Nucleic Acids Research, 2025, **53**, D678–D690
https://doi.org/10.1093/nar/gkae1115
Advance access publication date: 9 December 2024
Database issue

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

Received: September 16, 2024. **Revised:** October 16, 2024. **Editorial Decision:** October 16, 2024. **Accepted:** October 28, 2024

Noel Kubach[23], Aditya M. Kunjapur[126], Aleksandra K. Kushnareva[3,24], Andreja Kust[84,85], Jessica Lamber[3,24,127], Martin Larralde[128], Niels J. Larsen[2], Adrien P. Launay[129], Ngoc-Thao-Hien Le[26], Sarah Lebeer[89,130], Byung Tae Lee[123], Kyungha Lee[81], Katherine L. Lev[106,107], Shu-Ming Li[63], Yong-Xin Li[109], Cuauhtémoc Licona-Cassani[46,131], Annette Lien[1], Jing Liu[109], Julius Adam V. Lopez[132,133], Nataliia V. Machushynets[55], Marla I. Macias[134], Taifo Mahmud[34], Matiss Maleckis[135], Añadir Maharai Martinez-Martinez[46], Yvonne Mast[103,136], Marina F. Maximo[48], Christina M. McBride[106,107], Rose M. McLellan[137,138], Khyati Mehta Bhatt[82], Chrats Melkonian[1,139], Aske Merrild[140], Mikko Metsä-Ketelä[141], Douglas A. Mitchell[59], Alison V. Müller[3,24,127], Giang-Son Nguyen[113], Hera T. Nguyen[10], Timo H. J. Niedermeyer[142], Julia H. O'Hare[67], Adam Ossowicki[70,75], Bohdan O. Ostash[143,144], Hiroshi Otani[90,145], Leo Padva[61], Sunaina Paliyal[29], Xinya Pan[55,70,71], Mohit Panghal[29,30], Dana S. Parade[3], Jiyoon Park[116], Jonathan Parra[146,147,148], Marcos Pedraza Rubio[70,75], Huong T. Pham[81], Sacha J. Pidot[149], Jörn Piel[150], Bita Pourmohsenin[23], Malik Rakhmanov[39], Sangeetha Ramesh[151], Michelle H. Rasmussen[152], Adriana Rego[76], Raphael Reher[63], Andrew J. Rice[59], Augustin Rigolet[55], Adriana Romero-Otero[2,135], Luis Rodrigo Rosas-Becerra[55,57], Pablo Y. Rosiles[129], Adriano Rutz[153], Byeol Ryu[56], Libby-Ann Sahadeo[58], Murrel Saldanha[7], Luca Salvi[62], Eduardo Sánchez-Carvajal[154], Christian Santos-Medellin[115], Nicolau Sbaraini[10], Sydney M. Schoellhorn[155], Clemens Schumm[3,127,156], Ludek Sehnal[62,157], Nelly Selem[104], Anjali D. Shah[158,159], Tania K. Shishido[82,160], Simon Sieber[161], Velina Silviani[79,80], Garima Singh[162,163], Hemant Singh[29], Nika Sokolova[108], Eva C. Sonnenschein[52], Margherita Sosio[97], Sven T. Sowa[39], Karin Steffen[164], Evi Stegmann[62,165], Alena B. Streiff[150], Alena Strüder[62], Frank Surup[25,77], Tiziana Svenningsen[140], Douglas Sweeney[56], Judit Szenei[2], Azat Tagirdzhanov[3,24], Bin Tan[16,17], Matthew J. Tarnowski[52], Barbara R. Terlouw[1], Thomas Rey[166,167], Nicola U. Thome[55,168], Laura Rosina Torres Ortega[1], Thomas Tørring[140], Marla Trindade[169], Andrew W. Truman[170], Marie Tvilum[140], Daniel W. Udwary[90], Christoph Ulbricht[61], Lisa Vader[2], Gilles P. van Wezel[55,71], Max Walmsley[83], Randika Warnasinghe[58], Heiner G. Weddeling[39], Angus N.M. Weir[100,171], Katherine Williams[158,172], Sam E. Williams[2], Thomas E. Witte[173], Steffaney M. Wood Rocca[56,96], Keith Yamada[141], Dong Yang[174], Dongsoo Yang[175], Jingwei Yu[176], Zhenyi Zhou[177], Nadine Ziemert[23], Lukas Zimmer[61], Alina Zimmermann[103], Christian Zimmermann[178], Justin J.J. van der Hooft[1,41], Roger G. Linington[179], Tilmann Weber[2,*] and 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, Building 220, Søltofts Plads, 2800 Kongens Lyngby, Denmark
[3]Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Campus E8.1, 66123 Saarbrücken, Germany
[4]School of Chemistry, Chemistry Building, University of Sydney, Eastern Ave, Camperdown NSW 2050, Sydney, New South Wales, Australia
[5]Department of Chemistry, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany
[6]Myria Biosciences AG, Tech Park Basel, Hochbergstrasse 60C, 4057 Basel, Switzerland
[7]Department of Microbiology and Cell Science, University of Florida, 1355 Museum Drive, Gainesville, Florida, 32611, USA
[8]Ginkgo Bioworks, 27 Drydock Avenue, 8th Floor, Boston, MA 02210, USA
[9]University of Florida Genetics Institute, University of Florida, 2033 Mowry Rd, Gainesville, FL 32611, USA
[10]School of Molecular Sciences, University of Western Australia, 35 Stirling Highway, Perth 6009, Australia
[11]Chair of Technical Biochemistry, Technical University of Dresden, Bergstraße 66, 01069 Dresden, Germany
[12]School of Chemical Sciences, University of Auckland, 23 Symonds St, Auckland 1010, New Zealand
[13]Newcastle University, Biosciences Institute, Catherine Cookson Building, Newcastle upon Tyne, NE2 4HH, UK
[14]Sutro Biopharma, 111 Oyster Point Blvd, South San Francisco, CA, 94080, USA
[15]Department of Chemistry, Faculty of Mathematics and Natural Sciences, IPB University, Gedung Kimia Wing 1 Lantai 3, Jalan Tanjung Kampus IPB Dramaga, Bogor, Jawa Barat 16680, Indonesia
[16]Institute for Molecular Bio Science, Goethe University Frankfurt, Max-von-Laue Strasse 9, 60438 Frankfurt am Main, Germany
[17]LOEWE Center for Translational Biodiversity Genomics (TBG), Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[18]Senckenberg Society for Nature Research, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[19]Department of Chemistry, University of Warwick, Gibbet Hill Rd, Coventry, CV4 7AL, UK
[20]Warwick Integrative Synthetic Biology Centre (WISB), University of Warwick, Gibbet Hill Rd, Coventry, CV4 7AL, UK

