MIBiG 4.0: Advancing Biosynthetic Gene Cluster Curation through Global Collaboration

Mitja M. Zdouc*[1], Kai Blin*†[2], Nico L.L. Louwen[1], Jorge Navarro[1], Catarina Loureiro[1], Chantal D. Bader[3,4], Constance B. Bailey[5], Lena Barra[6], Thomas J. Booth[2], Kenan A. J. Bozhüyük[7,8], José D. D. Cediel-Becerra[9], Zachary Charlop-Powers[10], Marc G. Chevrette[9,11], Yit Heng Chooi[12], Paul M. D'Agostino[13,14], Tristan de Rond[15], Elena Del Pup[1], Katherine R. Duncan[16], Wenjia Gu[17], Novriyandi Hanif[18], Eric J. N. Helfrich[19,20,21], Matthew Jenner[22,23], Yohei Katsuyama[24,25], Aleksandra Korenskaia[26], Daniel Krug[27,28,29], Vincent Libis[30,31], George A. Lund[32], Shrikant Mantri[33,34], Kalindi D. Morgan[35], Charlotte Owen[36], Chin-Soon Phan[37], Benjamin Philmus[38], Zachary L. Reitz[39], Serina L. Robinson[40], Kumar Saurabh Singh[1,41,42], Robin Teufel[43], Yaojun Tong[44], Fidele Tugizimana[45,46], Dana Ulanova[47,48], Jaclyn M. Winter[49], César Aguilar[50,51], Daniel Y. Akiyama[52], Suhad A. A. Al-Salihi[53], Mohammad Alanjary[1], Fabrizio Alberti[54], Gajender Aleti[55], Shumukh A. Alharthi[56], Mariela Y. Arias Rojo[57], Amr A. Arishi[12,58], Hannah E. Augustijn[1,59], Nicole E. Avalon[60], J. Abraham Avelar-Rivas[61], Kyle K. Axt[38], Hellen B. Barbieri[52], Julio Cesar J. Barbosa[52], Lucas Gabriel Barboza Segato[62], Susanna E. Barrett[63,64], Martin Baunach[65], Christine Beemelmanns[3,28,], Dardan Beqaj[66], Tim Berger[67], Jordan Bernaldo-Agüero[68,69], Sandra M. Bettenbühl[70,71], Vincent A. Bielinski[72], Friederike Biermann[1,19,20], Ricardo M. Borges[73], Rainer Borriss[74,75], Milena Breitenbach[19,76], Kevin M. Bretscher[77,78,79], Michael Brigham[80], Larissa Buedenbender[81], Brodie W. Bulcock[12], Carolina CP. Cano-Prieto[2], João Capela[82], Victor J. Carrion[83,78,77], Riley S. Carter[63,84], Raquel Castelo-Branco[85], Gabriel Castro-Falcón[86], Fernanda O. Chagas[73], Esteban Charria-Girón[87,88], Ayesha Ahmed Chaudhri[19], Vasvi Chaudhry[89], Hyukjae Choi[90,91], Yukyung Choi[92], Roya Choupannejad[93], Jakub Chromy[94], Melinda S. Chue Donahey[9], Jerome Collemare[95], Jack A. Connolly[94], Kaitlin E. Creamer[96,97,] Max Crüsemann[65], Andres Arredondo Cruz[57], Andres Cumsille[9], Jean-Felix Dallery[98], Luis Caleb Damas-Ramos[2], Tito Damiani[99], Martinus de Kruijff[3], Belén Delgado Martín[83], Gerardo Della Sala[100], Jelle Dillen[101], Drew T. Doering[102], Shravan R. Dommaraju[63,84], Suhan Durusu[70,71], Susan Egbert[62], Mark Ellerhorst[103], Baptiste Faussurier[30,31], Artem Fetter[16], Marc Feuermann[104], David P. Fewer[105], Jonathan F. Foldi[94], Andri Frediansyah[106,107], Erin A. Garza[108], Athina Gavriilidou[26], Andrea Gentile[109,110], Jennifer Gerke[111], Hans Gerstmans[112,113,114], Juan Pablo Gomez-Escribano[115], Luz A. Gonzalez[116], Natalie E. Grayson[60], Claudio Greco[56], Juan E. Gris Gomez[116], Sebastian Guerra[117,9], Shaday Guerrero Flores[118], Alexey Gurevich[119,120], Karina Gutiérrez-García[121], Lauren Hart[122,123,] Kristina Haslinger[124], Beibei He[125], Teo Hebra[99], Jethro L. Hemmann[126], Hindra Hindra[127,128], Lars Höing[43], Darren C. Holland[12], Jonathan E. Holme[129], Therese Horch[2], Pavlo Hrab[1,130], Jie Hu[131], Thanh-Hau Huynh[132], Ji-Yeon Hwang[133], Riccardo Iacovelli[134,] Dumitrita Iftime[66], Marianna Iorio[109], Sidharth Jayachandran[2], Eunah Jeong[92], Jiayi J. Jing[1], Jung J. Jung[132], Yuya Kakumu[70,71], Edward Kalkreuter[135], Kyo Bin Kang[92], Sangwook Kang[132], Wonyong Kim[136], Geum Jin Kim[137,91], Hyunwoo Kim[138], Hyun Uk Kim[139], Martin Klapper[140], Robert A. Koetsier[1], Cassandra Kollten[14], Ákos T. Kovács[59,141], Yelyzaveta Kriukova[13,14], Noel Kubach[26], Aditya M. Kunjapur[142], Aleksandra K. Kushnareva[119,120], Andreja Kust[96,97], Jessica Lamber[4,29,143] Martin

