Metagenomics - Tools and other Points of Interest

Jonathan Jacobs / @bioinformer
jonathan.jacobs@gmail.com

:)

With thanks to

- Paul Gardner https://twitter.com/ppgardne
- Chris Bear https://twitter.com/bear_chris
- Rachael Lappan https://github.com/rachaellappan
- Rayan Chikhi https://twitter.com/rayanchikhi
- Anuradha Ravi https://twitter.com/anuradharavi10
- Ann Gregory https://twitter.com/gregory_annc
-
- Tons of other people.
- If you have contributed - please leave a comment and let me know.

Some kind of table of contents…

The goal of this document is to capture current tools, methods and the overall madness of metagenomics as a science and the emerging commercial field. Wherever possible, I'll add links / references to each resource, etc. but **THIS IS BY NO MEANS COMPLETE**.

PLEASE LEAVE A COMMENT TO ADD or EDIT SOMETHING IN ANY OF THE SECTIONS.

CONFERENCES / WORKSHOPS / VENUES

Check out my (incomplete) public Google Calendar for a list of meetings and conferences, etc.

https://goo.gl/ogArZ8

SHOTGUN METAGENOMICS ANALYSIS METHODS

The majority of the methods outlined below are intended for community profiling - not determining if a specific pathogen - is present/absent from the profile. It's implied that these tools will be used as the basis for an initial profiling of a sample, and then potential pathogens of interest will be assessed from the data using simple filtering.

KARP (2017)

1. Karp: Accurate and fast taxonomic classification using pseudoalignment. Mark Reppell, John Novembre doi: https://doi.org/10.1101/097949

metaCRAM (2017)

1. MetaCRAM: an integrated pipeline for metagenomic taxonomy identification and compression

metaBIT (2017)

1. metaBIT, an integrative and automated metagenomic pipeline for analysing microbial profiles from high-throughput sequencing shotgun data.

SLIMM (2017)

1. Dadi TH, Renard BY, Wieler LH, Semmler T, Reinert K. **SLIMM: species level identification of microorganisms from metagenomes.** PeerJ. 2017 Mar 28;5:e3138. doi: 10.7717/peerj.3138.

MetaSpark (2017)

1. Zhou W, Li R, Yuan S, Liu C, Yao S, Luo J, Niu B. **MetaSpark: a spark-based distributed processing tool to recruit metagenomic reads to reference genomes**. Bioinformatics. 2017 Apr 1;33(7):1090-1092. doi: 10.1093/bioinformatics/btw750. PubMed PMID: 28065898.

AFS (All-Food-Seq) (2017)

1. Liu Y, Ripp F, Koeppel R, Schmidt H, Hellmann SL, Weber M, Krombholz CF, Schmidt B, Hankeln T. **AFS: identification and quantification of species composition by metagenomic sequencing**. Bioinformatics. 2017 Jan 5. pii: btw822. doi: 10.1093/bioinformatics/btw822.

VIrusSeeker (2017)

1. Zhao G, Wu G, Lim ES, Droit L, Krishnamurthy S, Barouch DH, Virgin HW, Wang D. **VirusSeeker, a computational pipeline for virus discovery and virome composition analysis**. Virology. 2017 Mar;503:21-30. doi:10.1016/j.virol.2017.01.005.

MetaMLST (2017)

1. **MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples.** Moreno Zolfo, Adrian Tett, Olivier Jousson, Claudio Donati, Nicola Segata. Nucleic Acids Res. 2017 Jan 25; 45(2): e7. Published online 2016 Sep 19. doi://10.1093/nar/gkw837 PMCID: PMC5314789
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5314789/

deSPI (2016)

1. Guan, D., Liu, B., & Wang, Y. (2016). **deSPI: efficient classification of metagenomic reads with lightweight de Bruijn graph-based reference indexing**. bioRxiv. http://doi.org/10.1101/080200
2. https://github.com/hitbc/deSPI

k-SLAM (2016) (k-mer Sorted List Alignment and Metagenomics)

1. Ainsworth, D., Sternberg, M. J. E., Raczy, C., & Butcher, S. A. (2016). k-SLAM: accurate and ultra-fast taxonomic classification and gene identification for large metagenomic data sets. Nucleic Acids Research, gkw1248–8. http://doi.org/10.1093/nar/gkw1248
2. https://github.com/aindj/k-SLAM

MetaPalette (2016)

1. MetaPalette: a k-mer painting approach for metagenomic taxonomic profiling and quantification of novel strain variation Koslicki, D., Falush, D. bioRxiv; doi: http://dx.doi.org/10.1101/039909

