SPAAM2 - References and Resources Bibliography

Collection of article references and resources referred to during SPAAM2

Session 1

Comparative analysis between soil extraction (and lib. prep) methods

Armbrecht, L., Herrando-Pérez, S., Eisenhofer, R., Hallegraeff, G. M., Bolch, C. J. S., & Cooper, A. (2020). An optimized method for the extraction of ancient eukaryote DNA from marine sediments. *Molecular Ecology Resources*, *20*(4), 906–919. https://doi.org/10.1111/1755-0998.13162

Ligation biases in inline barcodes

Brealey, J. C., Leitão, H. G., van der Valk, T., Xu, W., Bougiouri, K., Dalén, L., & Guschanski, K. (2020). **Dental calculus as a tool to study the evolution of the mammalian oral microbiome**. Molecular Biology and Evolution. https://doi.org/10.1093/molbev/msaa135

Session 2

Variation Graphs

Martiniano, R., Garrison, E., Jones, E. R., Manica, A., & Durbin, R. (2020). **Removing reference bias and improving indel calling in ancient DNA data analysis by mapping to a sequence variation graph**. Genome Biology, 21(1), 250. https://doi.org/10.1186/s13059-020-02160-7

Co-infections and challenges in genotyping in strain separation

Kay, G. L., Sergeant, M. J., Zhou, Z., Chan, J. Z.-M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M., Loman, N. J., Achtman, M., Donoghue, H. D., & Pallen, M. J. (2015). **Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe**. Nature Communications, 6, 6717. https://doi.org/10.1038/ncomms7717

Sabin, S., Herbig, A., Vågene, Å. J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., & Bos, K. I. (2020). **A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex**. *Genome Biology*, *21*(1), 201. https://doi.org/10.1186/s13059-020-02112-1

Genotypers

Li, Heng. **Htsbox** (majority consensus caller) (Github: https://github.com/lh3/htsbox)

Garrison, E., & Marth, G. (2012). **Haplotype-based variant detection from short-read sequencing**. In *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1207.3907 (Github: https://github.com/ekg/freebayes)

Zhou, B., Wen, S., Wang, L., Jin, L., Li, H., & Zhang, H. (2017). **AntCaller: an accurate variant caller incorporating ancient DNA damage**. *Molecular Genetics and Genomics: MGG*, *292*(6), 1419–1430. https://doi.org/10.1007/s00438-017-1358-5
(Github: https://github.com/BoyanZhou/AntCaller)

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). **A universal SNP and small-indel variant caller using deep neural networks**. Nature Biotechnology, 36(10), 983–987. https://doi.org/10.1038/nbt.4235 (Github https://github.com/google/deepvariant)

SNP Caller comparison

Bush, S. J., Foster, D., Eyre, D. W., Clark, E. L., De Maio, N., Shaw, L. P., Stoesser, N., Peto, T. E. A., Crook, D. W., & Walker, A. S. (2020). **Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism-calling pipelines.** *GigaScience*, *9*(2). https://doi.org/10.1093/gigascience/giaa007

SNP Calling/Filtering/Table generation

Herbig, A. **MultiVCFAnalyzer** (Github: https://github.com/alexherbig/MultiVCFAnalyzer)

Seitz, A. **MUSIAL** (Github: https://github.com/Integrative-Transcriptomics/MUSIAL)

Seeman, T. **Snippy** (Github: https://github.com/tseemann/snippy)

Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). **SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments**. Microbial Genomics, 2(4), e000056. https://doi.org/10.1099/mgen.0.000056 (Github: https://github.com/sanger-pathogens/snp-sites)

Namouchi, A. **snpToolkit** (Github: https://github.com/Amine-Namouchi/snpToolkit)

Strain delineation

Van Rossum, T., Ferretti, P., Maistrenko, O. M., & Bork, P. (2020). **Diversity within species: interpreting strains in microbiomes.** *Nature Reviews. Microbiology*. https://doi.org/10.1038/s41579-020-0368-1

Pan genome reconstruction

Perrin, A., & Rocha, E. P. C. (2020). **PanACoTA: A modular tool for massive microbial comparative genomics.** In *bioRxiv* (p. 2020.09.11.293472). https://doi.org/10.1101/2020.09.11.293472

Recombination Detection

Didelot, X., & Wilson, D. J. (2015). **ClonalFrameML: efficient inference of recombination in whole bacterial genomes.** *PLoS Computational Biology*, *11*(2), e1004041. https://doi.org/10.1371/journal.pcbi.1004041

Lin, M., & Kussell, E. (2019). **Inferring bacterial recombination rates from large-scale sequencing datasets.** *Nature Methods*, *16*(2), 199–204. https://doi.org/10.1038/s41592-018-0293-7