Larralde[144], Niels J. Larsen[2], Adrien P. Launay[145], Ngoc-Thao-Hien Le[30], Sarah Lebeer[146,147], Byung Tae Lee[139,] Kyungha Lee[92], Katherine L. Lev[148], Shu-Ming Li[149], Yong-Xin Li[125], Cuauhtémoc Licona-Cassani[50,150], Annette Lien[1], Jing Liu[125], Julius Adam V. Lopez[151,152], Nataliia V. Machushynets[59], Marla I. Macias[153], Taifo Mahmud[38], Matiss Maleckis[154], Maharai Martinez[116], Yvonne Mast[115,155], Marina F. Maximo[52], Christina M. McBride[156], Rose M. McLellan[157,158], Khyati Mehta Bhatt[93], Chrats Melkonian[1,159], Aske Merrild[160], Mikko Metsä-Ketelä[161], Douglas A. Mitchell[63], Alison V. Müller[4,29,143], Giang-Son Nguyen[129], Hera T. Nguyen[12], Timo H. J. Niedermeyer[162], Julia H. O'Hare[73], Adam Ossowicki[78,163], Bohdan O. Ostash[164,165], Hiroshi Otani[102,166], Leo Padva[65], Sunaina Paliyal[33], Xinya Pan[77,78,79], Mohit Panghal[33,34], Dana S. Parade[13], Jiyoon Park[132], Jonathan Parra[167,168,169], Marcos Pedraza Rubio[78,163], Huong Thi Pham[92], Sacha J. Pidot[170], Jörn Piel[171], Bita Pourmohsenin[26], Malik Rakhmanov[43], Sangeetha Ramesh[172], Michelle H. Rasmussen[173], Adriana Rego[85], Raphael Reher[174], Andrew J. Rice[63], Augustin Rigolet[59], Adriana Romero-Otero[2,154], Luis Rodrigo Rosas-Becerra[59,61], Pablo Y. Rosiles[145], Adriano Rutz[175], Byeol Ryu[60], Libby Ann Sahadeo[62], Murrel Saldanha[9], Luca Salvi[66], Eduardo Sánchez-Carvajal[176], Christian Santos-Medellin[177], Nicolau Sbaraini[12], Sydney M. Schoellhorn[178], Clemens Schumm[4,29,143], Ludek Sehnal[179,180], Nelly Selem[118], Anjali D. Shah[181,182], Tania K. Shishido[183,93], Simon Sieber[184], Velina Silviani[91,185], Garima Singh[186,187], Hemant Singh[33], Nika Sokolova[124], Eva C. Sonnenschein[188], Margherita Sosio[109], Sven T. Sowa[43], Karin Steffen[189], Evi Stegmann[66,190], Alena B. Streiff[171], Alena Strüder[66], Frank Surup[87], Tiziana Svenningsen[160], Douglas A. Sweeney[60], Judit Szenei[2], Azat Tagirdzhanov[119,120], Bin Tan[70,71], Matthew J. Tarnowski[188], Barbara R. Terlouw[1], Rey Thomas[191,192], Nicola U. Thome[193,194], Laura Rosina Torres Ortega[1], Thomas Tørring[160], Marla Trindade[195], Andrew Truman[196], Marie Tvilum[160], Daniel W. Udwary[102], Christoph Ulbricht[65], Lisa Vader[2], Gilles P. van Wezel[79,197], Max Walmsley[94], Randika Warnasinghe[62], Heiner G. Weddeling[43], Angus N. M. Weir[198,112], Katherine Williams[199,200], Sam E. Williams[2], Thomas E. Witte[201], Steffaney M. Wood[60,202], Keith Yamada[161], Dong Yang[203], Dongsoo Yang[204], Jingwei Yu[205], Zhenyi Zhou[206], Nadine Ziemert[26], Lukas Zimmer[65], Alina Zimmermann[115], Christian Zimmermann[207], Justin J.J. van der Hooft[1,208,] Roger G. Linington[209], Tilmann Weber[2†], Marnix H. Medema[1†]

1. Bioinformatics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.
2. The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark.
3. Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), 66123 Saarbrücken, Germany.
4. Helmholtz Centre for Infection Research (HZI), 38124 Braunschweig, Germany.
5. School of Chemistry, University of Sydney, Sydney, New South Wales, Australia.
6. Department of Chemistry, University of Konstanz, 78464 Konstanz, Germany .
7. Synthetic Biology of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), 66123 Saarbrücken, Germany.
8. Myria Biosciences AG, Tech Park Basel, Hochbergstrasse 60C, 4057 Basel, Switzerland.
9. Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida, 32611, USA.
10. Ginkgo Bioworks, 27 Drydock Avenue, 8th Floor, Boston, MA 02210.
11. University of Florida Genetics Institute, University of Florida, Gainesville, FL, USA 32611.
12. School of Molecular Sciences, University of Western Australia, Perth 6009, Australia.
13. Department of Natural Product Biotechnology, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) and Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123 Saarbrücken, Germany.
14. Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany.
15. School of Chemical Sciences, University of Auckland, New Zealand.

16. Newcastle University, Biosciences Institute, Newcastle upon Tyne, UK. NE2 4HH.

17. Sutro Biopharma, 111 Oyster Point Blvd, South San Francisco, CA, 94080.

18. Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia.

19. Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue Strasse 9, 60438 Frankfurt am Main, Germany.

20. LOEWE Center for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

21. Senckenberg Society for Nature Research, Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

22. Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK.

23. Warwick Integrative Synthetic Biology Centre (WISB), University of Warwick, Coventry, CV4 7AL, UK.

24. Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo.

25. Collaborative Research Institute for Innovative Microbiology, The University of Tokyo.

26. Translational Genome Mining for Natural Products, Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT), Interfaculty Institute for Biomedical Informatics (IBMI), University of Tübingen, Tübingen, Germany.

27. Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) and Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123 Saarbrücken, Germany.

28. Universität des Saarlandes, Campus E8.1, 66123 Saarbrücken, Germany.

29. German Centre for Infection Research (DZIF), 38124 Hannover-Braunschweig, Germany.

30. Université Paris Cité - Inserm Unit 1284, Paris, France.

31. Generare Bioscience, Paris, France.

32. Sustainable Soils and Crops, Rothamsted Research, Harpenden, United Kingdom.

33. National Agri Food Biotechnology Institute, Mohali -140306 , Punjab ,India.

34. Regional Center For Biotechnology, Faridabad - 121001 Haryana , India.

35. Department of Chemistry and Biochemistry, University of Northern British Columbia, Prince George, BC, Canada.

36. Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK.

37. Latvian Institute of Organic Synthesis, Aizkraukles street 21, LV-1006 Riga, Latvia.

38. Department of Pharmaceutical Sciences, Oregon State University, Corvallis, OR 97331-3507, USA.

39. Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, CA 93117.

40. Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, 8600 Duebendorf, Switzerland.

41. Plant-Microbe Interactions, Institute of Environmental Biology, Utrecht University, The Netherlands.

42. Faculty of Environment, Science and Economy, University of Exeter, TR10 9FE Penryn Cornwall UK.

43. Department of Pharmaceutical Sciences, University of Basel, 4056 Basel, Switzerland.

44. State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China.

45. Department of Biochemistry, University of Johannesburg, South Africa.

46. International Research and Development (R&D) Division, Omnia Nutriology, Omnia Holdings, Ltd, South Africa.

47. Department of Marine Resource Science, Faculty of Agriculture and Marine Science, Kochi University, Nankoku, Kochi, Japan.

48. Marine Core Research Institute, Kochi University, Nankoku, Kochi, Japan.

49. Department of Pharmacology and Toxicology, University of Utah, Utah, 84112, United States.

50. Industrial Genomics Laboratory, Centro de Biotecnología FEMSA, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, N.L. México.

51. Department of Chemistry, Purdue University, West Lafayette, IN, 47907, USA.

52. Institute of Chemistry, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

53. Department of Applied Sciences, University of Technology-Iraq.

54. School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK.

55. Department of Food and Animal Sciences, College of Agriculture, Tennessee State University, Nashville, TN 37209, USA.

56. Department of Biosciences, Swansea University, Swansea, SA2 8PP, UK.

57. Escuela Nacional de Estudios Superiores Unidad Leon, Universidad Nacional Autonoma de Mexico.

58. Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia.

59. Institute of Biology, Leiden University, 2333BE Leiden, The Netherlands.

60. Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0212, USA.

