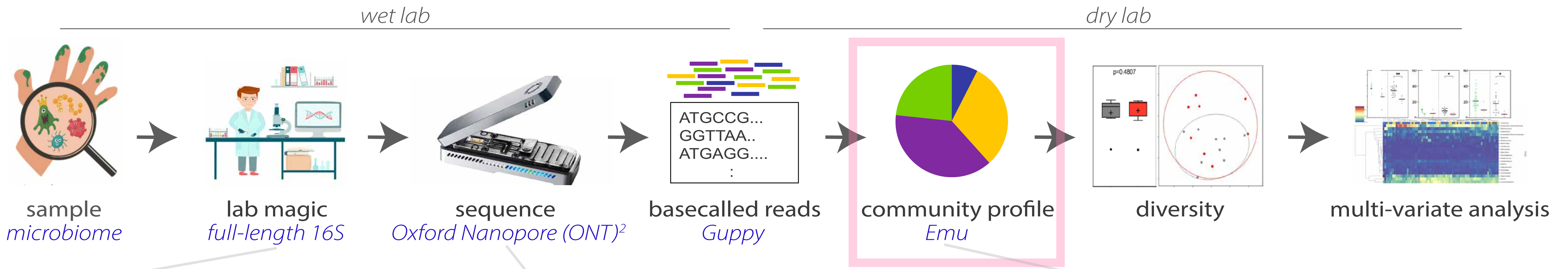
Emu: microbiome profiling software for 16S rRNA ONT reads

Kristen D. Curry[1], Qi Wang[2], Michael G. Nute[1], Alona Tyshaieva[3], Elizabeth Reeves[1], Sirena Soriano[4], Qinglong Wu[5], Enid Graeber[3], Patrick Finzer[3], Werner Mendling[6], Tor Savidge[5], Sonia Villapol[4], Alexander Dilthey[3,*], Todd J. Treangen[1,*]

[1]Rice University, Department of Computer Science, Houston, TX, USA
[2]Rice University, Department of Systems, Synthetic and Physical Biology Science, Houston, TX, USA
[3]Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University of Düsseldorf, Düsseldorf, Germany
[4]Houston Methodist Research Institute, Center for Neuroregeneration, Houston, TX, USA
[5]Baylor College of Medicine, Department of Pathology and Immunology, Houston, TX, USA
[6]German Center for Infections in Gynaecology and Obstetrics at Helios University Clinic Wuppertal, Wuppertal,Germany
*These authors contributed equally