[21]Department of Biotechnology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo, 113-8657, Japan
[22]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo, 113-8657, Japan
[23]Translational Genome Mining for Natural Products, Interfaculty Institute of Microbiology and Infection Medicine Tübingen (IMIT), Interfaculty Institute for Biomedical Informatics (IBMI), University of Tübingen, Sand 14, 72076 Tübingen, Germany
[24]Saarland University, Campus E8.1, 66123 Saarbrücken, Germany
[25]Department of Microbial Drugs, Helmholtz Centre for Infection Research (HZI), Inhoffenstr. 7, 38124 Braunschweig, Germany
[26]Université Paris Cité - Inserm Unit 1284, 75015 Paris, France
[27]Translational Genome Mining for Natural Products, Generare Bioscience, 75011 Paris, Île-de-France, France
[28]Sustainable Soils and Crops, Rothamsted Research, West Common, Harpenden, Hertfordshire, AL5 2JQ, UK
[29]Computational Biology Lab, National Agri-Food and Biomanufacturing Institute (NABI), Sector 81, S.A.S. Nagar, Mohali, Punjab 140306, India
[30]Regional Centre for Biotechnology, NCR Biotech Science Cluster, 3rd Milestone, Faridabad-Gurugram Expressway, Faridabad - 121001 Haryana (NCR Delhi), India
[31]Department of Chemistry and Biochemistry, University of Northern British Columbia, 3333 University Way, Prince George, BC, V2N 4Z9, Canada
[32]Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Colney Lane, Norwich, NR4 7UH, UK
[33]Latvian Institute of Organic Synthesis, Aizkraukles street 21, LV-1006 Riga, Latvia
[34]Department of Pharmaceutical Sciences, Oregon State University, 1601 SW Jefferson Way, Corvallis, OR 97331-3507, USA
[35]Department of Ecology, Evolution and Marine Biology, University of California, 1169 Biological Sciences II, Santa Barbara, CA 93106, USA
[36]Department of Environmental Microbiology, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, 8600 Duebendorf, Switzerland
[37]Plant-Microbe Interactions, Institute of Environmental Biology, Utrecht University, 3584 CH Utrecht, The Netherlands
[38]Faculty of Environment, Science and Economy, University of Exeter, TR10 9FE, Penryn, Cornwall, UK
[39]Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland
[40]State Key Laboratory of Microbial Metabolism, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Minhang District, Shanghai 200240, China
[41]Department of Biochemistry, University of Johannesburg, C2 Lab Building 224, Kingsway Campus, Cnr University & Kingsway Road, Auckland Park, Johannesburg 2006, South Africa
[42]International Research and Development (R&D) Division, Omnia Nutriology, Omnia Holdings Ltd, Building H, Monte Circle, 178 Montecasino Blvd, Fourways, Sandton, 2055, South Africa
[43]Department of Marine Resource Science, Faculty of Agriculture and Marine Science, Kochi University, 200 Otsu, Monobe, Nankoku-shi, Kochi, 783-8502, Japan
[44]Marine Core Research Institute, Kochi University, 200 Otsu, Monobe, Nankoku-shi, Kochi, 783-8502, Japan
[45]Department of Pharmacology and Toxicology, University of Utah, 30 S 2000 E, Salt Lake City, Utah, 84112, USA
[46]Industrial Genomics Laboratory, Centro de Biotecnología FEMSA, Escuela de Ingeniería y Ciencias, Tecnológico de Monterrey, Av. Eugenio Garza Sada 2501sur, Nuevo Leon, 64700, México
[47]Department of Chemistry, Purdue University, 610 Purdue Mall, West Lafayette, IN, 47907, USA
[48]Department of Organic Chemistry, Institute of Chemistry, University of Campinas (UNICAMP), Rua Monteiro Lobato 270, Campinas, São Paulo, 13.083-862, Brazil
[49]Department of Applied Sciences, University of Technology, Al-Sina'a St., 10066, Baghdad, Iraq
[50]School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK
[51]Department of Food and Animal Sciences, College of Agriculture, Tennessee State University, Nashville, TN 37209, USA
[52]Department of Biosciences, Swansea University, Singleton Park, Swansea, SA2 8PP, UK
[53]Escuela Nacional de Estudios Superiores Unidad Leon, Universidad Nacional Autonoma de Mexico, Blv. UNAM #2011, Predio El Saucillo y, Comunidad de los Tepetates, El Potrero, 37684 León de los Aldama, Gto., Mexico
[54]Department of Botany and Microbiology, College of Science, King Saud University, P.O. Box 2455, Riyadh, 11451, Saudi Arabia
[55]Institute of Biology, Leiden University, Sylviusweg 72, 2333BE Leiden, The Netherlands
[56]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0212, USA
[57]Unidad de Genómica Avanzada, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Km 9.6 Libramiento Norte Carretera Irapuato-León, 36824, Irapuato, Mexico
[58]Department of Chemistry, University of Manitoba, 66 Chancellors Circle, Winnipeg, MB R3T 2N2, Canada
[59]Department of Chemistry, University of Illinois Urbana-Champaign, 600 S. Matthews Ave, Urbana, IL 61801, USA
[60]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, 1206 W. Gregory Drive, Urbana, IL 61801, USA
[61]Institute of Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany
[62]Interfaculty Institute of Microbiology and Infection Medicine Tübingen, Microbial Bioactive Compounds, University of Tübingen, Auf der Morgenstelle 28, 72076, Tübingen, Germany
[63]Institute of Pharmaceutical Biology and Biotechnology, University of Marburg, Robert-Koch-Str. 4, 35037 Marburg, Germany
[64]Departamento de Microbiología Molecular, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Chamilpa, 62210 Cuernavaca, Morelos, México
[65]Facultad de Ciencias Biológicas, Universidad Nacional Mayor de San Marcos, Av. Germán Amézaga S/N, Ciudad Universitaria, UNMSM, Lima 15081, Perú
[66]Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, Avenida Athos da Silveira Ramos, no 149, Rio de Janeiro, 21941-909, Brazil
[67]Instituto de Pesquisas de Produtos Naturais Walter Mors, Universidade Federal do Rio de Janeiro, CCS - Av. Carlos Chagas Filho, 373 - Bloco H - Cidade Universitária, Rio de Janeiro - RJ, 21941-599, Brazil
[68]Department of Biology, Humboldt University Berlin, Invaliden-Str. 42, 10115 Berlin, Germany
[69]Institute of Marine Biotechnology (IMaB), University of Greifswald,Walter-Rathenau-Str. 49A, 17489 Greifswald, Germany