61. Unidad de Genómica Avanzada, Cinvestav, 36620 Irapuato, Mexico.

62. Department of Chemistry, University of Manitoba, Winnipeg, MB, Canada.

63. Department of Chemistry, University of Illinois Urbana-Champaign, Urbana, IL, USA.

64. Carl R. Woese Institute for Genomic Biology at the University of Illinois Urbana-Champaign, Urbana, IL, USA.

65. Institute of Pharmaceutical Biology, University of Bonn, Nussallee 6, 53121 Bonn, Germany.

66. Interfaculty Institute of Microbiology and Infection Medicine Tübingen, Microbial Bioactive Compounds, University of Tübingen, Auf der Morgenstelle 28, 72076, Tübingen, Germany.

67. Institute for Pharmaceutical Biology and Biotechnology, Department of Pharmacy, Philipps-University Marburg, Robert-Koch-Straße 4, 35037 Marburg, Germany.

68. Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, México.

69. Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Lima, Perú.

70. Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue Strasse 9, 60438 Frankfurt am Main, Germany.

71. LOEWE Center for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

72. Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 21941-909, Brazil.

73. Instituto de Pesquisas de Produtos Naturais Walter Mors, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

74. Institute of Biology, Humboldt University Berlin, Berlin, Germany.

75. Institute of Marine Biotechnology (IMaB), University of Greifswald, Greifswald, Germany.

76. LOEWE Center for Translational Biodiversity Genomics (TBG), Frankfurt, Germany.

77. Institute of Biology, Leiden University, Sylviusweg 72, 2333BE, Leiden, The Netherlands.

78. Departament of Crop Protection, Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Campus Universitario de Teatinos, 29010, Málaga, Spain.

79. Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10 6708 PB Wageningen, The Netherlands.

80. School of Cellular and Molecular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, United Kingdom.

81. CICA – Centro Interdisciplinar de Química e Bioloxía, Universidade da Coruña, 15071 A Coruña, Spain.

82. Centre of Biological Engineering, University of Minho, 4710-057, Braga, Portugal.

83. Department of Microbiology, Faculty of Science, Campus Universitario de Teatinos s/n, University of Málaga, 29010, Málaga, Spain.

84. Carl R. Woese Institute for Genomic Biology, Urbana, IL, USA.

85. Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, 4450-208 Matosinhos, Portugal.

86. Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA.

87. Department Microbial Drugs, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany.

88. Institute of Microbiology, Technische Universität Braunschweig, Spielmannstraße 7, 38106 Braunschweig, Germany.

89. Microbial Interactions in Plant Ecosystems, IMIT/ZMBP, Eberhard Karls University of Tübingen, 72076 Tübingen, Germany.

90. College of Pharmacy, Yeungnam University, 280 Daehak-ro Gyeongsan-Si, Gyeongsangbuk-do 38541, Republic of Korea.

91. Research Institute of Cell Culture , Yeungnam University, 280 Daehak-ro Gyeongsan-Si, Gyeongsangbuk-do 38541, Republic of Korea.

92. College of Pharmacy, Sookmyung Women's University, Seoul 04310, Korea.

93. Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

94. Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, Manchester M1 7DN, United Kingdom.

95. Westerdijk Fungal Biodiversity Institute, Upsalalaan 8, 3584CT Utrecht, Netherlands.

96. Innovative Genomics Institute, University of California, Berkeley, CA, USA.

97. Department of Earth and Planetary Science, University of California, Berkeley, Berkeley, CA, USA.

98. Université Paris-Saclay, INRAE, UR BIOGER, 91120 Palaiseau, France.

99. IOCB Prague, Czech Academy of Science, Flemingovo náměstí 542/2, 160 00 Praha 6, Czech Republic.

100. Department of Ecosustainable Marine Biotechnology, Stazione Zoologica Anton Dohrn, Giardini del Molosiglio, Via A.F. Acton 55, 80133 Naples, Italy.

101. Department of Bioscience Engineering, Research Group Laboratory of Applied Microbiology and Biotechnology, University of Antwerp, Antwerp, Belgium.

102. US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

103. Biochemistry of Microorganisms, Faculty of Life Sciences: Food, Nutrition, and Health, University of Bayreuth, Germany.

104. SIB Swiss Institute of Bioinformatics 1, rue Michel Servet - CH 1211 - Switzerland.

105. Department of Microbiology, University of Helsinki, Viikinkaari 9, 00017, Helsinki, Finland.

106. Research Center for Food Technology and Processing, Research Organization of Agriculture and Food, National Research and Innovation Agency (BRIN), Jl. Jogja - Wonosari Km 31.5, DI. Yogyakarta 55861, Indonesia.

107. Microbial Metabolites for Food Research Group, National Research and Innovation Agency (BRIN), Jl. Jogja - Wonosari Km 31.5, DI. Yogyakarta 55861, Indonesia.

108. J. Craig Venter Institute, La Jolla, CA, 92037, USA.

109. NAICONS Srl, 20139 Milan, Italy.

110. University of Parma, 43121 Parma, Italy.

111. Leibniz Universität Hannover, Institute for Organic Chemistry, Schneiderberg 38, 30167 Hannover, Germany.

112. VIB-KU Leuven Center for Microbiology, Flanders Institute for Biotechnology, 3001 Leuven, Belgium.

113. Laboratory for Biomolecular Discovery & Engineering, KU Leuven, 3001 Leuven, Belgium.

114. Biosensors Group, KU Leuven, 3001 Leuven, Belgium.

115. Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Inhoffenstraße 7B, 38124 Braunschweig, Germany.

116. Industrial Genomics Laboratory, Centro de Biotecnología FEMSA, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, N.L. México.

117. University of Florida Genetics Institute, University of Florida, Gainesville, Florida, 32610, USA.

118. Centro de Ciencias Matemáticas National Autonomous University of Mexico (UNAM) Antigua Carretera a Pátzcuaro # 8701, Sin Nombre, Residencial San José de la Huerta, 58089 Morelia, Mich.

119. Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarbrücken 66123, Germany.

120. Center for Bioinformatics Saar and Saarland University, Saarland Informatics Campus, Saarbrücken 66123, Germany.

121. Biosphere Sciences and Engineering Division, Carnegie Institution for Science; Baltimore, 21218, USA.

122. Program in Chemical Biology, University of Michigan, Ann Arbor, Michigan, USA. .

123. Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA.

124. Department of Chemical and Pharmaceutical Biology, University of Groningen, Antonius Deusinglaan 1, 9713AV Groningen, The Netherlands.

125. Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China.

126. Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology (HKI), Jena, Germany.

127. Institute of Infectious Disease Research, McMaster University, Hamilton, Ontario, Canada.

128. Department of Biology, McMaster University, Hamilton, Ontario, Canada.

129. Department of Biotechnology and Nanomedicine, SINTEF Industry, P.O.Box 4760 Torgard, N-7465 Trondheim, Norway.

130. Microbiology Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.

131. Corteva Agriscience, Farming Solutions and Digital, Indianapolis, Indiana, United States.

132. Natural Products Research Institute, College of Pharmacy, Seoul National University, Seoul 08826, Korea.

133. Molecular Targets Program, Center for Cancer Research, National Cancer Institute, Frederick, MD, United States.

134. VTT Technical Research Centre of Finland Ltd, 02150 Espoo, Finland.

135. Department of Chemistry and Biochemistry, Laboratories of Molecular Recognition, Florida State University, 95 Chieftan Way, Tallahassee, FL 32306, USA.

