

Unearthing soil biodiversity through collaborative genomic research and education

The BioDIGS Consortium, Tristen Alberts, Claude F. Albritton, Rosa Alcazar, Zainab Aljabri, Maria Alvarez, Anish Aradhey, Mentewab Ayalew, Nareh Azizian, Yasmeen Balayah, Destiny D. Ball, Efren Barragan, Corey Beshoar, Lyle Best, Emily Biggane, Joseph Biggane, Jesse Blick, Myron Blosser, Alex Kenneth Brown, Michael C. Campbell, Zoe Canizares, Faith N. Chanhuhwa, Yu Chen, Daniel R. Chin, Kamal Chowdhury, Tyler Collins, Blair Compton, Jefferson Da Silva, Nia R. Davis, Natalie DeCaro, Frida Delgadillo, Youping Deng, Joseph Duncan, Arinzechukwu C. Egwu, Grace D. Ekalle, Noha Elnawam, Ray Enke, Naomi Ewhe, Marco A. Ferrel, Janna Fierst, Grace Freymiller, Karla Fuller, Lena Fulton-Wright, Valeriya Gaysinskaya, Torrence Gill, Ellie Gillespie, Perla Gonzalez Moreno, Sara Goodwin, Natajha Graham, Madeline E. Graham, Joseph L. Graves Jr., Emily Grob, Rachael Gutierrez, Aisha Hager, Shazia Tabassum Hakim, Aaliyah Harris, Ava M. Hoffman, Tobias Hoffmann, Alani M. Horton, Allison Hughes, Elizabeth M. Humphries, Josh-Samuel Ikechi-Konkwo, Aadil Ishtiaq, Ryan Jackson, Joshua Ronnie James, Kaitlan James, Sydney A. Jamison, Armando Jimenez, Rachel Johnson, Abigail Kauffman, Harkiran Kaur, Kritika Kc, Analyse Keeton, Olivia E. Kelly, Jennifer Kerr, Nataliya Kucher, Donna Lee Kuehu, Wendy A. Larson, Joslynn Lee, Andrew Lee, Jeffrey T. Leek, Danilo Lemaic, Lincoln E. Liburd II, Alan Fernando Lopez, Mohammadamin Mahmanzar, Karwitha Mamae, Raffi Manjikian, Michael Marone, Katerin Marquez, Amara Martinson, Senem Mavruk Eskipehlivan, Ashley Medrano, Melanie Melendrez-Vallard, Robert Meller, Loyda B. Méndez, Miguel P. Mendez Gonzalez, Nicolle Mesquita, Concepcion Martinez Miller, Isam Mohd-Ibrahim, Peter Mortensen, Stephen Mosher, Alketa Muja, Nadia Nasrin, Masaki Nasu, Matthew H. Nguyen, Ba Thong Nguyen, Michele Nishiguchi, Lance M. O'Connor, Disomi Okie, Tolulope Olowookorun, Alex Ostrovsky, Keyan Ozuna, Asmita Pandey, Shiv B. Patel, Gauri Paul, Shrikant Pawar, Andrea Pearson, Deborah Petrik, Jordan Platero, Carl Pontino, Arjun P. Pratap, Siddharth Pratap, Yujia Qin, Sudhir Kumar Rai, Nistha Ray, Ethan Repesh, Kristen Rhinehardt, Brennan Roche, Ariana Rodriguez, Shriya Roy, Sourav Roy, Alexa Sawa, Michael C. Schatz, Shurjo K. Sen, Randon Serikawa, Tyler Smith, Loraye Smith, James Snizek, Ryley D. Stewart, Edu B. Suarez-Martinez, Joelle Taganna, Frederick J. Tan, Nikolaos Tsotakos, Nwanneka Udolisa, Katherine Ulbricht, Tanner Veo, Jennifer Vessio, Lia Walker, Oscar Wang, Qingguo Wang, Robert Wappel, Kalynn Wesby, Malachi Whitford, Nicole Wild, Xianfa Xie, Hua Yang, Sayumi York & Lindsay Zirkle



The BioDIGS project is a nationwide initiative involving students, researchers and educators across more than 40 research and teaching institutions. Participants lead sample collection, computational analysis and results interpretation to understand the relationships between the soil microbiome, environment and health.

Soil is the most biodiverse habitat on the planet, home to around 59% of all species on Earth¹. These species include many millions of vertebrates, arthropods, annelids, nematodes, plants and fungi, as well as billions of bacteria, archaea, bacteriophages and other microorganisms. Through their interactions with their physical environment, themselves and other species, soil microorganisms maintain many essential ecological functions. For example, some soil microorganisms are involved in fundamental geochemical processes, such as nitrogen and carbon

cycles, whereas others closely interact with the animals, plants, fungi and other species in soil or in neighboring environments to exert vital roles in nutrition, development and health². Besides participating in these crucial ecological processes, soils are important reservoirs of both beneficial microorganisms and pathogens, including those that affect human health^{3,4}. Notably, the One Health perspective recognizes soil as a major source of antibiotic resistance genes, requiring a coordinated effort by scientists and other stakeholders to monitor trends, threats and ongoing changes in the soil microbiome⁵.