[70]Department of Crop Protection, Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora" (IHSM-UMA-CSIC), Campus Universitario de Teatinos, 29010, Málaga, Spain
[71]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10 6708 PB Wageningen, The Netherlands
[72]School of Cellular and Molecular Biology, Faculty of Biological Sciences, University of Leeds, 6 Clarendon Way, Woodhouse, Leeds, LS2 3AA, UK
[73]CICA – Centro Interdisciplinar de Química e Bioloxía, Universidade da Coruña, As Carballeiras, s/n, Campus de Elviña, 15071 A Coruña, Spain
[74]Centre of Biological Engineering, University of Minho, 4710-057, Braga, Portugal
[75]Department of Microbiology, Faculty of Science, Campus Universitario de Teatinos s/n, University of Málaga, 29010 Málaga, Spain
[76]Interdisciplinary Centre of Marine and Environmental Research (CIIMAR/CIMAR), University of Porto, 4450-208 Matosinhos, Portugal
[77]Institute of Microbiology, Technische Universität Braunschweig, Spielmannstraße 7, 38106 Braunschweig, Germany
[78]Microbial Interactions in Plant Ecosystems, IMIT/ZMBP, Eberhard Karls University of Tübingen, 72076 Tübingen, Germany
[79]College of Pharmacy, Yeungnam University, 280 Daehak-ro Gyeongsan-Si, Gyeongsangbuk-do 38541, Republic of Korea
[80]Research Institute of Cell Culture, Yeungnam University, 280 Daehak-ro Gyeongsan-Si, Gyeongsangbuk-do 38541, Republic of Korea
[81]College of Pharmacy and Drug Information Research Institute, Sookmyung Women's University, 100 Cheongpa-ro 47 gil, Seoul 04310, Republic of Korea
[82]Fungal Natural Products, Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584CT Utrecht, The Netherlands
[83]Manchester Institute of Biotechnology, Department of Chemistry, School of Natural Sciences, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
[84]Innovative Genomics Institute, University of California Berkeley, 2151Berkeley Way, Berkeley, CA 94720, USA
[85]Department of Earth and Planetary Science, University of California, 307 McCone Hall Berkeley, CA 94720-4767, USA
[86]Université Paris-Saclay, INRAE, UR BIOGER, 22 place de l'Agronomie, 91120 Palaiseau, France
[87]IOCB Prague, Czech Academy of Science, Flemingovo náměstí 542/2, 160 00 Praha 6, Czech Republic
[88]Department of Ecosustainable Marine Biotechnology, Stazione Zoologica Anton Dohrn, Giardini del Molosiglio, Via A.F. Acton 55, 80133 Naples, Italy
[89]Department of Bioscience Engineering, Research Group Laboratory of Applied Microbiology and Biotechnology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium
[90]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[91]Biochemistry of Microorganisms, Faculty of Life Sciences: Food, Nutrition, and Health, University of Bayreuth, Fritz-Hornschuch-Straße 13, 95326 Kulmbach, Germany
[92]SIB Swiss Institute of Bioinformatics, Centre Medical Universitaire, 1 rue Michel Servet, 1211 Geneva 4, Switzerland
[93]Department of Microbiology, University of Helsinki, Viikinkaari 9, 00017, Helsinki, Finland
[94]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
[95]Microbial Metabolites for Food Research Group, National Research and Innovation Agency (BRIN), Jl. Jogja - Wonosari Km 31.5, Gunungkidul, DI. Yogyakarta 55861, Indonesia
[96]Microbial and Environmental Genomics Group, J. Craig Venter Institute, 4120 Capricorn Lane, La Jolla, CA 92037, USA
[97]NAICONS Srl, viale Ortles 22/4, 20139 Milan, Italy
[98]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Biosciences complex – Pavilion 02, Sciences and Technology Campus 11/A, 43124 Parma, Italy
[99]Institute for Organic Chemistry, Leibniz Universität Hannover, Schneiderberg 38, 30167 Hannover, Germany
[100]VIB-KU Leuven Center for Microbiology, Flanders Institute for Biotechnology, Kasteelpark Arenberg 31, 3001 Leuven, Belgium
[101]Department of Biology, Laboratory for Biomolecular Discovery & Engineering, KU Leuven, Kasteelpark Arenberg 31, 3001 Leuven, Belgium
[102]Department of Biosystems, Biosensors Group, KU Leuven, Willem de Croylaan 42, box 2428, 3001 Leuven, Belgium
[103]Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Inhoffenstraße 7B, 38124Braunschweig, Germany
[104]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.
[105]Biosphere Sciences and Engineering Division, Carnegie Institution for Science, 3520 San Martin Dr, Baltimore, MD, 21218, USA
[106]Program in Chemical Biology, University of Michigan, 210 Washtenaw Avenue, Ann Arbor, MI, USA
[107]Life Sciences Institute, University of Michigan, 210 Washtenaw Avenue, Ann Arbor, Michigan, USA
[108]Department of Chemical and Pharmaceutical Biology, University of Groningen, Antonius Deusinglaan 1, 9713AV Groningen, The Netherlands
[109]Department of Chemistry, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[110]Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology (HKI), Adolf-Reichwein-Straße 23, 07745 Jena, Germany
[111]Institute of Infectious Disease Research, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S4L8, Canada
[112]Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S4K1, Canada
[113]Department of Biotechnology and Nanomedicine, SINTEF Industry, P.O.Box 4760 Torgard, N-7465 Trondheim, Norway
[114]Microbiology Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[115]Corteva Agriscience, 9330 Zionsville Road, Indianapolis, Indiana 46268, USA
[116]Natural Products Research Institute, College of Pharmacy, Seoul National University, 101 Daehak-ro, Jongro-gu, Seoul 110-744, Korea
[117]Molecular Targets Program, Center for Cancer Research, National Cancer Institute, Frederick, Maryland 21702-1201, USA
[118]Production Host Engineering Team, VTT Technical Research Centre of Finland Ltd, Maarintie 3, 02150 Espoo, Finland
[119]Department of Chemistry and Biochemistry, Laboratories of Molecular Recognition, Florida State University, 95 Chieftan Way, Tallahassee, FL 32306, USA
[120]Department of Applied Biology, College of Agriculture and Life Sciences, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju, 61186, South Korea