136. Department of Applied Biology, College of Agriculture and Life Sciences, Chonnam National University, Gwangju, 61186, South Korea.

137. Department of Pharmacology, College of Medicine, Dongguk University, Gyeongju, Gyeongsangbuk-do, Republic of Korea.

138. College of Pharmacy and Integrated Research Institute for Drug Development, Dongguk University-Seoul, Goyang 10326, Republic of Korea.

139. Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea.

140. Department of Paleobiotechnology, Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute.

141. DTU Bioengineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

142. Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA.

143. Helmholtz Institute for Pharmaceutical Research Saarland (HIPS) and Saarland University Department of Pharmacy, 66123 Saarbrücken, Germany.

144. Center of Infectious Diseases, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden,9 Netherlands.

145. Bioinformatics department, Endogenomiks, El Marqués, Querétaro, Mexico.

146. Department of Bioscience Engineering, University of Antwerp, 2020 Antwerp, Belgium.

147. U-MaMi Centre of Excellence, 2020 Antwerp, Belgium.

148. Program in Chemical Biology, University of Michigan, Ann Arbor MI, USA.

149. Institut für Pharmazeutische Biologie und Biotechnologie, Fachbereich Pharmazie, Philipps-Universität Marburg, Robert-Koch Straße 4, 35037 Marburg, Germany.

150. Integrative Biology Research Unit, The Institute for Obesity Research, Tecnológico de Monterrey, N.L. México.

151. School of Innovation and Sustainability, De La Salle University, Laguna Boulevard, LTI Spine Road, Brgys. Biñan and Malamig, 4024 Biñan City, Laguna, Philippines.

152. Department of Chemistry, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines.

153. University of Strathclyde, Strathclyde Instutute of Pharmacy and Biomedical Sciences, Glasgow. UK. G4 0RE.

154. Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark, Søltofts Plads, Building 221, 2800 Kgs. Lyngby, Denmark.

155. Technical University of Braunschweig, Institute of Microbiology, Rebenring 56, 38106 Braunschweig, Germany.

156. Program in Chemical Biology, University of Michigan, Ann Arbor, MI, USA.

157. Ferrier Research Institute, Victoria University of Wellington, Wellington 6012, New Zealand.

158. Maurice Wilkins Centre for Molecular Biodiscovery, Victoria University of Wellington, Wellington 6012, New Zealand.

159. Theoretical Biology and Bioinformatics, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands.

160. Department of Biological and Chemical Engineering, Aarhus University, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark.

161. Department of Life Technologies, University of Turku, Tykistökatu 6, FIN-20520 Turku, Finland.

162. Institute of Pharmacy, Freie Universität Berlin, Königin-Luise-Str. 2+4, 14195 Berlin, Germany.

163. Departament of Microbiology, Faculty of Science, Campus Universitario de Teatinos s/n, University of Málaga, 29010, Málaga, Spain.

164. Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Hrushevskoho st 4, Lviv 79005, Ukraine.

165. German-Ukrainian Core of Excellence in Natural Products Research, Zelena str. 20, Lviv 79005, Ukraine.

166. Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

167. Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San José, 11501-2060, Costa Rica.

168. Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, 1174-1200, San José, Costa Rica.

169. Instituto de Investigaciones Farmacéuticas (INIFAR), Facultad de Farmacia, Universidad de Costa Rica, San José, 11501-2060, Costa Rica.

170. Department of Microbiology and Immunology at the Doherty Institute, Unviersity of Melbourne, Victoria, Australia, 3000.

171. Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog Weg 4, 8093 Zurich, Switzerland.

172. Department of Plant Pathology, University of California Davis, Davis, CA, USA.

173. Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000, Aarhus C, Denmark.

174. Institute of Pharmaceutical Biology and Biotechnology,Philipps-University Marburg, Marburg, Germany.

175. Institute for Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland.

176. Laboratorio de Microbiología Molecular y Biotecnología Ambiental, Centro de Biotecnología DAL, Universidad Técnica Federico Santa María, Valparaíso, Chile.

177. Corteva Agriscience, Indianapolis, Indiana 46268, United States.

178. Department of Chemistry and BioDiscovery Institute, University of North Texas, 1155 Union Circle, Denton, Texas 76203, United States.

179. Interfaculty Institute of Microbiology and Infection Medicine Tübingen, Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Tübingen, Germany.

180. Masaryk University, Faculty of Science, RECETOX, Kamenice 753/5, 625 00 Brno, Czech Republic.

181. School of Biological Sciences, Life Sciences Building, University of Bristol, 24 Tyndall Ave, Bristol BS8 1TQ, UK.

182. School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

183. Institute of Biotechnology, Helsinki Institute of Life Science, University of Helsinki, Viikinkaari 5, 00790, Helsinki, Finland.

184. Department of Chemistry, University of Zurich, 8057 Zurich, Switzerland.

185. Yeungnam University, 280 Daehak-ro Gyeongsan-Si, Gyeongsangbuk-do 38541, Republic of Korea.

186. Department of Biology, University of Padova, Via U. Bassi, 58/B, 35121 Padova, Italy.

187. Botanical Garden, University of Padova, Padua, Italy.

188. Department of Biosciences, Geography and Physics, Swansea University, SA2 8PP, UK.

189. Department of Biological Sciences and Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN, USA.

190. German Centre for Infection Research (DZIF), Partner Site Tübingen, Tübingen, Germany.

191. DE SANGOSSE, Bonnel, 47480, Pont-Du-Casse, France.

192. Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, Université Toulouse III, Toulouse INP, 24 Chemin de Borde Rouge, Auzeville, 31320, Auzeville-Tolosane, France.

193. Department of Microbial Biotechnology, Institute of Biology, Leiden University, Sylviusweg 72, 2333BE Leiden, The Netherlands.

194. Laboratory for Biomolecular Discovery & Engineering, VIB-KU Leuven Center for Microbiology, KU Leuven, 3001 Leuven, Belgium.

195. Institute for Microbial Biotechnology and Metagenomics, University of the Western Cape, Bellville 7535, South Africa.

196. Department of Molecular Microbiology, John Innes Centre, Norwich, United Kingdom.

197. Institute of Biology, Leiden University, 2333BE Leiden, The Netherlands.

198. The Rosalind Franklin Institute, Harwell Science and Innovation Campus, Didcot, Oxfordshire, OX11 0QX, UK.

199. University of the West of England, School of Applied Sciences, Bristol, BS16 1QY, UK.

200. University of Bristol, School of Biological Sciences, Bristol, BS8 1TQ, UK.

201. Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario, Canada.

202. Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, CA 92037.

203. Natural Products Discovery Center, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, University of Florida, Jupiter, Florida 33458, United States.

204. Synthetic Biology and Enzyme Engineering Laboratory, Department of Chemical and Biological Engineering, Korea University, Seoul 02841, Republic of Korea.