To address current health and environmental concerns, researchers must catalog soil biodiversity and understand microbial interactions with environmental factors. Metagenomic analyses of 16S gene sequences, a common technique to identify and classify bacteria and archaea, and short-read sequencing have provided important insights into soil bacterial communities and the environmental variables that influence their biodiversity (for example, soil type, heavy metals, climate, vegetation and land-use practices)^{5,6}. However, researchers still face several challenges⁷, particularly the existence of uncultured and uncharacterized microorganisms (that is, the microbial 'dark genomes'), which comprise up to 99% of soil microorganisms⁸.

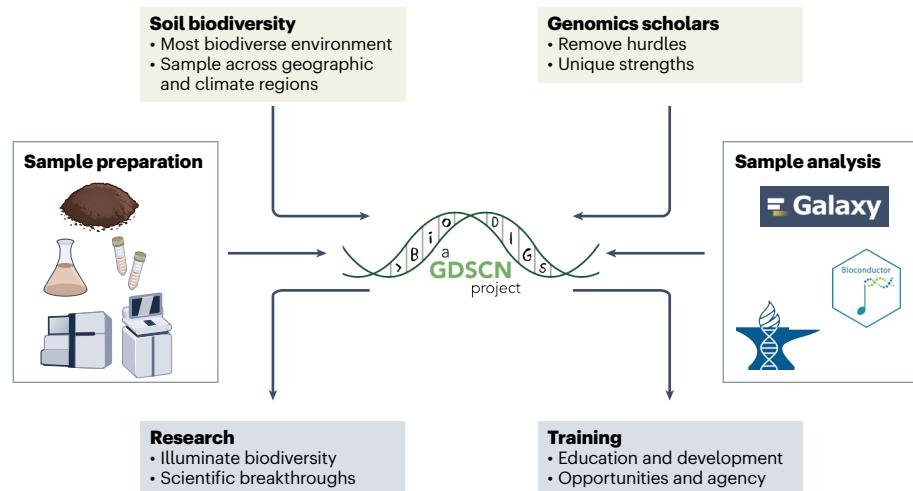


Fig. 1 | BioDIGS project overview. BioDIGS is an innovative research initiative aimed at understanding the interconnection between the environment, natural resources and public health. This collaborative project will lead to advancements in soil-based research, while at the same time providing training opportunities within this research space for genomics scholars. BioDIGS is an outgrowth of

the GDSCN, which works toward a vision in which all researchers, educators and students are able to fully participate in genomic data science. Logos are reproduced with permission from Galaxy Project, Bioconductor and NHGRI AnVIL Project.

Other major challenges include limited ability to link specific functions to particular microorganisms and the complexity of interactions among species to abiotic environmental factors, as well as the spatial and temporal dynamics of these interactions⁸. Further advancements in understanding the soil environment require new technologies and approaches that enable more refined data acquisition and analysis, such as long-read sequencing and artificial intelligence-based analysis methods.

Confronting these grand challenges requires a well-trained and innovative scientific workforce that is geographically distributed and represents different types of research and teaching institutions to bring multiple perspectives and resources. The BioDiversity and Informatics for Genomics Scholars (BioDIGS) Consortium (<https://biodigs.org>) was designed to address knowledge gaps in soil biodiversity, while training the next generation of scientists in genomic data science (GDS). BioDIGS is an initiative of the Genomic Data Science Community Network (GDSCN), a national collaboration in the USA with a unified goal of increasing access, training and opportunities in GDS for faculty and students⁹ (Fig. 1). Specifically, GDSCN aims to provide enhanced training opportunities in genomics and data science research, consisting of faculty, students and affiliates spanning from PhD-granting institutions (R1 institutions, very high research activity; R2 institutions, high research activity; and graduate-only institutions) to primarily undergraduate institutions, which are the most limited with regard to resources (2-year colleges granting associate's degrees, 4-year colleges granting bachelor's degrees) as classified by the Carnegie Classification of Institutions of Higher Education. Each group participating in BioDIGS determines relevant soil collection sites and contributes to the analysis, with the goals of broadening our understanding of soil biodiversity and increasing the distribution of participants in a nation-wide project. Furthermore, BioDIGS encourages institutions to develop student-driven, locally relevant research questions, thereby giving students an opportunity to be invested in the project and increase their connection to their environment. Equally important is that BioDIGS data can be used in the classroom to train students in fundamental GDS techniques.