[121]Department of Pharmacology, College of Medicine, Dongguk University, Dongdae-ro 123, Gyeongju-si, Gyeongsangbuk-do 38066, Republic of Korea
[122]College of Pharmacy and Integrated Research Institute for Drug Development, Dongguk University-Seoul, Dongguk-ro 32, Goyang 10326, Republic of Korea
[123]Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea
[124]Department of Paleobiotechnology, Leibniz Institute for Natural Product Research and Infection Biology Hans Knöll Institute, Beutenbergstr. 11a, 07745 Jena, Germany
[125]DTU Bioengineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[126]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, USA
[127]German Centre for Infection Research (DZIF), Inhoffenstr. 7, 38124 Hannover-Braunschweig, Germany
[128]Center of Infectious Diseases, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, Netherlands
[129]Bioinformatics department, Endogenomiks, El Marqués, Querétaro, Mexico
[130]U-MaMi Centre of Excellence, Middelheimlaan 1, 2020 Antwerp, Belgium
[131]Integrative Biology Research Unit, The Institute for Obesity Research, Tecnológico de Monterrey, Av. Eugenio Garza Sada 2501 sur, Nuevo Leon, 64700. México
[132]School of Innovation and Sustainability, De La Salle University, Laguna Boulevard, LTI Spine Road, Brgys. Biñan and Malamig, 4024 Biñan City, Laguna, Philippines
[133]Department of Chemistry, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines
[134]University of Strathclyde, Strathclyde Institute of Pharmacy and Biomedical Sciences, Glasgow, G4 0RE UK
[135]Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 221, 2800 Kgs. Lyngby, Denmark
[136]Technical University of Braunschweig, Institute of Microbiology, Rebenring 56, 38106 Braunschweig, Germany
[137]Ferrier Research Institute, Victoria University of Wellington, Kelburn Parade, Wellington 6012, New Zealand
[138]Maurice Wilkins Centre for Molecular Biodiscovery, Victoria University of Wellington, Kelburn Parade, Wellington 6012, New Zealand
[139]Theoretical Biology and Bioinformatics, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands
[140]Department of Biological and Chemical Engineering, Aarhus University, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark
[141]Department of Life Technologies, University of Turku, Tykistökatu 6, FIN-20520 Turku, Finland
[142]Institute of Pharmacy, Freie Universität Berlin, Königin-Luise-Str. 2+4, 14195 Berlin, Germany
[143]Department of Genetics and Biotechnology, Ivan Franko National University of Lviv, Hrushevskoho st 4, Lviv 79005, Ukraine
[144]German-Ukrainian Core of Excellence in Natural Products Research, Zelena str. 20, Lviv 79005, Ukraine
[145]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
[146]Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San José, 11501-2060, Costa Rica
[147]Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, 1174-1200, San José, Costa Rica
[148]Instituto de Investigaciones Farmacéuticas (INIFAR), Facultad de Farmacia, Universidad de Costa Rica, San José, 11501-2060, Costa Rica
[149]Department of Microbiology and Immunology at the Doherty Institute, University of Melbourne, Grattan Street, Parkville Victoria, Victoria 3000, Australia
[150]Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog Weg 4, 8093 Zurich, Switzerland
[151]Department of Plant Pathology, University of California Davis, One Shields Avenue, Davis, CA 95616-8751, USA
[152]Department of Chemistry, Aarhus University, Langelandsgade 140, DK-8000, Aarhus C, Denmark
[153]Institute for Molecular Systems Biology, ETH Zürich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland
[154]Laboratorio de Microbiología Molecular y Biotecnología Ambiental, Centro de Biotecnología DAL, Universidad Técnica Federico Santa María, Avenida España 1680, 2390123, Valparaíso, Chile
[155]Department of Chemistry and BioDiscovery Institute, University of North Texas, 1155 Union Circle, Denton, Texas 76203, USA
[156]Helmholtz Centre for Infection Research (HZI), Inhoffenstr. 7, 38124 Braunschweig, Germany
[157]Masaryk University, Faculty of Science, RECETOX, Kamenice 753/5, 625 00 Brno, Czech Republic
[158]School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK
[159]School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK
[160]Institute of Biotechnology, Helsinki Institute of Life Science, University of Helsinki, Viikinkaari 5, 00790, Helsinki, Finland
[161]Department of Chemistry, University of Zurich,Winterthurerstrasse 190, 8057 Zurich, Switzerland
[162]Department of Biology, University of Padova, Via U. Bassi, 58/B, 35121 Padova, Italy
[163]Botanical Garden, University of Padova, Via Orto Botanico 15, 35123 Padova, Italy
[164]Department of Biological Sciences and Evolutionary Studies Initiative, Vanderbilt University, 465 21st Ave S, Nashville, TN 37232, USA
[165]German Centre for Infection Research (DZIF), Partner Site Tübingen, Hoppe-Seyler-Straße 3, 72076 Tübingen, Germany
[166]DE SANGOSSE, Bonnel, 47480, Pont-Du-Casse, France
[167]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
[168]Department of Biology, Laboratory for Biomolecular Discovery & Engineering, KU Leuven, Kasteelpark Arenberg 31, 3001 Leuven, Belgium
[169]Institute for Microbial Biotechnology and Metagenomics, Department of Biotechnology, University of the Western Cape, Robert Sobukwe Rd, Bellville, 7535, South Africa
[170]Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdome, Norwich, UK
[171]The Rosalind Franklin Institute, R113 Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Didcot, Oxfordshire, OX11 0QX, UK
[172]School of Applied Sciences, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol, BS16 1QY, UK
[173]Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie Curie Private, Ottawa, Ontario, K1N 6N5, Canada
[174]Natural Products Discovery Center, The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, University of Florida, Jupiter, Florida 33458, USA
[175]Synthetic Biology and Enzyme Engineering Laboratory, Department of Chemical and Biological Engineering, Korea University, Anam-ro 145, Seongbuk-gu, Seoul 02841, Republic of Korea

[176]Institute of Plant and Food Science, Department of Biology, School of Life Sciences, Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen 518055, P.R. China
[177]College of Pharmaceutical Science & Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Chaowang Raod 18, Hangzhou 310014, China
[178]Institute of Chemical, Environmental and Bioscience Engineering, TU Wien, Gumpendorfer Strasse 1a, 1060 Vienna, Austria
[179]Department of Chemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, V5A 1S6, Canada

*To whom correspondence should be addressed. Email: marnix.medema@wur.nl
Correspondence may also be addressed to Tilmann Weber. Email: tiwe@biosustain.dtu.dk
Correspondence may also be addressed to Kai Blin. Email: kblin@biosustain.dtu.dk
†The first two authors should be regarded as Joint First Authors.

Abstract

Specialized or secondary metabolites are small molecules of biological origin, often showing potent biological activities with applications in agriculture, engineering and medicine. Usually, the biosynthesis of these 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 prototype, 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://mibig.secondarymetabolites.org/.