205. Institute of Plant and Food Science, Department of Biology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China.

206. College of Pharmaceutical Science & Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou 310014, China.

207. Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, 1060 Vienna, Austria.

208. Department of Biochemistry, University of Johannesburg, 2006 Johannesburg, South Africa.
209. Department of Chemistry, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada.

* Co-first authors   † Co-corresponding authors
MHM: marnix.medema@wur.nl TW: tiwe@biosustain.dtu.dk KB: kblin@biosustain.dtu.dk

ABSTRACT

Natural products, small molecules of biological origin, often show specific and potent biological activities. These specialized or secondary metabolites have important applications in agriculture, engineering, and medicine. Usually, the biosynthesis of natural products is governed by sets of co-regulated and physically clustered genes known as biosynthetic gene clusters (BGCs). To share information about BGCs in a standardized and machine-readable way, the Minimum Information about a Biosynthetic Gene cluster (MIBiG) data standard and repository was initiated in 2015. Since its conception, MIBiG has been regularly updated to expand data coverage and remain up to date with innovations in natural product research. Here, we describe MIBiG version 4.0, an extensive update to the data repository and the underlying data standard. In a massive community annotation effort, 267 contributors performed 8304 edits, creating 557 new entries and modifying 590 existing entries, resulting in a new total of 3059 curated entries in MIBiG. Particular attention was paid to ensuring high data quality, with automated data validation using a newly developed custom submission portal paired with a novel peer-reviewing model. MIBiG 4.0 also takes steps towards a rolling release model and a broader involvement of the scientific community. MIBiG 4.0 is accessible online at https://dev.mibig.secondarymetabolites.org/.

INTRODUCTION

Many organisms are prolific producers of small molecules commonly referred to as natural products (NPs). Also known as specialized or secondary metabolites, these molecules often show a diversity of potent biological activities, which have been leveraged for the development of numerous drugs (1, 2). NPs are generally hypothesized to increase the fitness of the producing organism or its host. In microbes, the biosynthetic genes required for the production of a NP are co-regulated and frequently physically clustered in the genome, in a so-called biosynthetic gene cluster (BGC), and often transferred horizontally (3). BGCs, which by definition consist of two or more genes, encode the proteins/enzymes used in natural product biosynthesis of NPs and are the object of "genome mining" strategies that leverage analysis of genome sequence data for the discovery of (novel) metabolites (4). In rare cases, a single gene may be responsible for the biosynthesis of a natural product, such as a large NRPS; these standalone genes are also entered into MIBiG due to their relevance to specialised metabolism.

Over the last decades, various methods using manually curated detection rules based on prior knowledge (5–7), and more recently, machine learning-based tools for genome mining have been developed (8–12). These tools rely on accurately curated and machine-readable experimental data for annotation, rule definition, and training purposes. Unfortunately, machine-readable data is not readily available from the scientific literature nor is it universally required by journals to be directly deposited in databases. While there are efforts to mine data from the literature using computational methods (13, 14), these approaches currently often come with limitations when compared to human curators and may not be compatible with copyright laws. Therefore, manual data curation performed by researchers remains the gold standard for the generation of machine-readable data.

The largest manually curated resource on NP BGCs is the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) data repository (15). Initiated in 2015 and based on the MIBiG Data Standard, it now holds over 2500 hand-curated entries of experimentally validated BGCs and their products, alongside additional information such as biological activities and gene annotations. Conceptualized as an open data repository curated by and for the NP community, it has seen three iterations of online community-driven data annotation and curation hackathons (also known as "annotathons"), with more than 250 participants from 33 countries (16, 17). Despite its size, the MIBiG repository still only describes a part of the continuously growing known biosynthetic space, which motivates further efforts in curating and systemizing information on BGCs.

Here, we present version 4.0 of the MIBiG data standard and repository. Besides a thorough update of the underlying MIBiG data standard, we have substantially grown the number of available entries by initiating a large-scale community curation effort. In the first half of 2024, 267 contributors created 557 new entries and modified 590 existing entries in the scope of eight community annotathons (six general open events and two final data curation sessions with a more dedicated team). In this version of MIBiG, we focused on maintaining and further improving data quality in terms of completeness and accuracy. We encouraged contributors to fully complete entries before submission, which has significantly decreased the number of so-called minimum entries (entries with only the minimally required information) in the database. We also introduced a new peer-review model where modifications to entries are examined and approved by one or more expert reviewers, who can request corrections. Additionally, we have established a framework for efficient and standardized data submission, introducing a web interface (MIBiG Submission Portal) that allows for parallel, distributed data input featuring automated input validation. The latter refers to the tests that are performed by the submission portal itself to ensure the correct data types and formats are filled in. Together, these efforts further consolidate MIBiG as the leading database on experimentally characterized BGCs and prepare for the transition to a dynamic, rolling-release curation model.

METHODS AND IMPLEMENTATION

Rework of the MIBiG Data Standard

The MIBiG Data Standard (from here: Data Standard) is the "blueprint" of all allowed data in the MIBiG repository. It defines mandatory and optional data fields, allows the use of controlled vocabularies and automated validation, and enables the organization of complex data in a consistent, human- and machine-readable way. In this update, we extensively updated the Data Standard to accommodate advances in NP research and to extend the scope and ease of (re-)use of covered (meta)data.

Literature references and evidence qualifiers. Previously, all literature references associated with a MIBiG entry were collected in a single block, making it difficult to locate the origin of specific experimental data. In this update, we reorganized the Data Standard such that each data category (e.g. biosynthetic information, compound details) has its own list of literature references. Furthermore, evidence qualifiers can be selected from a controlled vocabulary (e.g.

"Heterologous expression") that concisely summarizes the experimental support for the claims.

Biosynthesis information, multiple loci, and class updates. Biosynthetic information is now organized in a "biosynthesis" section, tracking biosynthetic types, modules, operons, and newly introduced "biosynthetic path", which allows contributors to describe cases where a single BGC can lead to multiple products or describe sub-clusters of genes that produce building blocks. The "multiple loci" system has been re-introduced, allowing the specification of satellite genes or gene clusters that are involved in the biosynthesis but are not clustered with the 'main' BGC. Nevertheless, we still require that multiple biosynthetic genes are clustered in the same genomic region, to exclude non-clustered pathways. Furthermore, it is now possible to mark genes that are located within the boundaries of a BGC but do not partake in the biosynthesis, such as pseudogenes or transposable elements. Additionally, we have separated biosynthetic classification from compound classification (e.g., we removed "alkaloid" as a biosynthetic class) and introduced a custom biosynthesis-inspired chemical ontology for NPs (Supplementary Data 1, section 3.4) based on the work by Dewick (1). Notably, we have newly defined the nonribosomal peptide synthetase Type VI (modular, non-condensation-domain peptide-bond-forming), extending the current classification (18).