CENTRIFUGE (2016)

1. **Centrifuge: rapid and sensitive classification of metagenomic sequences** Daehwan Kim, Li Song, Florian P Breitwieser, Steven Salzberg bioRxiv 054965; doi: http://dx.doi.org/10.1101/054965
2. http://www.ccb.jhu.edu/software/centrifuge/

TAXONOMER (2016)

1. Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T, Graf EH, Tardif KD, Kapusta A, Rynearson S, Stockmann C, Queen K, Tong S, Voelkerding KV, Blaschke A, Byington CL, Jain S, Pavia A, Ampofo K, Eilbeck K, Marth G, Yandell M, Schlaberg R. **Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling**. Genome Biol. 2016 May 26;17(1):111. doi: 10.1186/s13059-016-0969-1
2. http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0969-1
3. https://github.com/Yandell-Lab/taxonomer_0.5/releases

BRACKEN (2016)

1. Bracken: Estimating species abundance in metagenomics data. Jennifer Lu, Florian P Breitwieser, Peter Thielen, Steven L Salzberg doi: http://dx.doi.org/10.1101/051813
2. https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-3-r46
3. http://ccb.jhu.edu/software/bracken/

MGMapper (2016)

1. http://cge.cbs.dtu.dk/services/MCLARK-SGmapper
2. Publication in press. Presented at #SFAF16

CLARK-S (2016)

1. Need reference. TBD.

Lattice-METAge (2016)

1. Jha M, Malhotra R, Acharya R. **A Generalized Lattice based Probabilistic Approach for Metagenomic Clustering.** IEEE/ACM Trans Comput Biol Bioinform. 2016 May 5. [Epub ahead of print] PubMed PMID: 27168602
2. https://github.com/lattclus/lattice-metage

MEGAN (2016; MEGAN5, 2013, MEGAN1 2007)

1. http://ab.inf.uni-tuebingen.de/software/megan65/
2. Huson DH, Beier S, Flade I, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. Poisot T, ed. PLoS Computational Biology. 2016;12(6):e1004957. doi:10.1371/journal.pcbi.1004957.Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. Genome Res. 17, 377–86 (2007).

Bayesian Identification of Bacterial Strains (BIB) (2015)

1. Sankar et al. **Bayesian Identification of Sequencing Strains from Sequencing Data**
2. http://arxiv.org/pdf/1511.06546v2.pdf
3. https://github.com/PROBIC/BIB

Diamond (2015)

1. https://github.com/bbuchfink/diamond
2. http://www.nature.com/nmeth/journal/v12/n1/full/nmeth.3176.html

PathoSphere (2015)

1. Kilianski, A. *et al.* Pathosphere.org: pathogen detection and characterization through a web-based, open source informatics platform. *BMC Bioinformatics* 16, 416 (2015).
2. http://sourceforge.net/projects/pathosphere/?source=directory
3. http://www.pathosphere.org

Seed-Kraken (2015)

1. K. Břinda, M. Sykulski, and G. Kucherov. Spaced seeds improve *k*-mer-based metagenomic classification. Bioinformatics (2015) 31 (22): 3584-3592. doi: 10.1093/bioinformatics/btv419
2. http://seed-kraken.readthedocs.org/
3. Extended version of the paper: http://arxiv.org/pdf/1502.06256v3.pdf
4. https://github.com/macieksk/seed-kraken
5. http://cbio.ensmp.fr/~nvaroquaux/documents/abs4ngs/kucherov.pdf

Kaiju (2015)

1. Peter Menzel, Kim Lee Ng, Anders Krogh, “Kaiju: Fast and sensitive taxonomic classification for metagenomics” doi: http://dx.doi.org/10.1101/031229
2. https://github.com/bioinformatics-centre/kaiju

kallisto (2015)

1. Originally developed for RNASeq - this EM algorithm has been shown to be effective with metagenomics data as well. https://pachterlab.github.io/kallisto/about.html
2. http://arxiv.org/pdf/1510.07371v1.pdf

CoMeta (2015)

1. Kawulok J, Deorowicz S. CoMeta: classification of metagenomes using k-mers. PLoS One. 2015 Apr 17;10(4):e0121453. doi: 10.1371/journal.pone.0121453. eCollection 2015.
2. https://github.com/jkawulok/cometa

metaMix (2015)

