

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Sanders, Sheri Anne

eRA COMMONS USER NAME (credential, e.g., agency login): SSANDERS93

POSITION TITLE: Associate Professor of the Practice, Bioinformatics; Associate Director of Bioinformatics, Genomics and Bioinformatics Core Facility

EDUCATION/TRAINING (*Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.*)

INSTITUTION AND LOCATION	DEGREE (if applicable)	Completion Date MM/YYYY	FIELD OF STUDY
Michigan State University, Lansing MI	B.S.	08/2006	Zoology
Michigan State University, Lansing MI	B.S.	08/2007	Environmental Science and Management
University of Texas Tyler, Tyler TX	M.S.	08/2010	Biology
University of Notre Dame, Notre Dame IN	certification	05/2016	Teaching in Biological Sciences
University of Notre Dame, Notre Dame IN	Ph.D.	05/2016	Biological Sciences
Indiana University, Bloomington, IN	certification	05/2021	Data Science

A. Personal Statement

During my six years working with the National Center for Genome Analysis Support, I successfully conceived, organized, and orchestrated several large projects for the team, the largest of which was building an education program. Leveraging my 20 years of experience in teaching university-level biologists, I developed domain-specific educational materials for the genomics community in the R, basic computing skills in linux-based HPC systems, and basic system administration. In addition to these basic skills, I also developed a streamlined version of a transcriptome assembly workflow based on our research collaborations. I led the effort on these materials being converted into a well-trafficked blog and multi-day in-person and online courses, managing the planning, instruction, evaluation, and reporting. This service grew to be an international offering, with several hundred students participating in courses every year and three dozen authors using the workflows in published papers. During this time, I also provided expert consultation on several research projects, three of which were significant enough to warrant co-authorship of resulting products.

My ability to plan, organize, and successfully manage both research and educational projects resulted in my conversion to a Co-PI within three years of joining the National Center for Genome Analysis Support team, and promotion to manager and PI within four. I leveraged my time within the Research Technologies department to further develop my technical skills, including a data science certification (Spring 2021) and system administration training (Summer 2018). These skills further enhance my ability to serve as a bridge between education, technology, and genomic research.

Now, as a Professor of the Practice in Bioinformatics, my **primary mission is to support and advance the practice of bioinformatics in research.** My role enables me to continue to bridge the gap between complex genomic data and its practical application, making it more accessible and understandable for a diverse array of researchers. I focus on developing coursework and community materials that emphasize transparency and ease of use for a diverse audience. My goal is to ensure that strategies behind the best practices are clearly understood, and also made the easiest initial option for researchers. This includes leveraging national hardware resources to remove computational roadblocks, developing user-friendly tools like genome browsers and web applications for hard to use tools, and ensuring software is available and easy

to use in the context of other software. Where needed, this also makes providing direct support in the form of collaboration and analysis of data more efficient, transparent, and replicable.

Continuing education to deepen understanding and keep pace with the rapid advancements in bioinformatics is critical. Focusing on the intersections of statistical concepts, such as the similarities between module identification in Weighted Gene Co-expression Network Analysis (WGCNA) and cell annotation in single-cell sequencing (scRNAseq) allows for faster knowledge transfer, which I translate into courses. Similar approaches are taken for extending experience in shared analytical frameworks, such as the vast applications of the same basic analysis of count data, or the similar nature of new technologies such as scRNAseq and Visium spatial transcriptomics. This approach naturally leads to combined analysis of novel datatypes and a deeper understanding of datasets in research projects. My commitment to these principles ensures that I not only contribute to the field of bioinformatics through research and development but also empower and develop the communities practice of bioinformatics.