To take part in BioDIGS, participants are sent a pre-assembled kit to collect soil samples from locations of their choice, for which they have obtained explicit permission (for example, campus, parks, urban spaces and other sites with local importance). Students and collaborating faculty members collect GPS coordinates, metadata and images of each location through free, cloud-based forms, before collecting several small volumes of soil 10–12 cm below the surface within a 10 m² area, which are aggregated following a shared, harmonized protocol (<https://biodigs.org/#protocols>). The team then ships the prepared samples for soil characterization (for example, heavy metal profiling and elemental analysis) and whole-genome sequencing using both short (for example, Illumina and Element) and long (for example, Oxford Nanopore and PacBio) reads. This hybrid sequencing approach allows a breadth of taxa to be detected, while enabling high-quality *de novo* genome assembly, and will enable analysis to compare the technologies. After sequencing, human reads are deleted and the data and metadata are uploaded to National Human Genome Research Institute (NHGRI) AnVIL¹⁰ and Galaxy¹¹. As the number of participating institutions grows, the BioDIGS team is actively working on documentation for data-sharing agreements to ensure that proper approvals are in place prior to widely sharing the data. This approach will promote authentic partnerships by establishing transparency and awareness of results, respecting data sovereignty (particularly tribal data sovereignty), and fostering an environment for responsible data sharing.

The use of pre-assembled soil-collection kits and cloud-based analysis makes BioDIGS flexible and accessible, so that any student can participate without the need for extensive lab or computing resources. We encourage local teams to perform analyses that interest them, including metagenome assembly and novel species detection, taxonomic and functional diversity analyses, statistical approaches (for example, permutational analysis of variance (PERMANOVA)) linking taxonomic diversity to heavy metal concentration, antimicrobial resistance gene discovery and enrichment of biosynthetic gene clusters. Exploring local health and environmental challenges is