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., Parkhill, J., & Harris, S. R. (2015). **Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.** *Nucleic Acids Research*, *43*(3), e15. https://doi.org/10.1093/nar/gku1196

Differential genome preservation, characteristics and effect on genotyping

Mann, A. E., Sabin, S., Ziesemer, K., Vågene, Å. J., Schroeder, H., Ozga, A. T., Sankaranarayanan, K., Hofman, C. A., Fellows Yates, J. A., Salazar-García, D. C., Frohlich, B., Aldenderfer, M., Hoogland, M., Read, C., Milner, G. R., Stone, A. C., Lewis, C. M., Jr, Krause, J., Hofman, C., … Warinner, C. (2018). **Differential preservation of endogenous human and microbial DNA in dental calculus and dentin.** *Scientific Reports*, *8*(1), 9822. https://doi.org/10.1038/s41598-018-28091-9

Session 3

Generating reference datasets (e.g. soil)

Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R.,

Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2019). **MGnify: the microbiome analysis resource in 2020.** *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkz1035

Bradley, P., den Bakker, H. C., Rocha, E. P. C., McVean, G., & Iqbal, Z. (2019). **Ultrafast search of all deposited bacterial and viral genomic data.** *Nature Biotechnology*, *37*(2), 152–159. https://doi.org/10.1038/s41587-018-0010-1

Alternatives to damage patterns (when exceptional preservation damage, but still truly old)?

Roller, M., Lucić, V., Nagy, I., Perica, T., & Vlahovicek, K. (2013). **Environmental shaping of codon usage and functional adaptation across microbial communities.** *Nucleic Acids Research*, *41*(19), 8842–8852. https://doi.org/10.1093/nar/gkt673

Skoglund, P. **PMDtools** (Github: https://github.com/pontussk/PMDtools; `--platypus`)

Adding additional confounding factors to authentication cut offs between samples E.g. batch effects

Morton, J. T., Marotz, C., Washburne, A., Silverman, J., Zaramela, L. S., Edlund, A., Zengler, K., & Knight, R. (2019). **Establishing microbial composition measurement standards with reference frames.** *Nature Communications*, *10*(1), 2719. https://doi.org/10.1038/s41467-019-10656-5

Zhang, Y., Jenkins, D. F., Manimaran, S., & Johnson, W. E. (2018). **Alternative empirical Bayes models for adjusting for batch effects in genomic studies.** *BMC Bioinformatics*, *19*(1), 262. https://doi.org/10.1186/s12859-018-2263-6

Tran, H. T. N., Ang, K. S., Chevrier, M., Zhang, X., Lee, N. Y. S., Goh, M., & Chen, J. (2020). **A benchmark of batch-effect correction methods for single-cell RNA sequencing data**. *Genome Biology*, *21*(1), 12. https://doi.org/10.1186/s13059-019-1850-9

Wang, Y., & LêCao, K.-A. (2019). **Managing batch effects in microbiome data**. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbz105 (Analysis Vignette: https://evayiwenwang.github.io/Managing_batch_effects/)

Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., Beghini, F., Malik, F., Ramos, M., Dowd, J. B., Huttenhower, C., Morgan, M., Segata, N., & Waldron, L. (2017). **Accessible, curated metagenomic data through ExperimentHub.** Nature Methods, 14(11), 1023–1024. https://doi.org/10.1038/nmeth.4468 (curatedMetagenomicData: https://bioconductor.org/packages/release/data/experiment/html/curatedMetagenomicData.html)

Session 4

Fort Lauderdale for Genomic data

Kaye, J., Heeney, C., Hawkins, N., de Vries, J., & Boddington, P. (2009). **Data sharing in genomics--re-shaping scientific practice.** *Nature Reviews. Genetics*, *10*(5), 331–335. https://doi.org/10.1038/nrg2573

Radiocarbon Dating Reporting Conventions

Millard, A. R. (2014). **Conventions for Reporting Radiocarbon Determinations**. *Radiocarbon*, *56*(2), 555–559. https://journals.uair.arizona.edu/index.php/radiocarbon/article/view/17455

(also at: https://c14.arch.ox.ac.uk/calibration.html#conventions_reporting)

CAMI challenge (benchmarking datasets for modern microbiome work)

https://data.cami-challenge.org/

LEMMI Challenge (similar)

https://lemmi.ezlab.org/#/ (with post from Maxime Borry asking about aDNA datasets: https://gitlab.com/ezlab/lemmi/-/issues/5)

Lab Benchmarking Challenge (XMP)