1. Morfopoulou and Piagnol. "**Bayesian mixture analysis for metagenomic community profiling."** Bioinformatics 31 I (18): 2930-2938 (2015) doi: 10.1093/bioinformatics/btv317
2. https://cran.r-project.org/web/packages/metaMix/index.html
    a. This is a similarity based community profiling method. It is extremely computationally intensive, and most likely not suitable for routine testing of clinical or environmental samples in a high-throughput / high-case load lab.
    b. From the documentation: "*Uses a mixture model based approach with parallel Monte Carlo Markov chains for the exploration of the species space to identify the set of species likely to contribute to the community mixture of a sample*

MetaPhlan2 (2015)

1. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. *MetaPhlAn2 for enhanced metagenomic taxonomic profiling.* Nat Methods. 2015 Sep 29;12(10):902-3. doi: 10.1038/nmeth.3589. PubMed PMID: 26418763.
2. https://bitbucket.org/biobakery/metaphlan2
3. http://huttenhower.sph.harvard.edu/metaphlan2

One Codex (2015)

1. http://onecodex.com
2. S. S. Minot, N. Krumm, N. B. Greenfield, One Codex: A Sensitive and Accurate Data Platform for Genomic Microbial Identification. *bioRxiv* (2015) (available at http://biorxiv.org/content/early/2015/09/28/027607.abstract).

MetaPORE (2015)

1. Greninger, A. L. et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 7, 99 (2015).
2. specific for Oxford Nanopore data
3. https://github.com/chiulab/metaPORE (*repo currently empty*, 30-SEP-2015)

CLARK (2015)

1. Ounit, R., Wanamaker, S., Close, T. J. & Lonardi, S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. BMC Genomics 16, 236 (2015).
2. http://clark.cs.ucr.edu/Overview/

RIEMS (2015)

1. Scheuch, M., Höper, D. & Beer, M. RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets. BMC Bioinformatics 16, 69 (2015).
2. http://www.fli.de/en/institutes/institut-fuer-virusdiagnostik/labore-arbeitsgruppen/labor-fuer-ngs-und-microarray-diagnostik/ (GERMAN)

GOTTCHA (2015)

1. https://github.com/LANL-Bioinformatics/GOTTCHA
2. Freitas TA, Li PE, Scholz MB, Chain PS. Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. Nucleic Acids Res. 2015 Mar 12. pii: gkv180. [Epub ahead of print] PubMed PMID: 25765641.

ChainMapper / kmer Finder (2014)

1. Hasman, H. *et al.* Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J. Clin. Microbiol.* **52,** 139–46 (2014).
2. http://www.genomicepidemiology.org
3. https://bitbucket.org/genomicepidemiology/

ezVIR (2014)

1. https://www.ncbi.nlm.nih.gov/pubmed/25009045

RealTime Genomics (2014)

1. John G. Cleary, et.al. "Joint Variant and De Novo Mutation Identification on Pedigrees from High-Throughput Sequencing Data." Journal of Computational Biology. June 2014, 21(6): 405-419. doi:10.1089/cmb.2014.0029.
2. http://www.realtimegenomics.com (COMMERCIAL)

MetaGenie (2014)

1. Rawat, A., Engelthaler, D. M., Driebe, E. M., Keim, P. & Foster, J. T. MetaGeniE: characterizing human clinical samples using deep metagenomic sequencing. *PLoS One* **9,** e110915 (2014).
2. https://github.com/ngsclinical/metagenie

MePIC (2014)

1. Takeuchi, F. *et al.* MePIC, metagenomic pathogen identification for clinical specimens. *Jpn. J. Infect. Dis.* **67,** 62–65 (2014).
2. https://mepic.nih.go.jp

CensuScope (2014)

1. Shamsaddini, A. *et al.* Census-based rapid and accurate metagenome taxonomic profiling. *BMC Genomics* **15,** 918 (2014).
2. https://hive.biochemistry.gwu.edu/dna.cgi?cmd=censuscope

GroopM (2014)

1. Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. (2014) GroopM: an automated tool for the recovery of population genomes from related metagenomes. *PeerJ* 2:e603 http://dx.doi.org/10.7717/peerj.603
2. http://ecogenomics.github.io/GroopM/

PhyloSift (2014)

1. Darling, A. E. et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2, e243 (2014).
2. https://github.com/gjospin/PhyloSift
3. https://peerj.com/articles/243/

SIGMA (2014)

1. Ahn, T., Chai, J. & Pan, C. Sigma: Strain-level Inference of Genomes from Metagenomic Analysis for Biosurveillance. Bioinformatics 1–8 (2014).