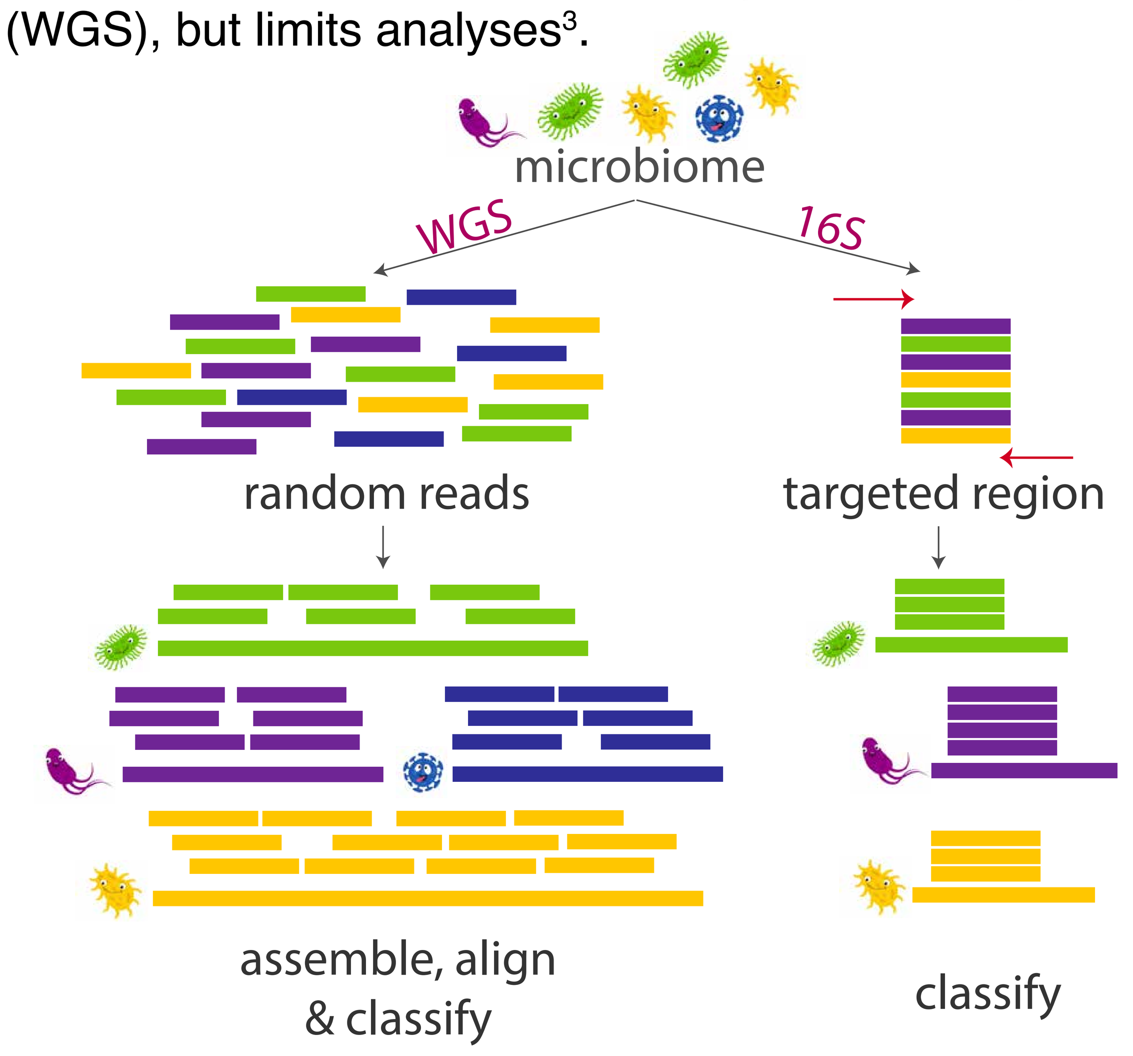
Motivation: democratized platform for species-level microbiome analyses

Fig 1: The ONT MinION[1] device can be used for inexpensive species-level microbiome analyses from full-length 16S rRNA sequences; however the software for generating community profiles was missing. We developed Emu to fill this void.