Biological activity and resource integration. MIBiG also accepts additional BGC-related data. In this update, we have reworked fields registering the biological activity of BGC-associated NPs: activities are now considered properties of a specific assay, and a controlled vocabulary (Supplementary Data 1, section 3.3) is available for defining bioactivity in a reproducible way. Additionally, we have included an optional "concentration" field, allowing submission of both qualitative and quantitative bioactivity data. At the same time, additional metadata parameters increase the scope of the already extensive Data Standard, and as such MIBiG references external resources where possible. Newly introduced links include references to the Minimum Information about a Tailoring Enzyme (MITE) data repository for annotation of tailoring enzyme-encoding genes (19), and CyanoMetDB for compound information on cyanobacterial NPs (20).

Community mobilization and data curation

Inspired by the contributions made to MIBiG 3.0, we again sought participation from the scientific community. Following calls on social media, 398 researchers signed up to participate in a series of eight three-hour online annotation sessions, accommodating different time zones (Figure 1). This enormous interest posed organizational challenges in terms of coordination and

communication, prompting us to develop a new model for community participation. Individual contributors were part of one or more Interest Groups that communicated using the MIBiG Slack (https://mibigannotathons.slack.com/) channel and were headed by Interest Group Coordinators: topic matter experts responsible for answering biosynthesis- and chemistry-related questions. Kanban-style boards (free version of Trello, https://trello.com/) were employed to coordinate work on entries. Data submission was performed using the MIBiG Submission Portal, a bespoke web interface that uses validated fields for data processing. Several curators with relevant expertise volunteered to take Reviewer roles, focusing on assessing the quality of newly generated or modified entries using the newly introduced peer review system. Aimed towards further improving the quality and confidence of entries, Reviewers could leverage the Kanban-style boards (Figure 2) to request revisions of entries if errors were found. To facilitate data curation, we prepared extensive online documentation (Supplementary Data 1) and instructional videos, and trained Interest Group Coordinators and Reviewers for their roles in online meetings. Participants who made a significant contribution (defined as participating in at least two three-hour sessions or an equivalent time investment) were invited to be co-authors in the present publication.
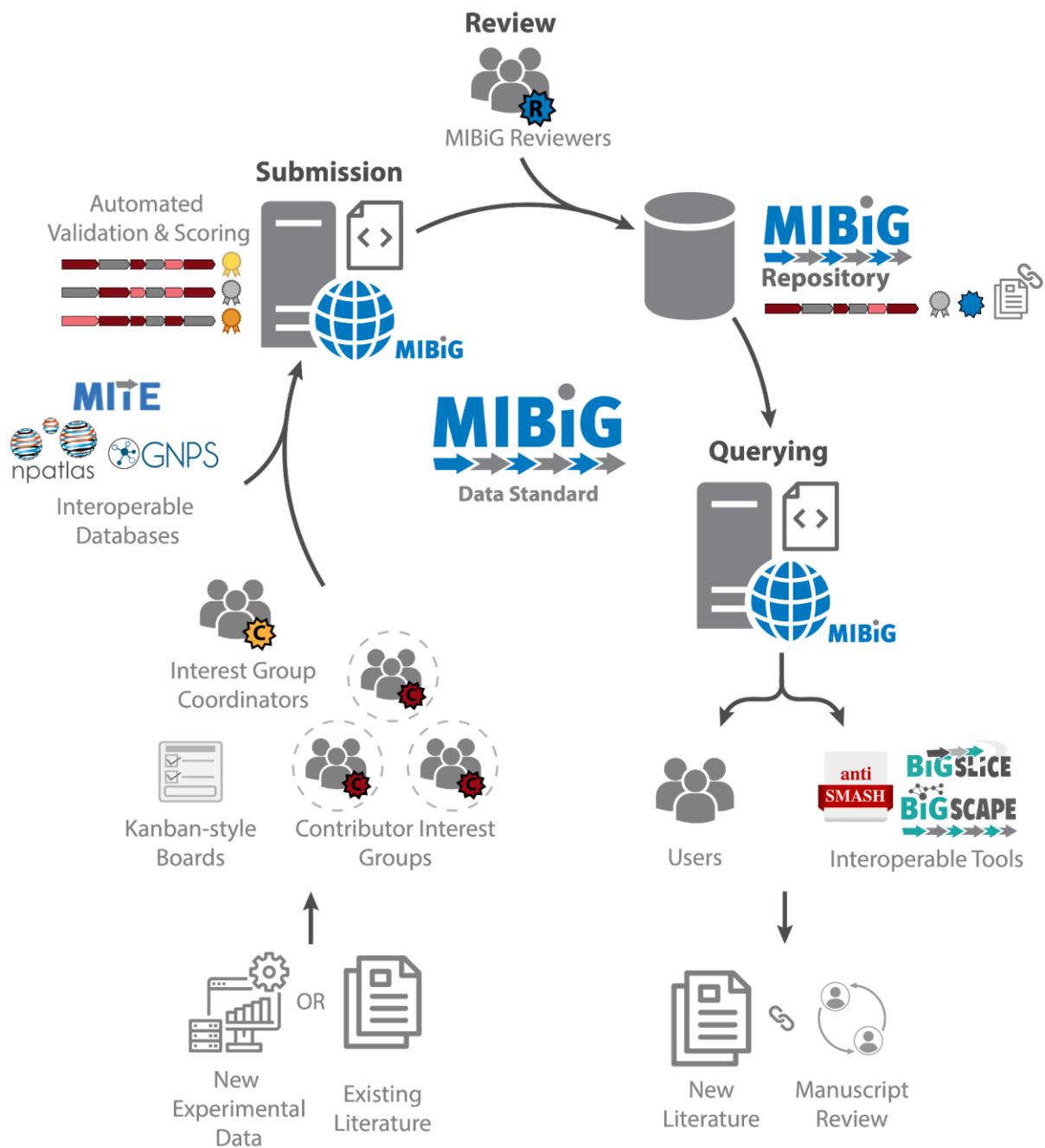
Figure 1. General workflow of the MIBiG annotation process. Data is submitted by annotathon contributors (organised by expertise into interest groups) or independent submitters to the database from new experimental data or existing/recent literature. The entries are then assessed by reviewers and revised when needed. Finally, they end up in the online MIBiG repository and become accessible by querying them online on the MIBiG web page or via interoperable tools.
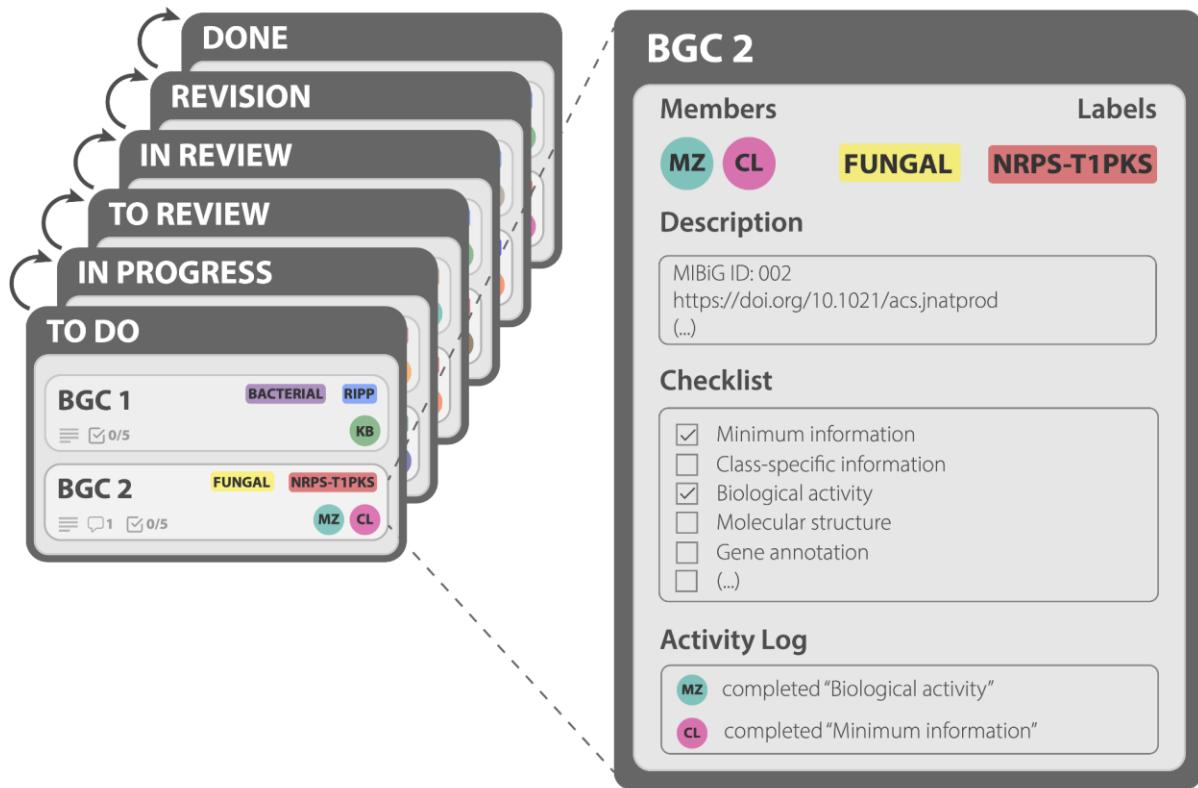
Figure 2. Architecture of the Kanban board used for the MIBiG annotathons. Every BGC would have its own 'card', where annotators with specific expertise could fill in and then check a specific part of its annotation. Once the checklist was complete, the card would move to review and, potentially, revision to repair any issues identified by the reviewers.