2. http://sigma.omicsbio.org

PathoScope 2.0 (2014)

1. Hong, C. *et al.* PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome* **2,** 33 (2014).
2. Francis, O. E. *et al.* Pathoscope: species identification and strain attribution with unassembled sequencing data. *Genome Res.* **23,** 1721–9 (2013).
3. http://sourceforge.net/projects/pathoscope/
4. Commercial implementation: http://www.aperiomics.com

SURPI (2014)

1. http://chiulab.ucsf.edu/surpi/
2. Naccache, S. N. *et al.* A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res.* (2014). doi:10.1101/gr.171934.113

Amordad (2014)

1. http://smithlabresearch.org/software/amordad/
2. Behnam, E. & Smith, A. D. The Amordad database engine for metagenomics. *Bioinformatics* 1–8 (2014). doi:10.1093/bioinformatics/btu405
3. not a classifier / but allows a MASH-style sample:sample comparison / lookup

GSMer (2014)

1. https://github.com/qichao1984/GSMer
2. Tu, Q., He, Z. & Zhou, J. Strain/species identification in metagenomes using genome-specific markers. Nucleic Acids Res. 42, e67 (2014).

KRAKEN (2014)

1. http://ccb.jhu.edu/software/kraken
2. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46 (2014).

eXpress (2013)

1. Roberts, A. and Pachter, L. (2013). **Streaming fragment assignment for real-time analysis of sequencing experiments.** Nature methods, 10(1):71–73.
2. Originally designed for RNA-seq - has been adapted for use in metagenomics profiling (see kalisto, 2015 above)

PathogenFinder (2013)

1. S. Cosentino, M. Voldby Larsen, F. Møller Aarestrup, O. Lund, PathogenFinder--distinguishing friend from foe using bacterial whole genome sequence data. *PLoS One*. **8**, e77302 (2013).
2. https://cge.cbs.dtu.dk//services/PathogenFinder/ (web based)

specI (2013)

1. http://vm-lux.embl.de/~mende/specI/
2. Mende, D. R., Sunagawa, S., Zeller, G. & Bork, P. Accurate and universal delineation of prokaryotic species. Nat. Methods 10, 881–4 (2013).

GaSiC (2013)

1. Lindner, M. S. & Renard, B. Y. Metagenomic abundance estimation and diagnostic testing on species level. Nucleic Acids Res. 41, 1–8 (2013).

2. http://sourceforge.net/projects/gasic/

READSCAN (2013)

1. http://cbrc.kaust.edu.sa/readscan/
2. Naeem, R., Rashid, M. & Pain, A. READSCAN: a fast and scalable pathogen discovery program with accurate genome relative abundance estimation. *Bioinformatics* **29,** 391–2 (2013).

LMAT (2013)

1. http://sourceforge.net/projects/lmat/
2. Ames, S. K. et al. Scalable metagenomic taxonomy classification using a reference genome database. Bioinformatics 1–7 (2013). doi:10.1093/bioinformatics/btt389
3. aka MTCP

MOCAT / mOTU (2013)

1. http://www.bork.embl.de/software/mOTU/
2. http://vm-lux.embl.de/~kultima/MOCAT/
3. Sunagawa, S. et al. Metagenomic species profiling using universal phylogenetic marker genes. Nat. Methods 10, 1196–9 (2013).
4. Kultima, J. R. et al. MOCAT: a metagenomics assembly and gene prediction toolkit. PLoS One 7, e47656 (2012).

MetaBEETL (2013)

1. https://github.com/BEETL/BEETL
2. Ander C, Schulz-Trieglaff OB, Stoye J, Cox AJ: metaBEETL: high-throughput analysis of heterogeneous microbial populations from shotgun DNA sequences. BMC Bioinformatics 2013, 14 Suppl 5:S2

SPANNER (2013)

1. http://kiwi.cs.dal.ca/Software/SPANNER
2. Porter MS, Beiko RG: SPANNER: taxonomic assignment of sequences using pyramid matching of similarity profiles. Bioinformatics 2013, 29:1858–1864.

TaxyPro (2013)

1. http://gobics.de/TaxyPro/
2. H. Klingenberg, K.P. Asshauer, T. Lingner and P. Meinicke. Protein signature-based estimation of metagenomic abundances including all domains of life and viruses. Bioinformatics, 29(8):973-80, 2013.

MetaCV (2013)

1. http://metacv.sourceforge.net/
2. Liu J, Wang H, Yang H, Zhang Y, Wang J, Zhao F, Qi J: Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms. Nucleic Acids Res 2013, 41:e3..

MetaPhlan (2012)

1. http://huttenhower.sph.harvard.edu/metaphlan
2. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9,** 811–4 (2012).

Sequedex (2012)

1. http://sequedex.lanl.gov

2. Berendzen, J. *et al.* Rapid phylogenetic and functional classification of short genomic fragments with signature peptides. *BMC Res. Notes* **5,** 460 (2012).