B. Positions, Scientific Appointments, and Honors

2023-	Associate Director of Bioinformatics, Genomics and Bioinformatics Core Facility, University of Notre Dame
2023-	Associate Professor of the Practice, Bioinformatics, University of Notre Dame
2022-2023	Consultant, Kallista Consulting
2021-2022	Subaward Co-Principal Investigator, Genomics and Eco-evolution of Multi-Scale Symbioses (GEMS)
2020-2022	Manager, National Center for Genome Analysis Support (NCGAS)
2020-2022	Principal Investigator, National Center for Genome Analysis Support (NCGAS)
2019-2020	Co-Principal Investigator, National Center for Genome Analysis Support (NCGAS)
2018-2020	Principal Bioinformatic Analyst, National Center for Genome Analysis Support (NCGAS)
2016-2017	Bioinformatic Analyst, National Center for Genome Analysis Support (NCGAS)
2015-2016	Course Instructor, University of Notre Dame Computer Science & Engineering/Biology
2014	Fellow, Center for Aquatic Conservation, University of Notre Dame
2011-2015	Fellow, Arthur J. Schmitt Presidential Fellow in Science and Engineering, Arthur J. Schmitt Foundation
2010-2016	Research Assistant, University of Notre Dame, with Michael Pfrender
2010-2014	Graduate Teaching Assistant, University of Notre Dame
2010	Fellow, Global Linkages of Biology, the Environment, and Society (GLOBES, NSF IGERT)
2008-2010	Graduate Teaching Assistant, University of Texas, Tyler, with John Placyk III
2006-2008	Laboratory Technician, Michigan State University, with Sharleen Sakai

Other Experience and Professional Memberships

2024-	Association for Computing Machinery (ACM) Member
2021-	Sigma XI Full Member
2020	Sustaining Biological Infrastructure Training, ESA workshop for NSF ABI PIs
2019-2021	Director of Conservation, Westarctica Inc.
2018-2022	Lead Content Developer, National Center for Genome Analysis Support (NCGAS) Outreach
2017-	Domain Champion, Genomics and Field Stations, XSEDE/ACCESS
2017-2021	Lead writer, NCGAS Blog
2017-2022	Member, Organization of Biological Field Stations
2016-2020	Member, Society for the Study of Amphibians and Reptiles
2016-2019	Member, Society for the Study of Evolution
2013	Communication and Policy Training, GLOBES
2013	Student, Policy Training and Communication, GLOBES
2012	Student, Communication for Scientific and Technology Issues, GLOBES
2011-2012	President, Society of Schmitt Fellows, Arthur J. Schmitt Foundation

C. Contributions to Science

1. My direct research contributions have primarily centered around **RNA-seq based transcriptome assembly and analysis**, initially in polyploid systems. During my time at the National Center for Genome Analysis Support, my work focused on the application of *de novo* assembled transcriptomes in management applications and describe specific complications in assembling and analyzing homeologs. This work has led to

broader additional work in combined *de novo* transcriptome assembly and annotation in polyploid crop systems (e.g. Coffee, Peanut, Sweet Potato). As a logical extension, I also became involved in the assembly and analysis of pantranscriptomes in plants, specifically the cultivated polyploid potato in Slovenia. I also provided analysis support on several non-polyploid projects, largely through collaborations generated through the workshops based on the analysis framework we had developed for *de novo* assembly, annotation, and analysis of RNAseq data.

Sanders S. Conservation transcriptomics of Ambystomatid salamanders and their polyploid hybrids.

University of Notre Dame, Ph.D. dissertation. 2016. Available from:

<https://curate.nd.edu/show/x633dz03m7r>.

Yang T, Fang L, **Sanders S**, Jayathi S, Rajan G, Ppdicheti R, Thallapuram SK, Mockaitis K, Medina-Bolica F. Stillbenoid prenyltransferases define key steps in the diversification of peanut phytoalexins.

Journal of Biochemistry. 2017. Available from:

<http://www.jbc.org/content/early/2017/11/17/jbc.RA117.000564>.

Papudeshi B, Chafin T, **Sanders S**, Ganote C, Reshetnikov A, Sokolov S, Doak T, Pummil JF, Douglas MR, Douglas M. Genome and transcriptome analysis of fish tapeworm *Nippotaenia percotti* through scientific collaboration between research labs and national cyberinfrastructure. In: American Fisheries Society and The Wildlife Society 2019 Joint Annual Conference; 2019. Available from:

<https://afs.confex.com/afs/2019/meetingapp.cgi/Paper/39888>.