Tighe, S., Afshinnekoo, E., Rock, T. M., McGrath, K., Alexander, N., McIntyre, A., Ahsanuddin, S., Bezdan, D., Green, S. J., Joye, S., Stewart Johnson, S., Baldwin, D. A., Bivens, N., Ajami, N., Carmical, J. R., Herriott, I. C., Colwell, R., Donia, M., Foox, J., … Mason, C. E. (2017). **Genomic Methods and Microbiological Technologies for Profiling Novel and Extreme Environments for the Extreme Microbiome Project (XMP)**. *Journal of Biomolecular Techniques: JBT*, *28*(1), 31–39. https://doi.org/10.7171/jbt.17-2801-004

Problems with Rarefaction of Taxonomic Profiles

McMurdie, P. J., & Holmes, S. (2014). **Waste not, want not: why rarefying microbiome data is inadmissible.** *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). **Microbiome Datasets Are Compositional: And This Is Not Optional.** *Frontiers in Microbiology*, *8*, 2224.

https://doi.org/10.3389/fmicb.2017.02224

'Data Brokers' for help in uploading data

https://www.gfbio.org/ (Possibly German/EU only)

Example Reporting Standards for Marine Biology

Ten Hoopen, P., Pesant, S., Kottmann, R., Kopf, A., Bicak, M., Claus, S., Deneudt, K., Borremans, C., Thijsse, P., Dekeyzer, S., Schaap, D. M., Bowler, C., Glöckner, F. O., & Cochrane, G. (2015). **Marine microbial biodiversity, bioinformatics and biotechnology (M2B3) data reporting and service standards.** *Standards in Genomic Sciences*, *10*, 20. https://doi.org/10.1186/s40793-015-0001-5

(Parodies) People reusing data are 'research parasites'

https://researchparasite.com/

Duvallet, C. (2020). **Data detectives, self-love, and humility: a research parasite's perspective.** *GigaScience*, *9*(1). https://doi.org/10.1093/gigascience/giz148

Greene, C. S., Garmire, L. X., Gilbert, J. A., Ritchie, M. D., & Hunter, L. E. (2017). **Celebrating parasites [Review of *Celebrating parasites*].** *Nature Genetics*, *49*(4), 483–484. https://doi.org/10.1038/ng.3830

What is considered 'authorship' - should we ask original uploaders when publishing re-used data?

http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html

"Examples of activities that alone (without other contributions) do not qualify a contributor for authorship are acquisition of funding; general supervision of a research group or general administrative support; and writing assistance, technical editing, language editing, and proofreading. Those whose contributions do not justify authorship may be acknowledged individually or together as a group under a single heading (e.g. "Clinical Investigators" or "Participating Investigators"), and their contributions should be specified (e.g., "served as scientific advisors," "critically reviewed the study proposal," "**collected data,**" "provided and cared for study patients", "participated in writing or technical editing of the manuscript")."

CREDIT: https://casrai.org/credit/

PEERJ: https://peerj.com/about/policies-and-procedures/#authorship-criteria

Session 5

How does one deal with novel unknown taxa identified from assembly

Tutorial example: http://merenlab.org/2016/11/08/pangenomics-v2/
Modern DNA Workshop example: https://pagesperso.univ-brest.fr/~maignien/ebame5.html
Blog post: http://merenlab.org/2020/07/01/dark-side/

Deep-learning alternative to k-mers for contig binning (variational autoencoders)

Nissen, J. N., Sønderby, C. K., Armenteros, J. J. A., Grønbech, C. H., Nielsen, H. B., Petersen, T. N., Winther, O., & Rasmussen, S. (2018). **Binning microbial genomes using deep learning** (p. 490078). https://doi.org/10.1101/490078

Protein assembly > metagenomic assembly for function

Steinegger, M., Mirdita, M., & Söding, J. (2019). **Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold**. Nature Methods, 16(7), 603–606. https://doi.org/10.1038/s41592-019-0437-4

Steinegger, M., Seidel, A., & Söding, J. (2020, September 16). **Protein-guided nucleotide viral genome assembly for huge metagenomic datasets**. https://doi.org/10.7490/f1000research.1118321.1

Established High-throughput Modern Assembly Methods

Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., Beghini, F., Manghi, P., Tett, A., Ghensi, P., Collado, M. C., Rice, B. L., DuLong, C., Morgan, X. C., Golden, C. D., Quince, C., Huttenhower, C., & Segata, N. (2019). **Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle.** *Cell*, *0*(0), 649–662.E20. https://doi.org/10.1016/j.cell.2019.01.001

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., De Filippis, F., Magnabosco, C., Bonneau, R., Lusingu, J., Amuasi, J., Reinhard, K., Rattei, T., Boulund, F., Engstrand, L., … Segata, N. (2019). **The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations**. *Cell Host & Microbe*, *26*(5), 666–679.e7. https://doi.org/10.1016/j.chom.2019.08.018

Is coverage variation on mapped genomes/assemblies bad?