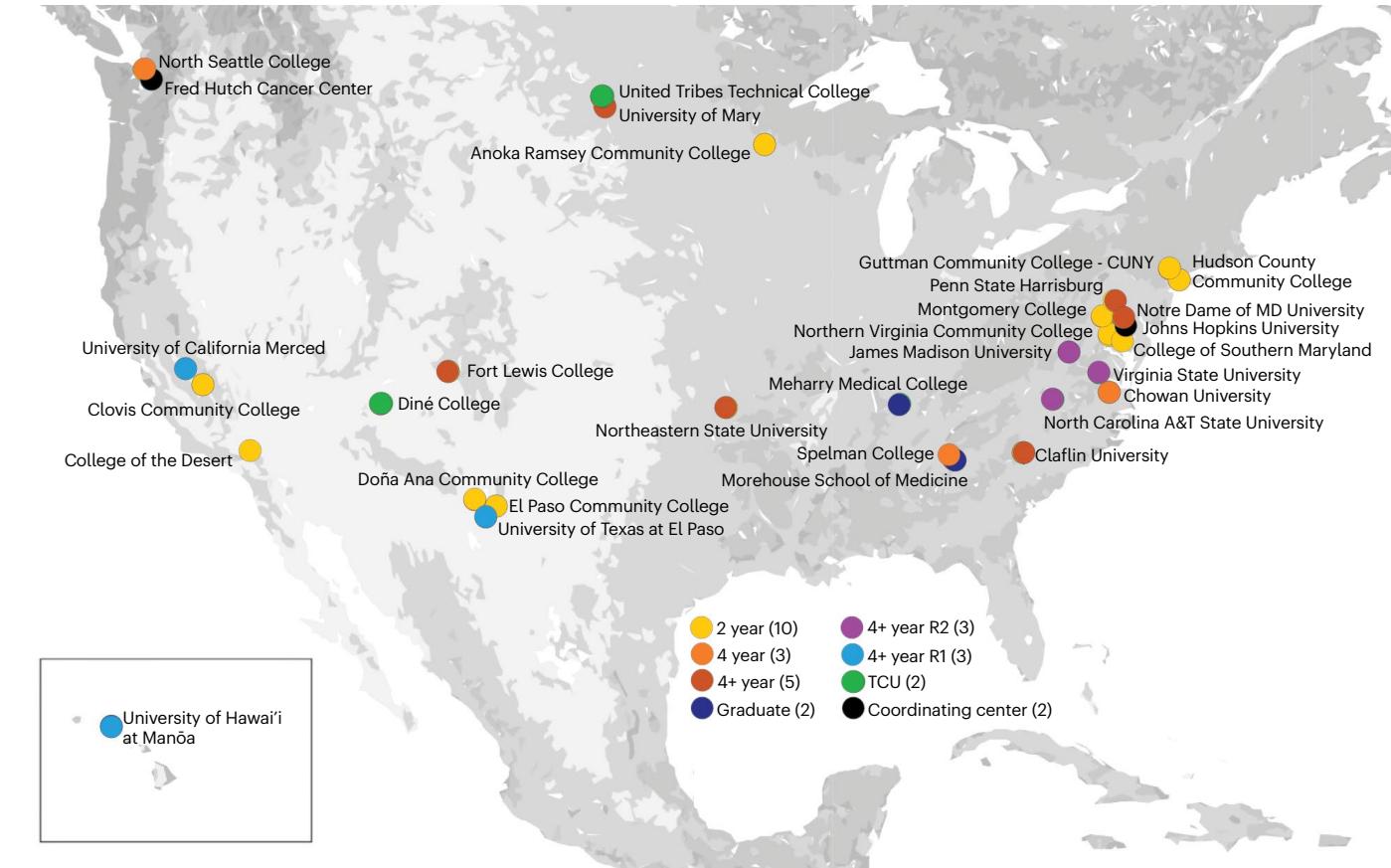


Fig. 2 | BioDIGS community engagement. Top, map of collection sites as of February 2025. The institutions involved in sample collection include ten 2-year institutions (2 year), three 4-year institutions (4 year), five primarily undergraduate institutions with a limited number of graduate programs (4+ year), two graduate-only institutions (Graduate), three R2 research institutions (4+ year R2), three R1 research institutions (4+ year R1), and two Tribal Colleges and Universities (TCUs) as classified by the Carnegie Classification of Institutions of Higher Education. There are also two R1 institutions that serve as coordinating centers. Bottom, sample of environments profiled. Participating faculty and students have chosen a variety of environments for sampling, including urban, rural and agricultural areas. Top row from left to right includes: Minnesota

(Anoka Ramsey Community College), California (Clovis Community College), New Jersey (Guttman Community College), Virginia (James Madison University), Georgia (Spelman College), Texas (University of Texas - El Paso) and Colorado (Fort Lewis College). Bottom row from left to right includes: Virginia (Virginia State University), Georgia (Claflin University, SC), Texas (El Paso Community College), North Dakota (United Tribes Technical College), Virginia (Northern Virginia Community College), Tennessee (Meharry Medical College) and California (University of California Merced). Images of additional sampling locations can be found on the sampling map at the biodigs.org website. Imagery ©2025, Map data ©2025 Google, INEGI.

encouraged to cultivate project ownership among students and create a connection to their locality through research. Overall, we strive for authentic partnerships, promoting scientific inquiry and student academic success.

The first of our major goals is to identify and illuminate the soil microbial dark genomes by characterizing novel species and gene sequences across diverse environments. Our preliminary results have shown that existing reference databases are severely lacking in

BOX 1

Student testimonials

"I chose to collect soil throughout my college campus at Notre Dame of Maryland University, where the landscape offers a diverse range of soil textures. Being part of the BioDIGS project is especially meaningful, as it integrates microbiology and the rapidly growing field of bioinformatics to analyze datasets of microbial populations in soil ecosystems. Engaging in this project is an honor to contribute to, allowing me to collaborate with driven, like-minded individuals who share a passion for microbiological research." — Loraye Smith, Notre Dame of Maryland University, currently attending Johns Hopkins University.

"Participating in this project was particularly exciting due to the unique biogeographical diversity of Hawai'i — a natural laboratory with varied ecosystems, soil types, and microclimates. I am honored to contribute to a nationwide effort analyzing soil microbiomes across diverse locations. My involvement reinforced the importance of collaborative, large-scale research in understanding Earth's microbiomes, and I am proud to support a project that bridges fieldwork, cutting-edge genomics, and education." — Asmita Pandey, University of Hawai'i.

"I had recently learned, to my surprise, that there was a US Environmental Protection Agency Hazardous Waste Cleanup site just 45 minutes from my high school, and I was curious about how the mercury contamination there affected soil microbes. With the help of Virginia Department of Environmental Quality scientist Calvin Jordan, my group sampled soil at two public parks along the river next to

the cleanup site, one park upstream and one downstream. It was fascinating how much data we could explore just through a simple question and a few soil samples." — Anish Aradhey, Harrisonburg High School, currently attending University of North Carolina.

"My participation in the BioDIGS project began with a simple curiosity about the soil in Lincoln Park, a heavily frequented public space in Jersey City. By contributing to this broader comparative study, I hope to gain insights into the potential variations and similarities in soil microbiomes, revealing how human activity and environmental factors shape these essential ecosystems. I am incredibly grateful to my professor, Raffi Manjikian, for allowing me to participate in this project, which has been a valuable opportunity to engage in real-world scientific inquiry and contribute to a larger understanding of our local environment." — Aadil Ishtiaq, Hudson County Community College.

"We selected four diverse sites for sample collection, including Claflin University in Orangeburg, South Carolina, and locations near my home in Augusta, Georgia, to ensure a broad representation of biodiversity. Collaborating with Dr. Chowdhury made the fieldwork both insightful and enjoyable, as we explored natural environments together. Collecting these samples was not only a valuable scientific experience but also a rewarding personal journey. We hope these samples reveal rich biodiversity of these regions and contribute to future innovations in science and biotechnology." — Kalynn B. Wesby, Claflin University.

soil microorganisms. Consequently, we use long-read sequencing to assemble high-quality genomes *de novo*, which is much more successful than short-read sequencing alone¹². Verification, analysis and annotation of these genomes will lead to individual genome announcements followed by broader studies across the entire dataset. Furthermore, uncovering the uncharacterized biodiversity of soil allows us to enhance our understanding of the connection between soil, soil biodiversity and other dimensions, including human health. For example, we will be able to better understand the dynamics of the soil microbiome over time and elucidate how soil biodiversity is affected by soil properties, such as pH and heavy-metal content. The presence of antimicrobial resistance (AMR) in soil is also of particular interest. The detection and tracing of AMR genes in soil can provide crucial information for managing and controlling AMR in other populations. Our long-term strategy is to build a network of regional soil characterization labs, scaling for additional consortium members that will allow us to track microbial populations over time and space.

