Research Culture Collection(TISTR), and UK National Collection of Type Cultures(NCTC))

1 did not have a representative present at the launch ceremony, but expressed their
2
3 willingness to join the collaboration.
4
5

6 It is expected that the project will be completed within 5 years, including a pilot stage
7
8 in the first year. GCM 2.0 includes two core subprojects, sequencing 10,000 bacterial
9
10 and archaeal type strains, and sequencing fungal type strains. It will also embrace
11
12 several satellite projects on specific scientific targets. GCM 2.0 is coordinated by a
13
14 steering committee and five interlinked working groups: 1) Bacteria Selection, 2)
15
16 Fungi Selection, 3) Standard Operational Procedures, 4) Databases, and 5) Intellectual
17
18 Property Rights and Legal Issues.
19
20
21
22
23

24
25 The project has established standard operational procedures for DNA extraction,
26
27 sample submission, sequencing and data processing to ensure that all genetic
28
29 resources, data and metadata associated with type strains are appropriately obtained,
30
31 recorded and stored. A project proposed by the WDCM, 'AWI 20170: specification on
32
33 data integration and publication in microbial resource centers', which would meet
34
35 International Organization for Standardization standards, is under development. The
36
37 raw data and annotation results generated from this project will be published on the
38
39 GCM portal. Following norms established for genome projects coming from the
40
41 Bermuda Principles and Fort Lauderdale agreement the resulting genomic data will
42
43 also be made freely available in public databases including those maintained by the
44
45 National Center for Biotechnology Information, the European Molecular Biology
46
47 Laboratory, the DNA Data Bank of Japan after completion of data analysis and
48
49 annotation and ensuring it has met a set of quality criteria.
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 All validly published bacterial and archaeal type strains, as well as selected reference
2
3 fungal type strains that are frequently used for functional or phylogenetic studies, will
4
5 be on the list of candidates for sequencing. Each of these strains has documentation
6
7 issued by the providing culture collection, which assures the purity and identity of the
8
9 type strain. BGI-Shenzhen will support the microbial genome sequencing and assist
10
11 with the data analysis for this project. Sampling works for the pilot stage have been
12
13 initiated. The project has established a global network to collect approximately 800
14
15 candidate type strain samples from American, British, Belgian, Chinese, Dutch,
16
17 Japanese, Korean, Portuguese, Russian, Swedish and Thai collections. Although
18
19 extracted DNA samples are much preferred, it is also acceptable to send cultured cell
20
21 samples of the strains. Importantly, under the terms of Nagoya Protocol, GCM 2.0
22
23 will respect the access and benefit-sharing regulations of all countries.
24
25

26
27
28 The GCM type strain sequencing project encourages all culture collections to
29
30 participate in this international collaborative project. Interested parties should be
31
32 willing to provide DNA for type strains held in their collections. All microbiologists
33
34 and institutions from related fields are welcome to submit subprojects for genomic
35
36 data-related research questions. Brief proposals, including questions to be addressed
37
38 and type strains to be sequenced and analyzed, should be emailed to Dr Juncai Ma at
39
40 ma@im.ac.cn. Once a proposal is granted as a subproject, the scientist(s) will be
41
42 asked to lead the full genome analysis and jointly publish the generated outcomes.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57

58 **Conclusion**

59
60
61
62
63
64
65

1 As a collaborative network of international culture collections, GCM 2.0 will
2
3 contribute to a genome-based microbial taxonomic framework, establishing high
4
5 quality complete genome sequences as the new gold standard. The resulting
6
7 knowledge and tools generated through this project will not only directly facilitate the
8
9 identification of microorganisms, but will also improve our ability to predict new
10
11 gene complexes and their functions from microbial communities. Thus, our
12
13 knowledge of the hitherto undiscovered microbial diversity will be expanded, which
14
15 may lead to the sustainable utilization of microbial resources for human benefit.
16
17
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.
29
30 Proc Natl Acad Sci USA. 2002; 99:10494–9. doi:10.1073/pnas.142680199.
31
32
- 33
34 2. Garrity GM. A New Genomics-Driven Taxonomy of Bacteria and Archaea: Are
35
36 We There Yet? J Clin Microbiol. 2016; 54: 1956-63. doi: 10.1128/JCM.00200-16.
37
38
- 39
40 3. Hawksworth DL. The magnitude of fungal diversity: the 1.5 million species
41
42 estimate revisited. Mycol Res. 2001; 105:1422–1432.
43
44
- 45
46 4. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al.
47
48 Proposed minimal standards for the use of genome data for the taxonomy of
49
50 prokaryotes. Int J Syst Evol Microbiol. 2018, 68:461–466. doi:
51
52 10.1099/ijsem.0.002516.
53
54
- 55
56 5. Kim Y, Koh I, Rho M. Deciphering the human microbiome using next-generation
57
58 sequencing data and bioinformatics approaches. Methods. 2015; 79-80: 52-9. doi:
59
60
61
62
63
64
65