WGS vs 16S reads

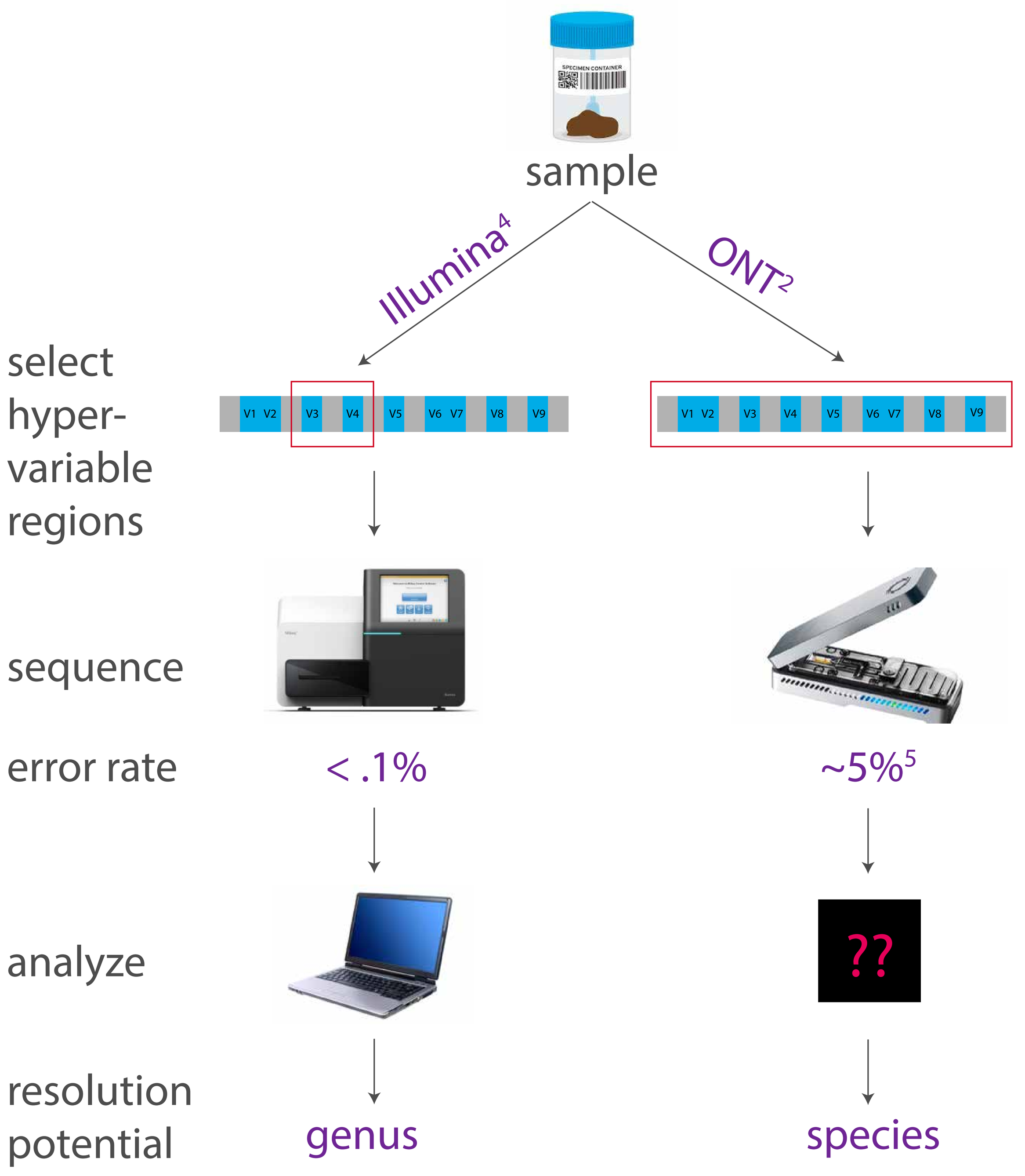
Fig 2: 16S rRNA targeted amplicon sequencing makes for a simpler approach over whole-genome shotgun (WGS), but limits analyses[3].



	WGS	16S
kingdoms	all	bacteria
composition	definite	relative
seq. depth	high	shallow
computation	complex	direct
cost	$$$	$
accessibility	external	in-house

Short vs long 16S reads

Fig 3: Short-read 16S rRNA analysis is limited to genus-level resolution due to the short length of each reads; however sequencing the full 16S gene has the potential for species-level results.