Real Time Metagenomics (2012)

1. https://edwards.sdsu.edu/RTMg/
2. Edwards RA, Olson R, Disz T, Pusch GD, Vonstein V, Stevens R, Overbeek R: Real time metagenomics: using k-mers to annotate metagenomes. Bioinformatics 2012, 28:3316–3317.

Genometa (2012)

1. http://genomics1.mh-hannover.de/genometa/
2. Davenport CF, Neugebauer J, Beckmann N, Friedrich B, Kameri B, Kokott S, Paetow M, Siekmann B, Wieding-Drewes M, Wienhöfer M, Wolf S, Tümmler B, Ahlers V, Sprengel F: Genometa--a fast and accurate classifier for short metagenomic shotgun reads. PLoS ONE 2012, 7:e41224.

MetaBin (2012)

1. http://metabin.riken.jp/
2. Sharma VK, Kumar N, Prakash T, Taylor TD: Fast and accurate taxonomic assignments of metagenomic sequences using MetaBin. PLoS ONE 2012, 7:e34030.

RITA (2012)

1. http://kiwi.cs.dal.ca/Software/RITA
2. MacDonald NJ, Parks DH, Beiko RG: Rapid identification of high-confidence taxonomic assignments for metagenomic data. Nucleic Acids Res 2012, 40:e111

GRAMMy (2011)

1. Xia, L. C., Cram, J. A., Chen, T., Fuhrman, J. A., and Sun, F. (2011). GRAMMy: Accurate Genome Relative Abundance Estimation Based on Shotgun Metagenomic Reads. PLoS ONE, 6(12).
2. https://bitbucket.org/charade/grammy/wiki/Home

PathSeq (2011)

1. Kostic, A. D. et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. Nat. Biotechnol. 29, 393–6 (2011).
2. http://www.broadinstitute.org/software/pathseq/index.html

CloVR (2011)

1. http://www.clovr.org
2. Angiuoli, S. V et al. CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. BMC Bioinformatics 12, 356 (2011).

PhymmBL (2011)

1. http://www.cbcb.umd.edu/software/phymm/
2. Brady, A. & Salzberg, S. L. Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. Nat. Methods 6, 673–6 (2009).
3. Brady, A. & Salzberg, S. PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nat. Methods* **8,** 367 (2011).

MetaPhyler (2011)

1. http://metaphyler.cbcb.umd.edu/

2. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M: Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. BMC Genomics 2011, 12(Suppl 2):S4.

CARMA3 (2011)

1. http://wwww.cebitec.uni-bielefeld.de/webcarma.cebitec.uni-bielefeld.de/manual.html
2. Gerlach W, Stoye J: Taxonomic classification of metagenomic shotgun sequences with CARMA3. Nucl Acids Res 2011:gkr225

DiScRIBinATE (2010)

1. http://metagenomics.atc.tcs.com/binning/DiScRIBinATE/
2. Ghosh TS, Monzoorul Haque M, Mande SS: DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. BMC Bioinformatics 2010, 11 Suppl 7:S14.

NBC (2008)

1. Software not available / requested.
2. Rosen, G., Garbarine, E., Caseiro, D., Polikar, R. & Sokhansanj, B. Metagenome fragment classification using N-mer frequency profiles. *Adv. Bioinformatics* **2008,** 205969 (2008).

MG-RAST (2008)

1. http://metagenomics.anl.gov/
2. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA: The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 2008, 9:386.

megaBLAST (2000)

1. http://blast.ncbi.nlm.nih.gov
2. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–14 (2000).

Targeted Amplicon Sequencing

Examples include 16S rRNA analysis, etc -- but I'm not actively updating this. [last edit ~2015]. Interesting that most commercial companies simply use these tools or some variant of these tools.

DADA2 (2016)

1) 1: Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. **DADA2: High-resolution sample inference from Illumina amplicon data.** Nat Methods. 2016 Jul;13(7):581-3. doi: 10.1038/nmeth.3869. Epub 2016 May 23. PubMed PMID: 27214047; PubMed Central PMCID: PMC4927377.
2) https://github.com/benjjneb/dada2

metagenomeSeq Bioconductor package (2013)

1. http://cbcb.umd.edu/software/metagenomeSeq

2. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. Differential abundance analysis for microbial marker-gene surveys. Nat. Methods 10, 1200–2 (2013).

MOTHUR (2009- 2014)

1. http://www.mothur.org
2. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75,** 7537–41 (2009).