Graphical abstract

Introduction

Many organisms are prolific producers of small molecules known as specialized or secondary metabolites (SMs). These molecules often show a diversity of potent biological activities, which have been leveraged for the development of numerous drugs (1,2). SMs are generally hypothesized to increase the fitness of the producing organism or its host. In microbes, the biosynthetic genes required for the production of an SM 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 biosynthesis, resistance and regulation of SMs and are the object of 'genome mining' strategies that leverage analysis of genome sequence data for the discovery of (novel) metabolites (4).

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 are neither readily available from the scientific literature nor 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 with 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 SM 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. In rare cases, a single gene may be responsible for the biosynthesis of a natural product, such as a large non-ribosomal peptide syn-

thases; these standalone genes are also entered into MIBiG due to their relevance to specialised metabolism. Conceptualized as an open data repository curated by and for the SM community, the MIBiG repository has seen three iterations of online community-driven data annotation and curation hackathons (also known as 'annotathons'), with >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 volunteer expert reviewers, who can request corrections from data submitters. Additionally, we have established an initial prototype for efficient and standardized data submission, and during the annotathons we utilized 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.

Materials and methods

Rework of the MIBiG Data Standard

The MIBiG Data Standard (from here onwards, 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 revised the Data Standard to accommodate advances in SM 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, etc.) 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. While newly added entries adhere to these changes, entries added in previous versions of MIBiG still follow the legacy format, and will be updated gradually over time. To summarize the data quality of an entry concisely, we also introduced a 'Quality' identifier, and it is possible to filter entries based on high, medium or questionable quality of data. Note that this label only reflects the presumed data quality of an MIBiG entry and does not address the quality of the underlying literature.

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 reintroduced, 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 pseudo-genes 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 SMs (Supplementary Data 1, section 3.4) based on the work by Dewick (1). Furthermore, we have newly defined the non-ribosomal 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 SMs: 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 SMs (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 3-h 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

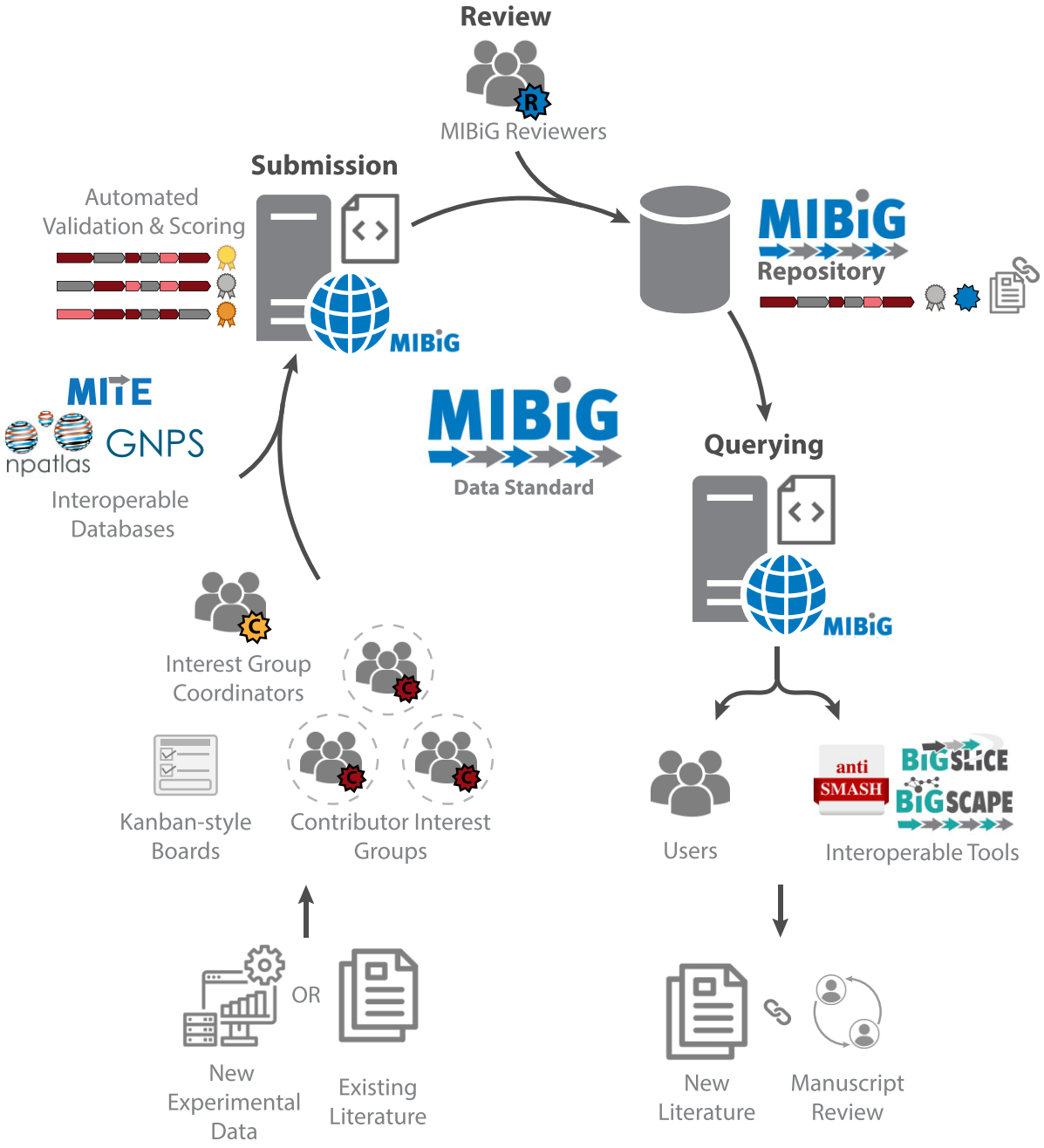

Figure 1. General workflow of the MIBiG annotation process. Data are submitted by annotathon contributors (organized by expertise into Interest Groups) or independent submitters to the database from new experimental data or existing/recent literature. The entries are then assessed by reviewers and revised 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.

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 a MIBiG Submission Portal prototype, a bespoke web interface that uses validated fields for data processing (code available at https://github.com/nlouwen/submission-prototype). 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 3-h sessions or an equivalent time investment) were invited to be co-authors in the present publication.