RESULTS AND DISCUSSION

Advancing the MIBiG Data Repository

In this iteration of the MIBiG annotathons, we put a greater emphasis on self-organization and facilitating motivated contributors to act independently and confidently when curating data. During the call for participation, researchers not only signed up to participate, but also contributed to assembling a list of recent publications associated with the biosynthesis of NPs. This initial effort yielded 552 publications supporting new entries and 266 publications for improvements of existing entries, which were used as a starting point for the curation process. Over the course of the annotathons, 267 contributors made a total of 8304 edits (e.g. adding an entirely new entry, adding biological activity to an existing entry, etc.), resulting in 557 new and 590 modified existing

entries. With the present update, MIBiG now contains a total of 3059 entries, a 22% increase in comparison to MIBiG 3.0. Of these, 1655 entries are now associated with 3604 biological activities, and 2634 entries have 5002 associated chemical structures. However, 672 entries still lack chemical structures; hence, future efforts will include attention to improving this aspect, especially with regard to structural information for ribosomally synthesized and post-translationally modified peptides (RiPPs). Additionally, 7677 references and 8582 evidence qualifiers were provided, 171 biosynthetic paths were described for 110 entries, and cross-references to 173 MITE and 93 CyanoMetDB entries were established.

Of the total 1147 contributed entries (557 new, 590 modified), 464 (40%) have been reviewed at the time of manuscript preparation. While all entries are available, those that are reviewed are highlighted in the MIBiG repository website to reflect the additional confidence. For applications using the MIBiG data where a high confidence level is required (e.g. machine learning applications), we recommend the use of reviewed entries only (the website facilitates filtering/sorting on this). We expect the "reviewed" part of the MIBiG repository to grow continuously once we have transitioned to the MIBiG rolling release model, and over time, we aim to formally review all entries in the MIBiG repository.
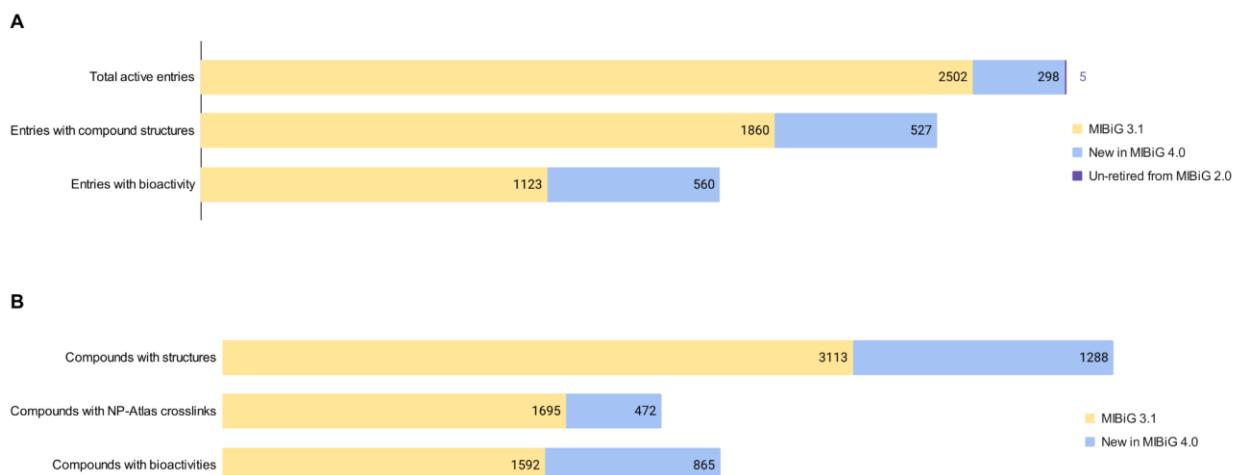
Figure 3. Quantitative overview of updates to the MIBiG database. Changes/updates compared to the previous version 3.1 are indicated in blue.

Initiating the MIBiG Rolling Release Model

The efforts described above demonstrate the value of leveraging large community initiatives such as the MIBiG annotathons. We estimate that contributors volunteered approximately 4000 hours in curating and reviewing entries, an effort in time and expertise that could not be raised by any single research group. Besides expanding the MIBiG repository, the annotathons were appreciated for their community-building aspect, fostering communication and exchange of ideas in the NP research community. In addition, the interaction with other resources prompted improvements to these databases as well, e.g. when curators could not find matching entries for a structure in the NP Atlas, thus encouraging wider cooperation beyond MIBiG itself. The intrinsic motivation of contributors is demonstrated by the fact that many edits were performed in between the annotathon sessions (Figure 2). The broad interest of the community motivated the planning of a "rolling release" model of MIBiG. In addition to the biennial efforts that will lead to "major" releases of MIBiG (e.g. the current v4.0, or the next major release v5.0), curators will be able to contribute new or modify existing entries on an ad hoc basis, leading to monthly "minor" releases (i.e. 4.1, 4.2). Contributors will be able to correct bugs and add references at any time, instead of waiting for the "major" release cycle to perform all edits at once. This new system is currently under development, and we invite the scientific community to participate in the discussion on how to structure contributions and governance (i.e., by communicating with the corresponding authors of this publication or using the MIBiG Slack). Furthermore, to facilitate future MIBiG updates and curation we encourage authors to release BGC sequence data during the publication submission and peer review process, or immediately thereafter, and to provide the respective accession details in the manuscript text.