1 10.1016/j.ymeth.2014.10.022.
2

3 6. Kang HS. Phylogeny-guided (meta)genome mining approach for the targeted
4 discovery of new microbial natural products. *J Ind Microbiol Biotechnol.* 2017;
5
6 44:285-293. doi: 10.1007/s10295-016-1874-z.
7
8

9 7. Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP,
10
11 Göker M, et al. 1,003 reference genomes of bacterial and archaeal isolates expand
12 coverage of the tree of life. *Nat Biotechnol.* 2017; 35:676–683. doi: 10.1038/nbt.3886.
13
14
15
16
17
18

19 8. Whitman WB, Woyke T, Klenk HP, Zhou Y, Lilburn TG, Beck BJ, et al. Genomic
20 Encyclopedia of Bacterial and Archaeal Type Strains, Phase III: the genomes of soil
21 and plant-associated and newly described type strains. *Stand Genomic Sci.* 2015;
22
23 10:26. doi: 10.1186/s40793-015-0017-x.
24
25
26
27
28
29

30 9. 1000 fungal genomes Project.
31

32 <https://jgi.doe.gov/our-science/science-programs/fungal-genomics/1000-fungal-genomes/>. Accessed March 1, 2017.
33
34
35
36
37
38

39 10. Wu L, Sun Q, Desmeth P, Sugawara H, Xu Z, McCluskey K, et al. World data
40 centre for microorganisms: an information infrastructure to explore and utilize
41 preserved microbial strains worldwide. *Nucleic Acids Res.* 2017; 45: D611-D618. doi:
42
43 10.1093/nar/gkw903.
44
45
46
47
48
49
50
51
52

53 **Abbreviations**

54
55 GEBA, Genomic Encyclopedia of Bacteria and Archaea; GCM, Global Catalogue of
56
57
58
59 Microorganisms; IJSEM, International Journal of Systematic and Evolutionary
60
61
62
63
64
65

1 Microbiology; US DOE JGI, US Department of Energy Joint Genome Institute;
2
3 WDCM, World Data Centre for Microorganisms; WFCC, World Federation of Culture
4
5
6 Collections.
7
8
9

10 11 **Acknowledgements**

12
13
14 The authors thank the WFCC Executive Board for its support, and also acknowledge
15
16 the members, partners, and advisors of the GCM type strain sequencing project who
17
18 have made this collaboration possible. We thank Takashi Gojobori from King
19
20 Abdullah University of Science and Technology, Guoping Zhao from Shanghai
21
22 Institutes for Biological Sciences, Chinese Academy of Sciences, Xuewei Xu from
23
24 State Oceanic Administration, China, Xiaoyang Zhi from Yunnan University, China
25
26 and Hua Xiang, and Lei Cai from the Institute of Microbiology, Chinese Academy of
27
28 Sciences for their insightful suggestions.
29
30
31
32
33
34
35
36
37
38

39 **Funding**

40
41
42 This work was supported by the Strategic Priority Research Program of the Chinese
43
44 Academy of Sciences (grant number XDA19050301), the Bureau of International
45
46 Cooperation of the Chinese Academy of Sciences (grant numbers
47
48 153211KYSB20160029 and 153211KYSB20150010), the National Key Research
49
50 Program of China (grant numbers 2017YFC1201202, 2016YFC1201303, and
51
52 2016YFC0901702), the 13th Five-year Informatization Plan of the Chinese Academy
53
54 of Sciences (grant number XXH13506), and the National Science Foundation for
55
56
57
58
59
60
61
62
63
64
65

1 Young Scientists of China (grant number 31701157).
2
3
4
5

6 **Availability of data and materials**
7

8
9 Not applicable.
10
11
12
13

14 **Conflict of interest**
15

16
17 The authors declare that they have no competing interests.
18
19
20
21

22 **Consent for publication**
23

24
25 Not applicable.
26
27
28
29
30

31 **Ethics approval and consent to participate**
32

33
34 Not applicable.
35
36
37
38

39 **Availability of data and material**
40

41
42 Not applicable.
43
44
45
46

47 **Author contributions**
48

49
50 LHW drafted the original manuscript text with detailed input from other authors. All
51
52 authors participated in the GCM type strain sequencing project and have read and
53
54 approved the final manuscript.
55
56
57
58
59
60
61
62
63
64
65