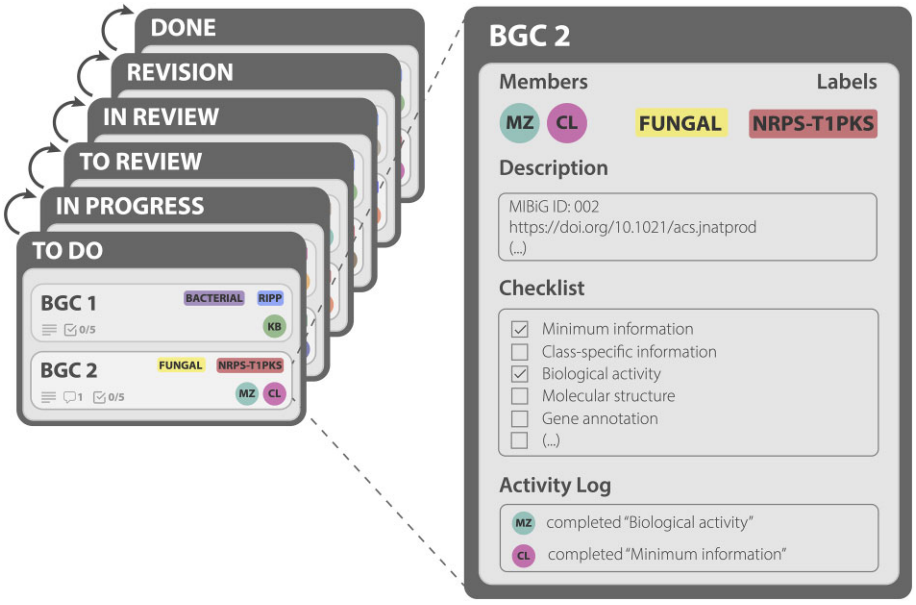

Figure 2. Architecture of the Kanban board used for the MIBiG annotathons. Every BGC would have its own 'card', where annotators with specific expertise could fill in and then check a specific part of its annotation. Once the checklist was complete, the card would move to review and, potentially, revision 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 SMs. 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. 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. A summary of the changes in comparison to MIBiG version 3.1 can be seen in Figure 3.

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.

Initiating the MIBiG rolling release model

The aforementioned efforts demonstrate the value of leveraging large community initiatives such as the MIBiG annotathons. We estimate that contributors volunteered ~4000 h 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 SM 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 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 quarterly 'minor' releases (i.e. 4.1 and 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 Workspace https://mibigannotathons.slack.com). 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.

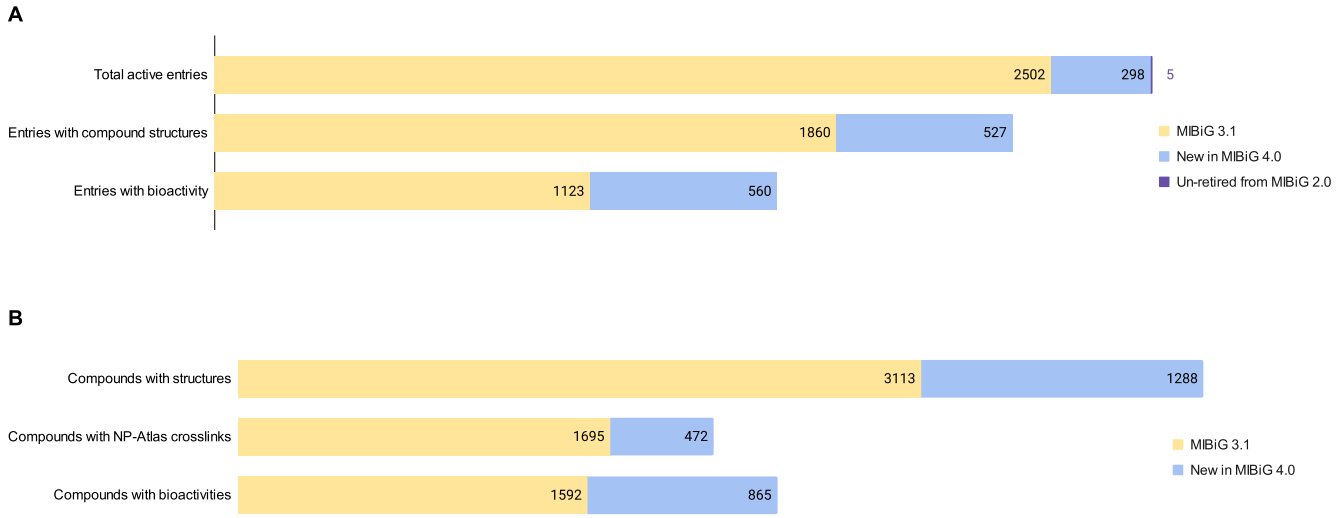

Figure 3. Quantitative overview of updates to the MIBiG database.in comparison with the previous version 3.1. Numbers in panel (**a**) refer to MIBiG entries, while numbers in panel (**b**) refer to individual compounds (a single MIBiG entry may contain more than one compound).

In summary, we have conducted a large-scale community effort to make experimental data on SM 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://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 (https://doi.org/10.5281/zenodo.13367755). Further materials are available on GitHub (https://github.com/mibig-secmet). All data are 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/).

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

The following contributors are acknowledged for their outstanding efforts as Interest Group Coordinators (in alphabetical order): Chantal D. Bader, Constance B. Bailey, Lena Barra, Thomas J. Booth, Kenan A. J. Bozhüyük, José D. D. Cediel-Becerra, Zachary Charlop-Powers, Marc G. Chevrette, Yit Heng Chooi, Paul M. D'Agostino, Tristan de Rond, Elena Del Pup, Katherine R. Duncan, Wenjia Gu, Novriyandi Hanif, Eric J. N. Helfrich, Matthew Jenner, Yohei K. Katsuyama, Aleksandra K. Korenskaia, Daniel Krug, Vincent Libis, Roger G. Linington, George A. Lund, Shrikant Mantri, Kalindi D. Morgan, Jorge Navarro, Charlotte Owen, Chin-Soon Phan, Benjamin Philmus, Zachary L. Reitz, Serina L. Robinson, Kumar Saurabh Singh, Robin Teufel, Yaojun Tong, Fidele Tugizimana, Dana Ulanova and Jaclyn M. Winter.