Emu method

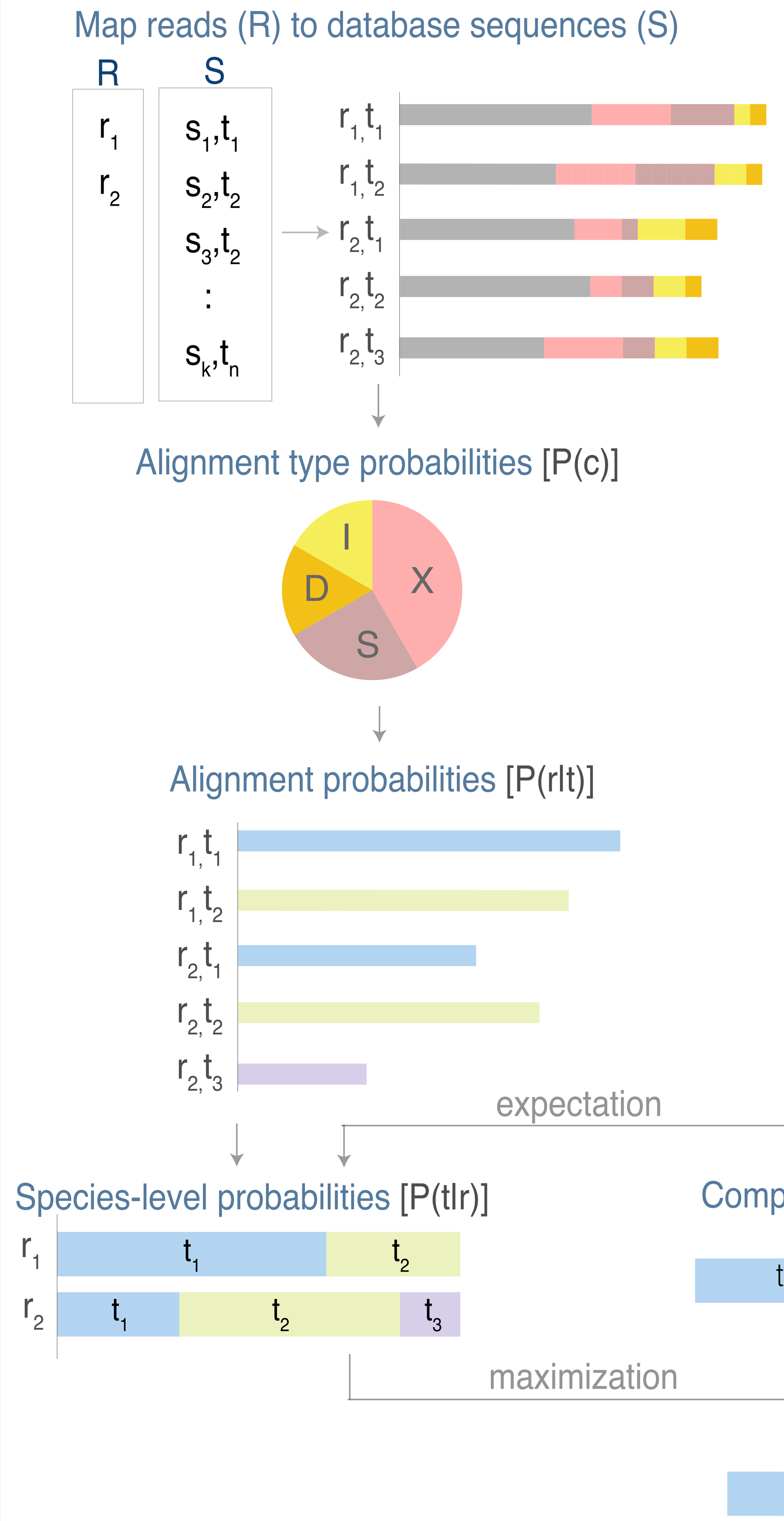