CloVR (2011)

3. http://www.clovr.org
4. Angiuoli, S. V et al. CloVR: a virtual machine for automated and portable sequence analysis from the desktop using cloud computing. BMC Bioinformatics 12, 356 (2011).

QIIME (2010)

1. http://qiime.org
2. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–6 (2010).

METHODS TO GENERATE SYNTHETIC READS

Bear (2014)

1. https://github.com/sej917/BEAR
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4168713/

NeSSM (2013)

1. http://cbb.sjtu.edu.cn/~ccwei/pub/software/NeSSM.php
2. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0075448

cMESSI (2012)

1. https://sourceforge.net/projects/cmessi/
2. http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0031386

Grinder (2012)

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3384353/

MetaSim (2008)

1. http://ab.inf.uni-tuebingen.de/software/metasim
2. Richter DC, Ott F, Auch AF, Schmid R, Huson DH (2008): MetaSim—A Sequencing Simulator for Genomics and Metagenomics. PLoS ONE 3(10): e3373. doi:10.1371/journal.pone.0003373

FUNCTIONAL CHARACTERIZATION PIPELINES

http://www.nature.com/nbt/journal/v31/n6/abs/nbt.2579.html

HUMAnN2 (2017)

1. https://bitbucket.org/biobakery/humann2/wiki/Home

2. HUMAnN2 manuscript submitted. HUMAnN1 reference: Abubucker, S. *et al.,* (2012). **Metabolic reconstruction for metagenomic data and its application to the human microbiome.** PLoS Computational Biology 13(8):e1002358 doi:10.1371/journal.pcbi.1002358

MetaPath (2015)

1. http://www.cbcb.umd.edu/software/metapath

Tax4Fun (2015)

1. http://tax4fun.gobics.de/
2. Aßhauer, K.P., Wemheuer B. , Daniel R. and Meinicke, P. (2015). **Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data.** Bioinformatics 31(17), 2015, 2882–2884 doi: 10.1093/bioinformatics/btv287

IslandViewer 3/GenomeD3Plot (2015)

1. http://pathogenomics.sfu.ca/islandviewer
2. Dhillon, B. *et al.* (2015). **IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis.** Nucleic Acids Research doi: 10.1093/nar/gkv401

Roary (2015)

1. http://sanger-pathogens.github.io/Roary
2. Page, A.J. *et al.* (2015). **Roary: rapid large-scale prokaryote pan genome analysis.** Bioinformatics, 31(22), 3691–3693 doi: 10.1093/bioinformatics/btv421

SUPER-FOCUS (2015)

1. https://edwards.sdsu.edu/SUPERFOCUS
2. Silva, G.G.Z., Green, K.T., Dutilh, B.E., Edwards, R.A. (2015). **SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data.** Bioinformatics 1-8 doi: 10.1093/bioinformatics/btv584

SEARS (2015)

1. http://computing.bio.cam.ac.uk/sear/SEAR_WEB_PAGE/SEAR.html
2. Rowe, W. *et al.* (2015). **Search Engine for Antimicrobial Resistance: A Cloud Compatible Pipeline and Web Interface for Rapidly Detecting Antimicrobial Resistance Genes Directly from Sequence Data.** PLoS ONE 10(7): e0133492. doi:10.1371/journal.pone.0133492

KvarQ (2014)

1. http://www.swisstph.ch/kvarq
2. Steiner A., Stucki D., Coscolla, M., Borrell S, Gagneux S. (2014). **KvarQ: targeted and direct variant calling from fastq reads of bacterial genomes** BMC Genomics 15:881 doi: 10.1186/1471-2164-15-881

kSNP v2 (2013)

1. http://sourceforge.net/projects/ksnp/
2. Gardner, S.N. & Hall, B.G. (2013) **When Whole-Genome Alignments Just Won't Work: kSNP v2 Software for Alignment-Free SNP Discovery and Phylogenetics of Hundreds of Microbial Genomes.** PLoS ONE 8(12): e81760. doi:10.1371/journal.pone.0081760

PICRUSt (2013)

1. http://picrust.github.io/picrust/
2. Langille, M.G.I *et al.,* (2013). **Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences.** Nature Biotechnology 31(9):814-821. doi:10.1038/nbt.2676

SmashCommunity (2010)

1. http://www.bork.embl.de/software/smash
2. Manimozhiyan A., Harrington E.D., Foerstner, K.U., Raes J. and Bork, P. (2010). **SmashCommunity: a metagenomic annotation and analysis tool.** Bioinformatics 26 (23):2977-2978. doi:10.1093/bioinformatics/btq536

MG-RAST (2008)