Funding

M.M.Z. was supported by the NWO Grant KICH1.LWV04.21.013 and by the Horizon 2020 Grant 101000392; C.L. was supported by the NWO Open Science Project 'BiG-CODEC' No. OSF.23.1.044; C.D.B was supported by the German Research Foundation Grant No. 547394769; C.B.B. was supported by the University of Sydney Drug Discovery Initiative; T.J.B was supported by the Novo Nodisk Foundation Grant NNF22OC0078997; Y.H.C. was supported by the Australian Research Council Industry Fellowship IM230100154; P.M.D. was supported by the Hans Fischer Society; K.R.D. was supported by the UK Government Department for Environment, Food & Rural Affairs (DEFRA) Global Centre on Biodiversity for the Climate and by the United Kingdom Research and Innovation (EP/X03142X/1) and by the Horizon Europe Marie Skłodowska-Curie grant agreement No 101072485; N.H. was supported by the Indonesia Endowment Fund for Education Agency (LPDP) and National Research and Innovation Agency (BRIN) of the Republic of Indonesia (106/IV/KS/11/2023 and 41644/IT3/PT.01.03/P/B/2023) and by the the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia (027/E5/PG.02.00.PL/2024); M.J. was supported by the UKRI Future Leaders Fellowship (MR/W011247/1); A.Korenskaia was supported by the Horizon Europe Marie Skłodowska-Curie grant agreement No 101072485; V.L. was supported by the ERC Starting Grant 101117891-MeDiSyn and by the Agence Nationale de la Recherche project ANR-22-CE44-0011-01 UMISYN; G.A.L. was supported by the Growing Health Institute Strategic Programme (BB/X010953/1; BBS/E/RH/230003B); S.M. was supported by the Department of Biotechnology

(DBT), Government of India and by the National Agri-Food Biotechnology Institute (NABI); C-S.P. was supported by the EU project No. 101087181 (Natural Products Research at Latvian Institute of Organic Synthesis as a Driver for Excellence in Innovation); R.T. was supported by the Swiss National Science Foundation (SNSF), grant 212747; Y.T. was supported by the National Key Research and Development Program of China (2021YFA0909500) and by the the National Natural Science Foundation of China (32170080 and 32370026) and by the Shanghai Pilot Program for Basic Research - Shanghai Jiao Tong University; D.U. was supported by the Japan Society for Promotion of Science KAKENHI grant number 21K06336; D.Y.A. was supported by the São Paulo Research Foundation (FAPESP) research scholarship (grant 21/07038-0); M.A. was supported by the NWO Talent programme Veni science domain (VI.Veni.202.130); F.A. was supported by the UKRI Future Leaders Fellowship (MR/V022334/1); G.A. was supported by the USDA Evans-Allen Research Grant (222676); N.E.A. was supported by the National Center for Complementary and Integrative Health of the NIH under award number F32AT011475; H.B.B. was supported by the São Paulo Research Foundation (FAPESP) research scholarship (grant 21/08947-3); S.E.B. was supported by the National Science Foundation Graduate Research Fellowship (DGE 21-46756) and by the the University of Illinois Urbana-Champaign Illinois Distinguished Fellowship; C.B. was supported by the European Union Horizon 2020 research and innovation program (ERC Grant number: 802736, MORPHEUS); J.B-A. was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) [735867]; K.M.B. was supported by the NWO Merian fund (Micro-GRICE); M.W.B. was supported by the United Kingdom Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) funded White Rose Mechanistic and Structural Biology Doctoral Training Program (BB/T007222/1); L.B. was supported by the Horizon Europe Marie Skłodowska-Curie Actions Postdoctoral Fellowship funded by the European Union (Project chelOMICS - grant agreement No. 101066127); V.J.C. was supported by the Ministerio de Ciencia, Innovación y Universidades project RYC2020-029240-I; R.C-B. was supported by the scholarship SFRH/BD/136367/2018 by Fundaçao para a Ciencia e Tecnologia (FCT); G.C-F. was supported by the National Institutes of Health (NIH)/NIGMS K12 GM068524 Award as a San Diego IRACDA Scholar.; E.C-G. was supported by the HZI POF IV Cooperativity and Creativity Project Call; V.C. was supported by the Alexander von Humboldt-Stiftung (Ref: 3.5-IND-1199743-HFST-P) and by the Cluster of Excellence: Controlling Microbes to Fight Infection (CMFI-YIG (EXC-2124/1-09.029_0; H.C. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1A6A1A03044512, and NRF-2021R1A2C1010727); Y.C. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1C1C1004046 and NRF-2022R1A5A2021216); J.A.C. was supported by the Signals in the Soil program via UK Research and Innovation (UKRI; NE/T010959/1); K.E.C. was supported by the Chan Zuckerberg Initiative Foundation grant number CZIF2022-007203; M.C. was supported by the German Research Foundation (DFG) Grant No. 495740318; J-F.D. was supported by the Agence Nationale de la Recherche (ShySM grant ANR-24-CE20-7299-01 and by the EUR Saclay Plant Sciences-SPS (ANR-17-EUR-0007); L.C.D-R. was supported by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN); T.D. was supported by the European Regional Development Fund, Programme Johannes Amos Comenius project 'IOCB MSCA PF Mobility' no. CZ.02.01.01/00/22_010/0002733; M.dK. was supported by the European Union's Horizon 2020 research and innovation program (ERC Grant number: 802736, MORPHEUS).; A.Fetter was supported by the United Kingdom Research and Innovation (EP/X03142X/1) and by the Horizon Europe Marie Skłodowska-Curie grant agreement No 101072485; M.F. was supported by the Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI; A.Frediansyah was supported by the Fulbright Grant (PS00349981); A.Gavriilidou was supported by the Deutsche Forschungsgemeinschaft [398967434-TRR 261]; A.Gentile was supported by the Italian Ministry of Research (Grant DM60066); H.G. was supported by the Research Foundation–Flanders (FWO) under the scope of junior postdoctoral fellowship (1229222N); L.A.G-S. was supported by the CONAHCYT scholarship (#971765); N.E.G. was supported by the NIGMS R01-GM146224, NERRS NA22NOS4200050; C.G. was supported by the BBSRC (BB/V005723/2); J.E.G.G. was supported by the CONAHCYT scholarship (#1347411); S.G. was supported by the National Institutes of Health (NIH), Grant T32GM136583. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.; L.H. was supported by the NIH F31 1F31ES036421-01; T.Hebra was supported by the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101130799; T.Horch was supported by the Novo Nordisk Foundation grant-number: CFB 2.0, NNF20CC0035580; M.I. was supported by the Italian Ministry of Research (Grant DM60066); E.J. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1C1C1004046 and NRF-2022R1A5A2021216); K.B.K. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1C1C1004046 and NRF-2022R1A5A2021216); S.K. was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-RS-2024-00408499) and by the National Research Foundation of Korea grants funded by the Republic of Korean Government (Ministry of Science and ICT) (NRF-RS-2024-00352229); W.K. was supported by the National Research Foundation of Korea grant funded by the Korea government (No. 2022R1C1C2004118); G.J.K. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1A6A1A03044512, and NRF-2021R1A2C1010727); H.K. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) Grants NRF 2018R1A5A2023127 and RS-2023-00211868; M.K. was supported by the Werner Siemens Foundation grant Paleobiotechnology; R.A.K. was supported by the NWO-XL (OCENW.XL21.XL21.088); A.T.K. was supported by the Danish National Research Foundation CeMiSt, DNRF137 and by the Novo Nordisk