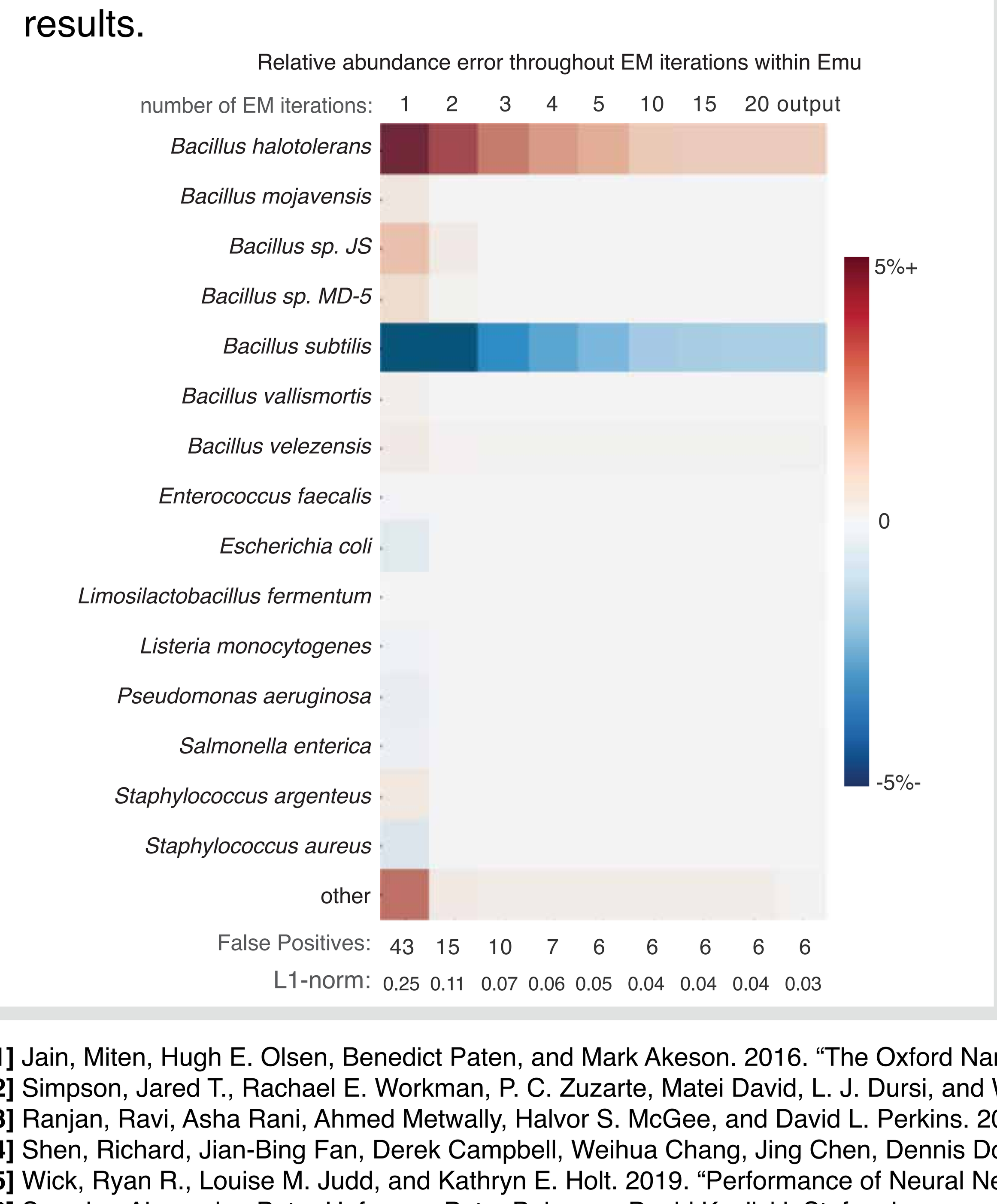