Petek M, Zagorscak M, Ramsak Z, **Sanders S**, Tseng E, Zouine M, Coll A, Gruden K. Cultivar-specific transcriptome and pan-transcriptome reconstruction of tetraploid potato. Scientific Data. 2020;7:249.

Available from: <https://www.nature.com/articles/s41597-020-00581-4>.

Lomheim HJ, Reyes Rodas L, Mulla L, Freeborn L, Sun DA, **Sanders SA**, Protas ME. Transcriptomic analysis of cave, surface, and hybrid samples of the isopod *Asellus aquaticus* and identification of chromosomal location of candidate genes for cave phenotype evolution. *EvoDevo*. 2023;14:9. Available from: <https://doi.org/10.1186/s13227-023-00213-z>.

2. More recently, I have shifted to developing support for the widening applications of RNA sequencing.

Transcriptomics has exploded in use and expanded to include new data-generating technologies (e.g. single-cell RNA sequencing and spatial transcriptomics) and consider new scopes of analysis (e.g. splice profiling of deep sequencing data versus broad coverage of whole microbiomes). Most of these projects are still in development. However, initial collaborations have yield both publications, workflow development, and course development as before.

Mansfield C, Tseng C, **Sanders S**, Custer TW, Custer CM, Matson CW. Genetic diversity comparison of tree swallow populations in the Great Lakes region using RNA-sequencing (Poster). In: SETAC North America 40th Annual Meeting; 2019 Nov 3-7; Toronto, ON, Canada.

Petek M, Zagorscak M, Ramsak Z, **Sanders S**, Tseng E, Zouine M, Coll A, Gruden K. Cultivar-specific transcriptome and pan-transcriptome reconstruction of tetraploid potato. Scientific Data. 2020;7:249.

Available from: <https://www.nature.com/articles/s41597-020-00581-4>.

Lomheim HJ, Reyes Rodas L, Mulla L, Freeborn L, Sun DA, **Sanders SA**, Protas ME. Transcriptomic analysis of cave, surface, and hybrid samples of the isopod *Asellus aquaticus* and identification of chromosomal location of candidate genes for cave phenotype evolution. *EvoDevo*. 2023;14:9. Available from: <https://doi.org/10.1186/s13227-023-00213-z>.

Shaikh M, Chao A, Lu X. Tackling Bone Metastasis in Prostate Cancer. 2024 Summer Undergraduate Research Symposium, University of Notre Dame. July 24, 2024, Notre Dame, Indiana. Available from: https://energy.nd.edu/assets/575366/summer_undergraduate_research_symposium_schedules_and_abstract_book_7242024.pdf.

Sanders SA. Topics in Bioinformatics: scRNAseq. University of Notre Dame; Fall Semester 2023.

Sanders SA. Topics in Bioinformatics: scRNAseq and spatial transcriptomics. University of Notre Dame; Fall Semester 2024.

Lomheim HJ, Reyes Rodas L, Price D, Sarbu SM, Bancila R, Freeborn L, **Sanders S**, Protas M.

Comparative embryology and transcriptomics of *Asellus infernus*, an isopod crustacean from sulfidic groundwater. *EvoDevo*. In revision.

3. Accelerating research by supporting computing infrastructure, which includes access to computing resources, access to common software, and support of common workflows. For six years, I installed and maintained software on Indiana University's compute clusters. I also worked with various groups to migrate tools to Jetstream cloud platform, which is a free to use cloud computing platform supported by NSF ACCESS. For the original Jetstream cloud, I produced tutorials on using these systems and a virtual machine library to support RStudio Server with Bioconductor, QIIME2, machine learning with sklearn, and Jbrowse/Tripal. In total, these images were used more than a thousand times. On Jetstream2, I have created and supported images to provide point-and-click, accessible analysis of count data (e.g. RNAseq, metagenomics, etc) for courses or research. I am also collaborating to build new genome browser platforms to integrate visualization of genomic data and point-and-click tools for crispr design, comparative genomics, etc. Additionally, I consult on a dozen of projects each year, helping researchers plan and troubleshoot workflows, originally through NCGAS but now through the Genomics and Bioinformatics Core Facility. Long- and short-term consulting on research projects ranging from transcriptomics to population genomics to sequencing troubleshooting, has led to my developing new tools for the community, such as genome browser add-ons (via Tripal software stack), web tutorials on various topics such as phylogenetic tree construction and third generation sequencing (via now retired NCGAS blog), and several streamlined workflows for common research analyses (such as transcriptome assembly, metagenomics, and variant calling).