http://merenlab.org/2016/12/14/coverage-variation/

Read extension to gain extra bits of information (Super-reads)

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). **The MaSuRCA genome assembler**. Bioinformatics , 29(21), 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

All about k-mers, potential ideas that can be applied on ancient metagenomes (i.e. filtering unique k-mers...)

https://khmer-recipes.readthedocs.io/en/latest/

Examples of successful (well published) ancient de-novo assemblies

Brealey, J. C., Leitão, H. G., van der Valk, T., Xu, W., Bougiouri, K., Dalén, L., & Guschanski, K. (2020). **Dental calculus as a tool to study the evolution of the mammalian oral microbiome.** *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msaa135

Luhmann, N., Doerr, D., & Chauve, C. (2017). **Comparative scaffolding and gap filling of ancient bacterial genomes applied to two ancient Yersinia pestis genomes.** *Microbial Genomics*, 3(9), e000123. https://doi.org/10.1099/mgen.0.000123

Tett, A., Huang, K. D., Asnicar, F., Fehlner-Peach, H., Pasolli, E., Karcher, N., Armanini, F., Manghi, P., Bonham, K., Zolfo, M., De Filippis, F., Magnabosco, C., Bonneau, R., Lusingu, J., Amuasi, J., Reinhard, K., Rattei, T., Boulund, F., Engstrand, L., … Segata, N. (2019). **The Prevotella copri Complex Comprises Four Distinct Clades Underrepresented in Westernized Populations**. *Cell Host & Microbe*, *26*(5), 666–679.e7. https://doi.org/10.1016/j.chom.2019.08.018

Krause-Kyora, B., Susat, J., Key, F. M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., Franzenburg, S., Heyne, H. O., Meier, T., Lösch, S., Meller, H., Friederich, S., Nicklisch, N., Alt, K. W., Schreiber, S., Tholey, A., … Krause, J. (2018). **Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B.** eLife, 7. https://doi.org/10.7554/eLife.36666

Martin, M. D., Cappellini, E., Samaniego, J. A., Zepeda, M. L., Campos, P. F., Seguin-Orlando, A., Wales, N., Orlando, L., Ho, S. Y. W., Dietrich, F. S., Mieczkowski, P. A., Heitman, J., Willerslev, E., Krogh, A., Ristaino, J. B., & Gilbert, M. T. P. (2013). **Reconstructing genome evolution in historic samples of the Irish potato famine pathogen.** Nature Communications, 4, 2172. https://doi.org/10.1038/ncomms3172

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., Langridge, G., Fotakis, A. K., Nair, S., Stenøien, H. K., Hamre, S. S., Casjens, S., Christophersen, A., Quince, C., Thomson, N. R., Weill, F.-X., Ho, S. Y. W., Gilbert, M. T. P., & Achtman, M. (2018). **Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the**

Invasive Para C Lineage for Millennia. *Current Biology: CB*, *28*(15), 2420–2428.e10. https://doi.org/10.1016/j.cub.2018.05.058

Lammers, Y., Heintzman, P. D., & Alsos, I. G. (2020). **Environmental palaeogenomic reconstruction of an Ice Age algal population** *bioRxiv* (p. 2020.04.10.035535). https://doi.org/10.1101/2020.04.10.035535

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K. R., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2020). **Hybridization capture of larch (Larix Mill) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forests** *bioRxiv* (p. 2020.01.06.896068). https://doi.org/10.1101/2020.01.06.896068

Session 6

Metagenomics Reporting Guidelines (Clinical)

Bharucha, T., Oeser, C., Balloux, F., Brown, J. R., Carbo, E. C., Charlett, A., Chiu, C. Y., Claas, E. C. J., de Goffau, M. C., de Vries, J. J. C., Eloit, M., Hopkins, S., Huggett, J. F., MacCannell, D., Morfopoulou, S., Nath, A., O'Sullivan, D. M., Reoma, L. B., Shaw, L. P., … Field, N. (2020). **STROBE-metagenomics: a STROBE extension statement to guide the reporting of metagenomics studies.** The Lancet Infectious Diseases. https://doi.org/10.1016/S1473-3099(20)30199-7

Inter-Lab Comparison/Competition

Doyle, R. M., O'Sullivan, D. M., Aller, S. D., Bruchmann, S., Clark, T., Coello Pelegrin, A., Cormican, M., Diez Benavente, E., Ellington, M. J., McGrath, E., Motro, Y., Phuong Thuy Nguyen, T., Phelan, J., Shaw, L. P., Stabler, R. A., van Belkum, A., van Dorp, L., Woodford, N., Moran-Gilad, J., … Harris, K. A. (2020). **Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study.** Microbial Genomics, 6(2). https://doi.org/10.1099/mgen.0.000335