1. http://metagenomics.anl.gov/
2. Meyer, F. *et al*. (2008) **The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes**. BMC Bioinformatics 9:386 doi:10.1186/1471-2105-9-386

SINGLE ISOLATE

The RAST Server (2008)

1. http://rast.nmpdr.org/
2. Aziz R.K. *et al*., (2008). **The RAST Server: Rapid Annotations using Subsystems Technology.** BMC Genomics 9:75 doi:10.1186/1471-2164-9-75

METAGENOMICS BENCHMARKING STUDIES

[[work in progress]] thanks to those who suggested adding this…

1. Vollmers, J. Wiegand, S. & Kasdter, A-K. (2017) **Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters!** PLoS One 12(1): e0169662. https://doi.org/10.1371/journal.pone.0169662.
2. **A comparative study of metagenomics analysis pipelines at the species level.** Yee Voan Teo, Nicola Neretti doi: https://doi.org/10.1101/081141
3. Siegwald L, Touzet H, Lemoine Y, Hot D, Audebert C, Caboche S. **Assessment of Common and Emerging Bioinformatics Pipelines for Targeted Metagenomics**. PLoS One. 2017 Jan 4;12(1):e0169563. doi: 10.1371/journal.pone.0169563
4. **Critical Assessment of Metagenome Interpretation – a benchmark of computational metagenomics software**. http://biorxiv.org/content/early/2017/01/09/099127.article-metrics
5. Peabody, M. A., Van Rossum, T., Lo, R., & Brinkman, F. S. L. (2015). **Evaluation of shotgun metagenomics sequence classification methods using in silico and in vitro simulated communities.** *BMC Bioinformatics*, ***16***(1), 363. doi:10.1186/s12859-015-0788-5

6. Lindgreen, S., Adair, K. L. & Gardner, P. P. **An evaluation of the accuracy and speed of metagenome analysis tools**. *Sci Rep*. (2016). http://www.nature.com/articles/srep19233
7. Critical Assessment of Metagenomic Interpretation (2015) http://www.cami-challenge.org/faq
8. Oulas, A. *et al.* **Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies**. *Bioinform. Biol. Insights* **9,** 75–88 (2015). http://dx.doi.org/10.4137%2FBBI.S12462
9. Garcia-Etxebarria K, Garcia-Garcerà M, Calafell F (2014) **Consistency of metagenomic assignment programs in simulated and real data.** *BMC Bioinformatics.*
10. Sun, Y. *et al.* **A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis**. *Brief. Bioinform.* **13,** 107–21 (2012).
11. Bazinet, A. L. & Cummings, M. P. **A comparative evaluation of sequence classification programs.** *BMC Bioinformatics* **13,** 92 (2012).
12. Martin, J., Sykes, S., Young, S., Kota, K., Sanka, R., Sheth, N., Orvis, J., Soder- gren, E., Wang, Z., Weinstock, G. M., and Mitreva, M. (2012). **Optimizing Read Mapping to Reference Genomes to Determine Composition and Species Prevalence in Microbial Communities**. PLoS ONE, 7(6):e36427.

COMPARATIVE METAGENOMICS

1. Jing G, Sun Z, Wang H, Gong Y, Huang S, Ning K, Xu J, Su X. Parallel-META 3: **Comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities**. Sci Rep. 2017 Jan 12;7:40371. doi: 10.1038/srep40371. PubMed PMID: 28079128; PubMed Central PMCID: PMC5227994.
2. Ban Y, An L, Jiang H. **Investigating microbial co-occurrence patterns based on metagenomic compositional data**. Bioinformatics. 2015 Oct 15;31(20):3322-9. doi: 10.1093/bioinformatics/btv364. Epub 2015 Jun 16. PubMed PMID: 26079350; PubMed Central PMCID: PMC4795632.
3. Ondov BD, Treangen TJ, Mallonee AB, Bergman NH, Koren S, Phillippy AM. "**Fast genome and metagenome distance estimation using MinHash**", doi: http://dx.doi.org/10.1101/029827
    a. http://mash.readthedocs.org/en/latest/ MASH (2015)
4. McMurdie, P. J. & Holmes, S. **Waste not, want not: why rarefying microbiome data is inadmissible.** *PLoS Comput. Biol.* **10,** e1003531 (2014).
5. Paulson, J. N., Stine, O. C., Bravo, H. C. & Pop, M. **Differential abundance analysis for microbial marker-gene surveys**. *Nat. Methods* **10,** 1200–2 (2013).
6. Evans, S. N. & Matsen, F. A. **The phylogenetic Kantorovich-Rubinstein metric for environmental sequence samples.** *J. R. Stat. Soc. Ser. B Stat. Methodol.* **74,** 569–592 (2012).
    a. http://matsen.fhcrc.org/pplacer/ (see "guppy" tool)