Sanders S, Ganote C, Papudeshi B, Doak T. National Center for Genome Analysis Support (NCGAS) use and development of Tripal Genome Browsers on XSEDE's Jetstream. In: Plant and Animal Genomics 2018; 2018 Jan 13-17; San Diego, CA. Available from: <http://hdl.handle.net/2022/21898>.

Papudeshi B, **Sanders S**, Ganote C, Fischer J, Doak T. Bioinformatic analysis using Jetstream, a cloud computing environment (Poster). In: Plant and Animal Genomics 2018; 2018 Jan 13-17; San Diego, CA. Available from: <http://hdl.handle.net/2022/21900>.

Papudeshi B, **Sanders S**, Ganote C, Doak T. Compute resources available to the research community for microbiome analysis. In: Plant and Animal Genomics 2019; 2019 Jan 12-16; San Diego, CA. Available from: <http://hdl.handle.net/2022/22655>.

Sanders S, Papudeshi B, Ganote C, Doak T. NCGAS Makes Robust Transcriptome Assembly Even Easier with Added Features to an Accessible de novo Transcriptome Assembly Workflow. In: Plant and Animal Genomics 2019; 2019 Jan 12-16; San Diego, CA. Available from: <http://hdl.handle.net/2022/22653>.

Brooks EM, **Sanders SA**, Pfrender ME. freeCount: A Coding Free Framework for Guided Count Data Visualization and Analysis. In: Practice and Experience in Advanced Research Computing (PEARC '24); 2024 Jul 21-25; Providence, RI. ACM, New York, NY, USA. Available from: <https://doi.org/10.1145/3626203.3670605>.

4. While collaborating in research and building community tools, I have produced **scalable complementary educational materials**, in modular form, to help train the national community of biologists in both HPC and bioinformatic topics. The streamlined transcriptome analysis pipeline was the basis for our first three-day national-level workshop through NCGAS, and it included approximately half HPC topics (job scheduling, choosing HPC resources, data management and movement) and half genomic analysis topics (assembly, pathway analysis, annotation). This interdisciplinary approach is intended to increase researcher buy-in on topics that are directly useful but often skipped (HPC topics) while providing immediate reinforcement of the material in useful analyses (analysis topics). This workshop was been consistently over-subscribed and led to seven papers, two dissertations and 12 presentations by 39 authors in the first two years. In total, 82 people from 43 institutions in 27 states/territories were trained on how to do RNA data analysis as well as how to use research clusters to do their work. Upon moving to the University of Notre Dame, these training materials were updated and converted into a graduate course and also integrated into research experience courses. A total of over 300 students have had exposure to RNA sequencing from wet lab through enrichment analysis in under two years using these materials. Additional workshop construction has followed the same general workflow (Sanders et al, 2019), resulting in another national workshop on metagenomics that has also since become a graduate course, and an online R course for biologists that has scaled to over 400 people in under two years.

Sanders S. Teaching HPC and genomics side-by-side. ScienceNode. 2018 Jul 24. Available from: <https://sciencenode.org/feature/Teaching%20HPC%20and%20Genomics%20Side-by-Side.php>. Accessed 2019 Dec 4.

Sanders S, Ganote C, Papudeshi B, Stewart C, Doak T. Summary of the National Center for Genome Analysis Support (NCGAS) 2018-2019 de Novo Transcriptome Workflow and Workshops. Indiana University, Bloomington, IN. PTI Technical Report. 2019. Available from: <https://scholarworks.iu.edu/dspace/handle/2022/24887>.