Our second goal is to promote academic, educational and career development in GDS through a variety of opportunities provided to students and faculty. Through the development of new educational materials and curation of existing materials for classroom use, the project will support open-source learning modules on statistical, algorithmic and machine learning principles of GDS, as well as the integration of BioDIGS in course-based undergraduate research experiences (CUREs). The curriculum includes primers on the basics of DNA sequencing, the importance of coverage and read length for *de novo*

genome assembly, and the algorithms and data structures needed for genome assembly and annotation, species classification and related bioinformatics challenges. Furthermore, the statistics for quantifying species abundances and associating genome abundance with the environmental factors of a site are discussed. The program will also support the use of BioDIGS as an independent research project for undergraduates. Importantly, the project provides students the opportunity to participate in the larger consortium effort, spanning from sampling through to analysis. Students will thus not only contribute to the research project and gain an appreciation for its scale and complexity, but also build community across institutions. Finally, students will have the opportunity to present their work at conferences, further establishing and solidifying connections within the GDSCN and the wider community of GDS researchers. It is expected that these varied modalities of student engagement will enhance learning and create a path to the field of GDS for students.

To date, BioDIGS has worked with over 100 student researchers from around the USA, spanning various types of institutions (Fig. 2), with opportunities to engage many more in the future. Students can participate in sampling and/or analysis as part of a class lab activity, through CUREs, or in the form of independent research projects. We have seen these efforts broadly promote education and research (Box 1), and we hope they will strengthen the GDS workforce. In this regard, an immediate need from the research community is to provide evolving professional development opportunities for both students and faculty as experimental and computational approaches advance.

Broadening the research and educational opportunities for faculty and students with limited resources will further advance the field of GDS, as well as science, technology, engineering and mathematics (STEM) at large, by increasing their participation and incorporating their viewpoints and knowledge⁹. To meet this need, we urge the genomics community to increase data findability, accessibility, interoperability and reusability (FAIRness), training and equipment access through resource sharing with faculty across a wide range of institutions to improve education and competitiveness in STEM¹³. In addition, funding is often a major hurdle to conducting research, especially at resource-limited institutions^{14,15}. Thus, the opportunities for faculty at these institutions to pursue independent research and for engaging students in genomic research are limited. Given this ongoing challenge, we encourage the genomics community to champion increased funding (for example, mini-sabbaticals, student internships and instrumentation mini-grants), and to promote collaborations with industry and private foundations to offset the expenses for faculty and students. These shared endeavors, emerging from authentic partnerships, will lay the groundwork for a better understanding of the dynamics of soil diversity, enhance genomics research and education, build trust and promote broader participation by increasing access to resources across the USA. For individuals interested in participating in the BioDIGS project, please find us at <https://biodigs.org>.

The BioDIGS Consortium*, Tristen Alberts¹, Claude F. Albritton², Rosa Alcazar³, Zainab Aljabri⁴, Maria Alvarez⁵, Anish Aradhey⁶, Mentewab Ayalew¹✉, Nareh Azizian¹✉, Yasmeen Balayah⁴, Destiny D. Ball², Efren Barragan⁵, Corey Beshoar⁶, Lyle Best⁹, Emily Biggane¹✉, Joseph Biggane¹⁰, Jesse Blick⁴, Myron Blosser⁶, Alex Kenneth Brown¹¹, Michael C. Campbell¹², Zoe Canizares¹³, Faith N. Chanhuwua⁵, Yu Chen¹⁴, Daniel R. Chin¹⁵, Kamal Chowdhury¹⁶, Tyler Collins¹⁷, Blair Compton¹⁸, Jefferson Da Silva¹⁹, Nia R. Davis⁷, Natalie DeCaro²⁰, Frida Delgadillo²¹, Youping Deng¹⁴, Joseph Duncan²², Arinzechukwu C. Egwu²³, Grace D. Ekalle¹⁹, Noha Elnawam²², Ray Enke¹⁸, Naomi Ewhe¹⁹, Marco A. Ferrel⁵, Janna Fierst¹✉, Grace Freymiller³, Karla Fuller¹✉, Lena Fulton-Wright⁶, Valeriya Gaysinskaya¹⁷, Torrence Gill¹¹, Ellie Gillespie⁴, Perla Gonzalez Moreno²⁶, Sara Goodwin²⁷, Natajha Graham⁷, Madeline E. Graham²⁸, Joseph L. Graves Jr.²⁹, Emily Grob³⁰, Rachael Gutierrez³, Aisha Hager⁴, Shazia Tabassum Hakim³¹, Aaliyah Harris²⁵, Ava M. Hoffman¹✉, Tobias Hoffmann⁷, Alani M. Horton²³, Allison Hughes¹⁰, Elizabeth M. Humphries³², Josh-Samuel Ikechi-Konkwo⁵, Aadil Ishtiaq¹³, Ryan Jackson¹⁸, Joshua Ronnie James¹¹, Kaitlan James³¹, Sydney A. Jamison², Armando Jimenez²¹, Rachel Johnson¹⁹, Abigail Kauffman⁶, Harkiran Kaur³, Kritika Kc¹⁹, Analyse Keeton⁴, Olivia E. Kelly⁵, Jennifer Kerr¹✉, Nataliya Kucher¹✉, Donna Lee Kuehu¹✉, Wendy A. Larson¹⁰, Joslynn Lee²⁰, Andrew Lee³³, Jeffrey T. Leek³², Danilo Lemaic¹✉, Lincoln E. Liburd II², Alan Fernando Lopez³, Mohammadamin Mahmanzar¹✉, Karwitha Mamae⁴, Raffi Manjikian¹³, Michael Marone²³, Katerin Marquez¹³, Amara Martinson¹, Senem Mavruk Eskipehlivan²⁷, Ashley Medrano¹³, Melanie Melendrez-Vallard¹✉, Robert Meller¹✉, Loyda B. Méndez³⁵, Miguel P. Mendez Gonzalez³⁶, Nicoll Mesquita²⁵, Concepcion Martinez Miller³⁷, Isam Mohd-Ibrahim¹⁴, Peter Mortensen¹, Stephen Mosher¹✉, Alketa Muja¹³, Nadia Nasrin²⁵, Masaki Nasu¹⁴, Matthew H. Nguyen¹⁷,