Foundation INTERACT, NNF19SA0059360; A.M.K. was supported by the U.S. National Science Foundation (CBET-2032243); A.Kust was supported by the Delta Stewardship Council Delta Science Program; S.L. was supported by the European Research Council under European Union's Horizon 2020 Research and Innovation Program ERC St grant 852600 Lacto-Be; S-M.L. was supported by the Deutsche Forschungsgemeinschaft LI844/11-1 and LI844/14-1; A.L. was supported by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN); M.I.M. was supported by the Conahcyt Mexico International PhD Studentship and by the Strathclyde University Global Research Scholarship; M.M. was supported by the Novo Nordisk Foundation (Grant NNF23OC0082881) and Innovation Fund Denmark (Grant 3141-00013A); Y.M. was supported by the Leibniz Association grant K445/2022; M.F.M. was supported by the São Paulo Research Foundation (FAPESP) research scholarship (grant 23/01956-2); C.M.M. was supported by the NSF GRFP (#DGE 2241144); C.M. was supported by the MiCRop Consortium (NWO/OCW grant no. 024.004.014); A.M. was supported by the Carlsberg Foundation (CF22-1239); G-S.N. was supported by the SINTEF internal projects, POP-SEP BiocatDB (102022750), SEP AGREE (102029187) and POS BIOINFO 2024 (102024676-14), European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 101000392 (MARBLES), no. 101081957 (BLUETOOLS), and no. 862923 (AtlantECO); A.O. was supported by the Marie Sklodowska-Curie grant No. 101106349; B.O.O. was supported by the BG-21F, Ministry of Education and Science of Ukraine and by the 57/0009, National Research Fund of Ukraine (partial support); H.O. was supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231; L.P. was supported by the German Academic Scholarship Foundation; X.P. was supported by the NWO-XL grant OCENW.GROOT.2019.063; M.P. was supported by the Department of Biotechnology (DBT), Government of India and by the University Grants Commission (UGC), Ministry of Education, Government of India; M.P.R. was supported by the Spanish "Junta de Andalucía" project PROYEXCEL_00012; H.T.P. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1C1C1004046 and NRF-2022R1A5A2021216); S.J.P. was supported by the National Health and Medical Research Council GNT2021638, Australian Research Council Discovery Project DP230102668; B.P. was supported by the SECRETed EU Project Horizon 2020 (101000794); M.H.R. was supported by the European Research Council (ERC), European Union's Horizon 2020 Research and Innovation Program (grant agreement no. 865738); A.J.R. was supported by the Chemical-Biology Interface Training Grant (Grant T32-GM136629) and a National Science Foundation Graduate Research Fellowship (Grant DGE 21-46756); A.R. was supported by the ERC Advanced Grant 101055020-COMMUNITY; L.R.R-B. was supported by the Consejo Nacional de Ciencia y Tecnología (CONACyT) [757173]; L.Salvi was supported by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN); E.S-C. was supported by the PhD scholarships ANID N° 21231991; L.Sehnal was supported by the Horizon Europe Marie Skłodowska-Curie Actions Postdoctoral Fellowship funded by the European Union (Project NAfrAM - grant agreement No. 10106428); A.D.S. was supported by the Biotechnology and Biological Sciences Research Council-funded South West Biosciences Doctoral Training Partnership [BB/T008741/1]; T.K.S. was supported by the Novo Nordisk Foundation (Grant number: NNF22OC0080109); V.S. was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean Government (MSIT) (Grant No. NRF-2020R1A6A1A03044512, and NRF-2021R1A2C1010727); E.C.S. was supported by the Pathfinder Open 2022, a European Innovation Council (EIC) work programme that is part of Horizon Europe (grant agreement no. 101099528) and UK Innovation Funding Agency (UKRI) (reference no. 10062709).; M.S. was supported by the Italian Ministry of Research (Grant DM60066); K.S. was supported by the Swedish Pharmaceutical Society PostDoc stipend; A.B.S. was supported by the Swiss National Science Foundation (SNSF, 205320_219638); T.S. was supported by the Carlsberg Foundation (CF22-1239); J.S. was supported by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN); A.T. was supported by the Saarland University through the NextAID project; M.J.T. was supported by the Pathfinder Open 2022, a European Innovation Council (EIC) work programme that is part of Horizon Europe (grant agreement no. 101099528) and UK Innovation Funding Agency (UKRI) (reference no. 10062709).; T.T. was supported by the Carlsberg Foundation (CF22-1239); A.W.T. was supported by the BBSRC Institute Strategic Program grant (BB/X01097X/1); M.T. was supported by the AUFF (AUFF-E-2022-9-42); L.V. was supported by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN); G.P.vW. was supported by the ERC Advanced Grant 101055020-COMMUNITY; K.W. was supported by the MR/N029909/1, Medical Research Council, UK; S.E.W. was supported by the Novo Nordisk Foundation Postdoctoral Fellowship (NNF22OC0079021); T.E.W. was supported by the Natural Science and Research Council of Canada PGS-D scholarship; D.Y. was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Grant RS-2024-00440975); A.Z. was supported by the German Center for Infection Research (DZIF) (TTU 09.826); C.Z. was supported by the Austrian Science Fund (FWF) [10.55776/P 34036]; R.G.L. was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant program; T.W. was supported by the Novo Nordisk Foundation, NNF20CC0035580 and by the Danish National Research Foundation CeMiSt, DNRF137 and by the European Union's Horizon Europe programme under the Marie Skłodowska-Curie grant agreement No 101072485 (MAGic-MOLFUN).

Conflict of interest statement

J.H. and C.S.M. are employees of Corteva Agriscience. B.R.T. is a consultant for BioConsortia Inc. J.J.J.vdH. is member of the Scientific Advisory Board of NAICONS Srl, Milano, Italy and consults for Corteva Agriscience, Indianapolis, IN, USA. M.H.M. is a member of the Scientific Advisory Board of Hexagon Bio.

References

1. Dewick,P.M. (2009) In: *Medicinal Natural Products: A Biosynthetic Approach*. 3rd edn. 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 doi: https://doi.org/10.1101/2023.05.23.540769, 02 June 2023, preprint: not peer reviewed.
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 doi: https://doi.org/10.1101/2021.05.03.442509, 04 May 2021, preprint: not peer reviewed.
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 doi: https://doi.org/10.1101/2024.08.01.606186, 02 August 2024, preprint: not peer reviewed.
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 doi: https://doi.org/10.26434/chemrxiv-2024-78mtl, 09 April 2024, 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.

Received: September 16, 2024. **Revised:** October 16, 2024. **Editorial Decision:** October 16, 2024. **Accepted:** October 28, 2024