Sanders S. Introduction to R for Biologists. E-book. National Center for Genome Analysis Support; 2019. Available from: https://ncgas.org/training/r_textbook_full.pdf.

Sanders S, Ganote C, Papudeshi B, Stewart C, Doak T. Summary Report on Scaling the Introduction to R for Biologists Workshop by National Center for Genome Analysis Support (NCGAS) to a Massive Open Online Course (MOOC). Indiana University, Bloomington, IN. PTI Technical Report. 2019. Available from: <http://hdl.handle.net/2022/24888>.

Sanders SA. Microbial Genome Assembly and Analysis. Indiana University and University of Illinois Urbana-Champaign (GEMS); Spring Semester 2021.

Sanders SA. Applied Bioinformatics. University of Notre Dame; Spring Semester 2023.

Brooks EM, **Sanders SA**, Pfreder ME. freeCount: A Coding Free Framework for Guided Count Data Visualization and Analysis. In: Practice and Experience in Advanced Research Computing (PEARC '24); 2024 Jul 21-25; Providence, RI. ACM, New York, NY, USA. Available from: <https://doi.org/10.1145/3626203.3670605>.

5. As data in biology has expanded, increasing use of machine learning and AI to analyze data has followed. I have been **studying, using, and training researchers in machine learning** applications to biology for several years now. I completed a graduate certificate in Data Science from Indiana University, which ended with a capstone project in biomedical science. I worked with researchers in the Department of Oncology to streamline data acquisition and cleaning, as well as assess their current model of radiation induced lymphopenia (RIL). We were able to predict absolute lymphocyte counts for full treatment courses given three initial time points. This was developed to help identification of patient at risk of RIL. Research was discontinued when faculty changed institutions and data access created limitations. I later directly mentored seven students through the Jetstream REU program, the Center of Excellence for Women and Technology (CEW&T) REU, the Bepko Internship at Indiana University. These students worked with me in developing computing and cloud skills and have worked toward developing automated raspberry pi based bioacoustic recording devices paired with frog call analysis with machine learning. Students learned machine learning methods (random forests), as well as deep learning methods (CNN, RNN). These projects led to several conference presentations, awards, and one student was given a position to extend her work with the bioacoustics workflow to bat surveys in Panama for Summer 2020 (canceled due to pandemic). Upon moving to the University of Notre Dame, I have developed an applied course in machine learning applications in biology, as well as support for network analysis outside of machine learning.

Anderson J, Slayton T, Guido E, Doak T, **Sanders S**. Harvesting Field Station Data; Raspberry Pi Sensors to Jetstream Databases. In: PEARC18 Conference; 2018 Jul 22-26; Pittsburgh, PA. Available from: https://pearc18.conference-program.com/presentation/?id=stu_pos135&sess=sess148.

Foran E, Underwood T, Suggs E, **Sanders S**, Snapp-Childs W. Automatic recognition of frog vocalizations using Jetstream. In: PEARC19 Conference; 2019 Jul 28-Aug 1; Chicago, IL. Available from: <https://scholarworks.iu.edu/dspace/handle/2022/23260>.

Foran E, Underwood TA, Snapp-Childs W, **Sanders S**. Automatic capture and classification of frog calls (Poster). In: Supercomputing 2020; 2020 Nov 9-19; Online. Available from: <https://hdl.handle.net/2022/25760>.

Cai JX, Weathers JG, Leffler H, Ganapaneni S, Papudeshi B, **Sanders S**, Doak TG. Navigating the Sequence Read Archive to identify crAssphage, an ubiquitous inhabitant of the human microbiome (Poster). In: Holland Summer Science Research Program Poster Session; 2019 Aug 9; Bloomington, IN.

Sanders SA. Topics in Bioinformatics: Critiquing Machine Learning Applications in Biology. University of Notre Dame; Fall Semester 2023.

Complete List of Published Work:

<https://www.ncbi.nlm.nih.gov/myncbi/1NOYgCGZB9FwDy/bibliography/public/>