Ba Thong Nguyen¹⁴, Michele Nishiguchi¹✉, Lance M. O'Connor¹✉, Disomi Okie¹⁹, Tolulope Olowookoran¹⁹, Alex Ostrovsky¹⁷, Keyan Ozuna²¹, Asmita Pandey¹⁴, Shiv B. Patel², Gauri Paul¹✉, Shrikant Pawar¹✉, Andrea Pearson³⁴, Deborah Petrik¹✉, Jordan Platero¹✉, Carl Pontino³⁹, Arjun P. Pratap², Siddharth Pratap², Yujia Qin¹✉, Sudhir Kumar Rai¹⁴, Nistha Ray¹⁹, Ethan Repesh¹✉, Kristen Rhinehardt²⁹, Brennan Roche¹✉, Ariana Rodriguez¹⁹, Shriya Roy⁴⁰, Sourav Roy²¹, Alexa Sawa¹✉, Michael C. Schatz¹✉, Shurjo K. Sen¹✉, Randon Serikawa¹✉, Tyler Smith⁴³, Loraye Smith¹⁹, James Sniezek⁴⁴, Ryley D. Stewart⁵, Edu B. Suarez-Martinez⁴⁵, Joelle Taganna³², Frederick J. Tan¹⁷, Nikolaos Tsotakos¹✉, Nwanneka Udolisa¹⁹, Katherine Ulbricht¹✉, Tanner Veo¹, Jennifer Vessio¹⁷, Lia Walker¹¹, Oscar Wang⁶, Qingguo Wang¹✉, Robert Wappel²⁷, Kalynn Wesby¹⁶, Malachi Whitford³, Nicole Wild¹⁰, Xianfa Xie¹✉, Hua Yang¹✉, Sayumi York¹⁹ & Lindsay Zirkle¹⁸

¹United Tribes Technical College, Bismarck, ND, USA. ²Meharry Medical College, Nashville, TN, USA. ³Clovis Community College, Fresno, CA, USA. ⁴Anoka Ramsey Community College, Coon Rapids, MN, USA.

⁵El Paso Community College, El Paso, TX, USA. ⁶Harrisonburg High School, Harrisonburg, VA, USA. ⁷Spelman College, Atlanta, GA, USA.

⁸Stanford University, Stanford, CA, USA. ⁹Turtle Mountain Community College, Belcourt, ND, USA. ¹⁰University of Mary, Bismarck, ND, USA.

¹¹Chowan University, Murfreesboro, NC, USA. ¹²University of Southern California, Los Angeles, CA, USA. ¹³Hudson County Community

College, Jersey City, NJ, USA. ¹⁴University of Hawaii - Mānoa, Honolulu, HI, USA. ¹⁵University of Maryland, College Park, College Park, MD, USA.

¹⁶Claflin University, Orangeburg, SC, USA. ¹⁷Johns Hopkins University, Baltimore, MD, USA. ¹⁸James Madison University, Harrisonburg, VA, USA. ¹⁹Notre Dame of Maryland University, Baltimore, MD, USA.

²⁰Fort Lewis College, Durango, CO, USA. ²¹University of Texas at El Paso, El Paso, TX, USA. ²²Northern VA Community College-Alexandria, Annandale, VA, USA. ²³Virginia State University, Petersburg, VA, USA. ²⁴Florida International University, Miami, FL, USA. ²⁵Guttman Community College, New York, NY, USA.