In summary, we have conducted a large-scale community effort to make experimental data on NP BGCs freely accessible and machine-readable. As a resource created for and by the scientific community, the MIBiG repository is freely accessed on an entry-by-entry basis or can be downloaded and parsed in bulk. MIBiG 4.0 also serves as the stepping stone for creating the infrastructure to establish a Wikipedia-like model of continuous community curation. Such a decentralized organization will guarantee continuous development of MIBiG and help in including the next generations of scientists in the annotation and development process.

DATA AVAILABILITY

The MIBiG Repository is available at https://dev.mibig.secondarymetabolites.org/. Files in JSON format following the MIBiG data standard (https://github.com/mibig-secmet/mibig-json) can be found on the MIBiG webpage (https://mibig.secondarymetabolites.org/download) and on the MIBiG Zenodo Community page (10.5281/zenodo.13367755). Further materials are available on GitHub (https://github.com/mibig-secmet). All data is freely available with no restrictions for academic and commercial reuse under the OSI-approved CC BY 4.0 Open Source license (https://creativecommons.org/licenses/by/4.0/).

FUNDING

COMPETING INTEREST

JH and CSM are employees of Corteva Agriscience. BRT is a consultant for BioConsortia Inc. JJJvdH is member of the Scientific Advisory Board of NAICONS Srl., Milano, Italy and consults for Corteva Agriscience, Indianapolis, IN, USA. MHM is member of the Scientific Advisory Board of Hexagon Bio.

REFERENCES

1. Dewick,P.M. (2009) Medicinal Natural Products: A Biosynthetic Approach 3rd ed. Wiley-Blackwell, Hoboken, NJ.

2. Newman,D.J. and Cragg,G.M. (2020) Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.*, **83**, 770–803.

3. Fischbach,M.A., Walsh,C.T. and Clardy,J. (2008) The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 4601–4608.

4. Medema,M.H., de Rond,T. and Moore,B.S. (2021) Mining genomes to illuminate the specialized chemistry of life. *Nat. Rev. Genet.*, **22**, 553–571.

5. Blin,K., Shaw,S., Augustijn,H.E., Reitz,Z.L., Biermann,F., Alanjary,M., Fetter,A., Terlouw,B.R., Metcalf,W.W., Helfrich,E.J.N., *et al.* (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.*, **51**, W46–W50.

6. van Heel,A.J., de Jong,A., Song,C., Viel,J.H., Kok,J. and Kuipers,O.P. (2018) BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.*, **46**, W278–W281.

7. Skinnider,M.A., Johnston,C.W., Gunabalasingam,M., Merwin,N.J., Kieliszek,A.M., MacLellan,R.J., Li,H., Ranieri,M.R.M., Webster,A.L.H., Cao,M.P.T., *et al.* (2020) Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun.*, **11**, 6058.

8. Merwin,N.J., Mousa,W.K., Dejong,C.A., Skinnider,M.A., Cannon,M.J., Li,H., Dial,K., Gunabalasingam,M., Johnston,C. and Magarvey,N.A. (2020) DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl. Acad. Sci. U. S. A.*, **117**, 371–380.

9. Liu,M., Li,Y. and Li,H. (2022) Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *J. Mol. Biol.*, **434**, 167597.

10. Hannigan,G.D., Prihoda,D., Palicka,A., Soukup,J., Klempir,O., Rampula,L., Durcak,J., Wurst,M., Kotowski,J., Chang,D., *et al.* (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.*, **47**, e110.

11. Sanchez,S., Rogers,J.D., Rogers,A.B., Nassar,M., McEntyre,J., Welch,M., Hollfelder,F. and Finn,R.D. (2023) Expansion of novel biosynthetic gene clusters from diverse environments using SanntiS. *bioRxiv*, 10.1101/2023.05.23.540769.

12. Carroll,L.M., Larralde,M., Fleck,J.S., Ponnudurai,R., Milanese,A., Cappio,E. and Zeller,G. (2021) Accurate *de novo* identification of biosynthetic gene clusters with GECCO. *bioRxiv*, 10.1101/2021.05.03.442509.

13. Kalmer,T.L., Ancajas,C.M.F., Cheng,Z., Oyedele,A.S., Davis,H.L. and Walker,A.S. (2024) Assessing the ability of ChatGPT to extract natural product bioactivity and biosynthesis data from publications. *bioRxiv*, 10.1101/2024.08.01.606186.

14. Rajan,K., Zielesny,A. and Steinbeck,C. (2020) DECIMER: towards deep learning for chemical image recognition. *J. Cheminform.*, **12**, 65.

15. Medema,M.H., Kottmann,R., Yilmaz,P., Cummings,M., Biggins,J.B., Blin,K., de Bruijn,I., Chooi,Y.H., Claesen,J., Coates,R.C., *et al.* (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.*, **11**, 625–631.

16. Kautsar,S.A., Blin,K., Shaw,S., Navarro-Muñoz,J.C., Terlouw,B.R., van der Hooft,J.J.J., van Santen,J.A., Tracanna,V., Suarez Duran,H.G., Pascal Andreu,V., *et al.* (2020) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.*, **48**, D454–D458.

17. Terlouw,B.R., Blin,K., Navarro-Muñoz,J.C., Avalon,N.E., Chevrette,M.G., Egbert,S., Lee,S., Meijer,D., Recchia,M.J.J., Reitz,Z.L., *et al.* (2023) MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res.*, **51**, D603–D610.

18. Dell,M., Dunbar,K.L. and Hertweck,C. (2022) Ribosome-independent peptide biosynthesis: the challenge of a unifying nomenclature. *Nat. Prod. Rep.*, **39**, 453–459.

19. Zdouc,M.M., Meijer,D., Biermann,F., Holme,J., Korenskaia,A., Lien,A., Louwen,N.L.L., Navarro-Muñoz,J.C., Nguyen,G.-S., Rutz,A., *et al.* (2024) The Minimum Information about a Tailoring Enzyme/Maturase data standard for capturing natural product biosynthesis. *ChemRxiv*, 10.26434/chemrxiv-2024-78mtl. PREPRINT NOT PEER REVIEWED

20. Jones,M.R., Pinto,E., Torres,M.A., Dörr,F., Mazur-Marzec,H., Szubert,K., Tartaglione,L., Dell'Aversano,C., Miles,C.O., Beach,D.G., *et al.* (2021) CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. *Water Res.*, **196**, 117017.