Fig 4: Emu corrects for ONT sequencing error through an expectation-maximization (EM) algorithm. Here, read classification are influenced by the community profile; when it is difficult to distinguish between species classifications for a given read, species that are in higher abundances in the sample are given a higher likelihood.

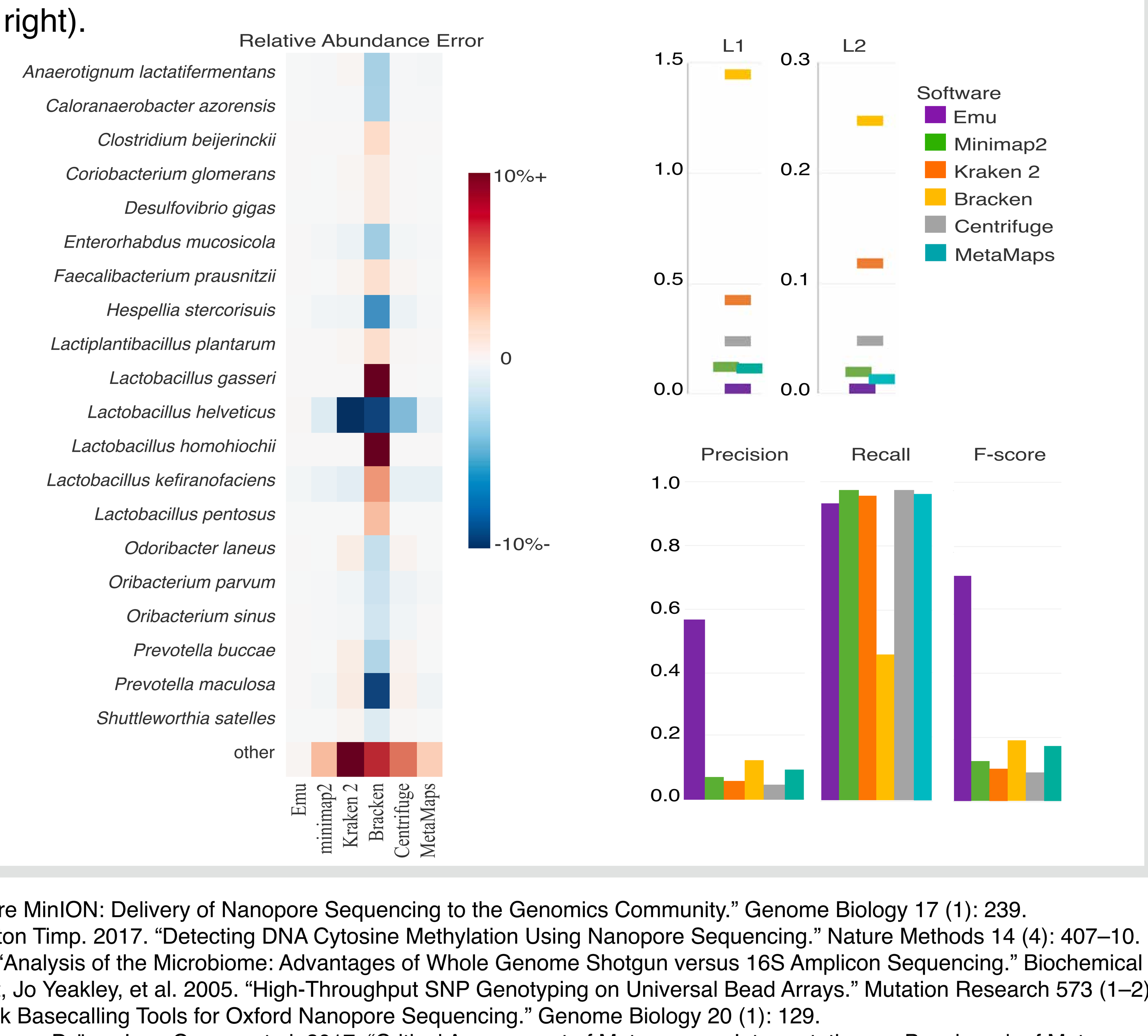
Error correction in Emu

Fig 5: Species-level error (overestimate in red; underestimate in blue) for test data set throughout EM iterations within Emu. Furthest left column is synonymous to classification without error-correction; furthest right is Emu results.



Evaluation of approaches

Fig 6: Performance of 6 different software on generating a species-level community profile from simulated full-length 16S rRNA sequence data of 350+ species mimicking a mouse gut microbiome[6]. Heat map of relative abundance error (left); L1- and L2-norm (top right); detection accuracy (bottom right).



Funding

NIH grant R21NS106640 from National Institute for Neurological Disorders and Stroke (NINDS), NIH grant 1P01AI152999-01 supported by National Institute of Allergy and Infectious Diseases (NIAID), Jürgen Manchot Foundation, Deutsche Forschungsgemeinschaft (DFG) award #428994620, fellowship from the National Library of Medicine Training Program in Biomedical Informatics and Data Science (T15LM007093, PI: Kavraki), Rice University Institute of Bioscience and Bioengineering Travel Award, Ken Kennedy Institute Computational Science & Engineering Recruiting Fellowship

Install

https://gitlab.com/treangenlab/emu

Emu
Project ID: 19618062

Publication

https://www.nature.com/articles/s41592-022-01520-4

Article | Published: 30 June 2022

Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data

[1] Jain, Miten, Hugh E. Olsen, Benedict Paten, and Mark Akeson. 2016. "The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community." Genome Biology 17 (1): 239.
[2] Simpson, Jared T., Rachael E. Workman, P. C. Zuzarte, Matei David, L. J. Dursi, and Winston Timp. 2017. "Detecting DNA Cytosine Methylation Using Nanopore Sequencing." Nature Methods 14 (4): 407–10.
[3] Ranjan, Ravi, Asha Rani, Ahmed Metwally, Halvor S. McGee, and David L. Perkins. 2016. "Analysis of the Microbiome: Advantages of Whole Genome Shotgun versus 16S Amplicon Sequencing." Biochemical and Biophysical Research Communications 469 (4): 967–77.
[4] Shen, Richard, Jian-Bing Fan, Derek Campbell, Weihua Chang, Jing Chen, Dennis Doucet, Jo Yeakley, et al. 2005. "High-Throughput SNP Genotyping on Universal Bead Array." Mutation Research 573 (1–2): 70–82.
[5] Wick, Ryan R., Louise M. Judd, and Kathryn E. Holt. 2019. "Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing." Genome Biology 20 (1): 129.
[6] Sczyrba, Alexander, Peter Hofmann, Peter Belmann, David Koslicki, Stefan Janssen, Johannes Dröge, Ivan Gregor, et al. 2017. "Critical Assessment of Metagenome Interpretation—a Benchmark of Metagenomics Software." Nature Methods 14 (11): 1063–71.