²⁶University of California, Merced, Merced, CA, USA. ²⁷Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA. ²⁸Towson University, Towson, MD, USA. ²⁹North Carolina A&T State University, Greensboro, NC, USA. ³⁰Loyola University Maryland, Baltimore, MD, USA.

³¹Diné College, Tsaiile, AZ, USA. ³²Fred Hutchinson Cancer Center, Seattle, WA, USA. ³³Georgetown University, Washington, DC, USA.

³⁴Morehouse School of Medicine, Atlanta, GA, USA. ³⁵Universidad Ana G. Méndez - Carolina, Carolina, Puerto Rico. ³⁶University of Puerto Rico - Aguadilla, Aguadilla, Puerto Rico. ³⁷Dona Ana Community

College, Las Cruces, NM, USA. ³⁸Northeast State University of Oklahoma, Tahlequah, OK, USA. ³⁹University of Washington - Seattle, Seattle, WA, USA. ⁴⁰Citrus Valley High School, Redlands, CA, USA.

⁴¹College of the Desert, Palm Desert, CA, USA. ⁴²National Human Genome Research Institute, Bethesda, MD, USA. ⁴³College of

Southern Maryland, La Plata, MD, USA. ⁴⁴University of Maryland Baltimore County, Baltimore, MD, USA. ⁴⁵University of Puerto Rico -

Ponce, Ponce, Puerto Rico. ⁴⁶Penn State Harrisburg, Middletown, PA, USA. *A list of authors and their affiliations appears at the end of the paper.

✉ e-mail: mayalew@spelman.edu; ebiggane@uttc.edu;

karla.fuller@guttman.cuny.edu; ahoffma2@fredhutch.org;

michael.schatz@gmail.com; xxie@vus.edu

References

1. Anthony, M. A., Bender, S. F. & van der Heijden, M. G. A. *Proc. Natl Acad. Sci. USA* **120**, e2304663120 (2023).
2. Jackson, R. B. et al. *Annu. Rev. Ecol. Evol. Syst.* **48**, 419–445 (2017).
3. Banerjee, S. & van der Heijden, M. G. A. *Nat. Rev. Microbiol.* **21**, 6–20 (2023).
4. Forsberg, K. J. et al. *Science* **337**, 1107–1111 (2012).
5. Thompson, L. R. et al. *Nature* **551**, 457–463 (2017).
6. Labouyrie, M. et al. *Nat. Commun.* **14**, 3311 (2023).
7. Guerra, C. A. et al. *Nat. Commun.* **11**, 3870 (2020).
8. Ma, B. et al. *Nat. Commun.* **14**, 7318 (2023).
9. Genomic Data Science Community Network. *Genome Res.* **32**, 1231–1241 (2022).
10. Schatz, M. C. et al. *Cell Genom.* **2**, 100085 (2022).
11. Galaxy Community *Nucleic Acids Res.* **52**, W83–W94 (2024).
12. Benoit, G. et al. *Nat. Biotechnol.* **42**, 1378–1383 (2024).
13. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
14. Ginther, D. K. et al. *Science* **333**, 1015–1019 (2011).
15. Hoppe, T. A. et al. *Sci. Adv.* **5**, eaaw7238 (2019).

Acknowledgements

We would like to thank all the educators and students who contributed to the project, the full list can be found at <https://biodigs.org/#team>. We would like to thank K. Moffat,

W. R. McCombie, S. Goodwin, K. Doheny, D. Mohr, M. Van Oene, K. Liu, J. Kortlach, R. Dokos, J. Brayer, N. Lakey, W. Timp and J. Rocha for their helpful discussions and support. We would also like to thank the Galaxy Community, the JHU Genetic Resources Core Facility, the Cold Spring Harbor Laboratory Genome Center, the Virginia State University Center for Biotechnology, Genomics, and Bioinformatics (CeBiGeBi), Pacific Biosciences, Oxford Nanopore and CosmosID/cmBio for their support of the project. BioDIGS and GDSCN are funded through NHGRI contracts 75N92023P00302 and 75N92022P00232. This work is also supported, in part, by National Institutes of Health (NIH) award U24HG013013 and 5T34GM151403, and National Science Foundation (NSF) awards 1839895, 2011934, 2000157 and 2221924. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the NSF.

Competing interests

M.C.S. is on the scientific advisory board for CosmosID/cmBio. BioDIGS has received preferred pricing from PacBio and Oxford Nanopore. All other authors declare no competing interests.

Additional information

Peer review information *Nature Genetics* thanks Rayan Chikhi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

The BioDIGS Consortium

Mentewab Ayalew⁷, Emily Biggane¹, Karla Fuller²⁵, Ava M. Hoffman³², Michael C. Schatz¹⁷, Lindsay Zirkle¹⁸, Tristen Alberts¹, Claude F. Albritton², Rosa Alcazar³, Zainab Aljabri⁴, Maria Alvarez⁵, Anish Aradhey⁶, Nareh Azizian⁸, Yasmeen Balayah⁴, Destiny D. Ball², Efren Barragan⁵, Corey Beshoar⁶, Lyle Best⁹, Joseph Biggane¹⁰, Jesse Blick⁴, Myron Blosser⁶, Alex Kenneth Brown¹¹, Michael C. Campbell¹², Zoe Canizares¹³, Faith N. Chanhuhwa⁵, Yu Chen¹⁴, Daniel R. Chin¹⁵, Kamal Chowdhury¹⁶, Tyler Collins¹⁷, Blair Compton¹⁸, Jefferson Da Silva¹⁹, Nia R. Davis⁷, Natalie DeCaro²⁰, Frida Delgadillo²¹, Youping Deng¹⁴, Joceph Duncan²², Arinzechukwu C. Egwu²³, Grace D. Ekalle¹⁹, Noha Elnawam²², Ray Enke¹⁸, Naomi Ewhe¹⁹, Marco A. Ferrel⁵, Janna Fierst²⁴, Grace Freymiller³, Lena Fulton-Wright⁶, Valeriya Gaysinskaya¹⁷, Torrence Gill¹¹, Ellie Gillespie⁴, Perla Gonzalez Moreno²⁶, Sara Goodwin²⁷, Natajha Graham⁷, Madeline E. Graham²⁸, Joseph L. Graves Jr.²⁹, Emily Grob³⁰, Rachael Gutierrez³, Aisha Hager⁴, Shazia Tabassum Hakim³¹, Aaliyah Harris²⁵, Tobias Hoffmann⁷, Alani M. Horton²³, Allison Hughes¹⁰, Elizabeth M. Humphries³², Josh-Samuel Ikechi-Konkwo⁵, Aadil Ishtiaq¹³, Ryan Jackson¹⁸, Joshua Ronnie James¹¹, Kaitlan James³¹, Sydney A. Jamison², Armando Jimenez²¹, Rachel Johnson¹⁹, Abigail Kauffman⁶, Harkiran Kaur³, Kritika Kc¹⁹, Analyse Keeton⁴, Olivia E. Kelly⁵, Jennifer Kerr¹⁹, Nataliya Kucher¹⁷, Donna Lee Kuehu¹⁴, Wendy A. Larson¹⁰, Joslynn Lee²⁰, Andrew Lee³³, Jeffrey T. Leek³², Danilo Lemaic¹⁸, Lincoln E. Liburd II², Alan Fernando Lopez³, Mohammadamin Mahmanzar¹⁴, Karwitha Mamae⁴, Raffi Manjikian¹³, Michael Marone²³, Katerin Marquez¹³, Amara Martinson¹, Senem Mavruk Eskipehlivan²⁷, Ashley Medrano¹³, Melanie Melendrez-Vallard⁴, Robert Meller³⁴, Loyda B. Méndez³⁵, Miguel P. Mendez Gonzalez³⁶, Nicolle Mesquita²⁵, Concepcion Martinez Miller³⁷, Isam Mohd-Ibrahim¹⁴, Peter Mortensen¹, Stephen Mosher¹⁷, Alketa Muja¹³, Nadia Nasrin²⁵, Masaki Nasu¹⁴, Matthew H. Nguyen¹⁷, Ba Thong Nguyen¹⁴, Michele Nishiguchi²⁶, Lance M. O'Connor¹⁷, Disomi Okie¹⁹, Tolulope Olowookorun¹⁹, Alex Ostrovsky¹⁷, Keyan Ozuna²¹, Asmita Pandey¹⁴, Shiv B. Patel², Gauri Paul³, Shrikant Pawar¹⁶, Andrea Pearson³⁴, Deborah Petrik³⁸, Jordan Platero²⁰, Carl Pontino³⁹, Arjun P. Pratap², Siddharth Pratap², Yujia Qin¹⁴, Sudhir Kumar Rai¹⁴, Nisttha Ray¹⁹, Ethan Repesh¹⁸, Kristen Rhinehardt²⁹, Brennan Roche¹⁰, Ariana Rodriguez¹⁹, Shriya Roy⁴⁰, Sourav Roy²¹, Alexa Sawa⁴¹, Shurjo K. Sen⁴², Randon Serikawa³², Tyler Smith⁴³, Loraye Smith¹⁹, James Sniezek⁴⁴, Ryley D. Stewart⁵, Edu B. Suarez-Martinez⁴⁵, Joelle Taganna³², Frederick J. Tan¹⁷, Nikolaos Tsotakos⁴⁶, Nwanneka Uadolisa¹⁹, Katherine Ulbricht⁷, Tanner Veo¹, Jennifer Vessio¹⁷, Lia Walker¹¹, Qingguo Wang², Oscar Wang⁶, Robert Wappel²⁷, Kalynn Wesby¹⁶, Malachi Whitford³, Nicole Wild¹⁰, Hua Yang¹⁴, Sayumi York¹⁹ & Lindsay Zirkle¹⁸