    b. https://liorpachter.wordpress.com/2013/09/18/unifrac-revealed/#more-471
7. Huson, D. H., Richter, D. C., Mitra, S., Auch, A. F., and Schuster, S. C. (2009). **Methods for comparative metagenomics.** BMC bioinformatics, 10(Suppl 1):S12.
8. Rodriguez-Brito, B., Rohwer, F., and Edwards, R. A. (2006). **An application of statistics to comparative metagenomics.** BMC bioinformatics, 7(1):162.
9. Tringe, S. G., Von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., et al. (2005). **Comparative metagenomics of microbial communities.** Science, 308(5721):554–557.

METAGENOMICS REFERENCE DATASETS

too few of these exist

1. Critical Assessment of Metagenomic Interpretation (2015) http://www.cami-challenge.org/faq
2. Mende, D. R., Waller, A. S., Sunagawa, S., Jrvelin, A. I., Chan, M. M., Aru- mugam, M., Raes, J., and Bork, P. (2012). **Assessment of Metagenomic Assembly Using Simulated Next Generation Sequencing Data**. PLoS ONE, 7(2):e31386.
3. Bokulich, N. A., Rideout, J. R., Mercurio, W. G., Wolfe, B., Maurice, C. F., Dutton, R. J., ... & Caporaso, J. G. (2016). **mockrobiota: a public resource for microbiome bioinformatics benchmarking** (No. e2065v1). PeerJ Preprints.

(Meta)GENOMICS STANDARDS Papers

[[articles for genomics, metagenomics and/or microbial forensics standards]]

1. Sinha, Rashmi, et al. "**The microbiome quality control project: baseline study design and future directions**." Genome biology 16.1 (2015): 276.
    a. The Microbiome Quality Control project http://www.mbqc.org
2. Budowle, B. *et al.* Validation of high throughput sequencing and microbial forensics applications. *Investig. Genet.* **5,** 9 (2014).
3. J. T. Ladner *et al.*, Standards for sequencing viral genomes in the era of high-throughput sequencing. *MBio*. **5**, e01360–14 (2014).

Metagenomics Assembly Tools

1. Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P.A. (2017). **metaSPAdes: a new versatile metagenomic assembler.** Genome Research 27(5):824-834, doi:10.1101/gr.213959.116.
2. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. **plasmidSPAdes: assembling plasmids from whole genome sequencing data.** Bioinformatics. 2016 Nov 15;32(22):3380-3387. Epub 2016 Jul 27. PubMed PMID: 27466620. https://doi.org/10.1093/bioinformatics/btw493
    a. http://spades.bioinf.spbau.ru/plasmidSPAdes/
3. Afiahayati,Sato, K., & Sakakibara., Y. (2015). **MetaVelvet-SL: an extension of the Velvet assembler to a de novo metagenomic assembler utilizing supervised learning.** DNA Research 22(1):69-77, doi: 10.1093/dnares/dsu041

4. Luo, C. et al. **ConStrains identifies microbial strains in metagenomic datasets**. Nat Biotech advance on, (2015). http://www.nature.com/nbt/journal/vaop/ncurrent/full/nbt.3319.html
   https://bitbucket.org/luo-chengwei/constrains
5. Cleary et al.. **Detection of low-abundance bacterial strains in metagenomic datasets by eigengenome partitioning**. Nature Biotechnology 33, 1053–1060 (2015) doi:10.1038/nbt.3329
   a. **Latent Strain Analysis (2015)**
   b. http://latentstrainanalysis.readthedocs.org/en/latest/
   c. Looks at covariance relationships between k-mers.
6. Li D. et al, **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.** (2015)

[Contig or Otherwise] Deduplication tools

Thanks Ann Gregory for adding this section!

1. Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F. (2017). **dRep: A tool for fast and accurate genome de-replication that enables tracking of microbial genotypes and improved genome recovery from metagenomes.** bioRxiv (preprint). 108142. DOI: 10.1101/108142

Commercial / Consumer Metagenomics / Microbiome Therapeutics & Services

Probably missed some… probably should start breaking these into separate categories… but… holy crap that's a lot of VC funding....

1. American Gut Project
2. AOBiome
3. Aperiomics
4. CeMeT GmbH
5. CosmosID
6. DayTwo
7. Diversigen
8. Eligo Bioscience
9. Enterome Bioscience
10. IDbyDNA
11. Karius
12. Maat Pharma
13. Metagenome Analytics, LLC

14. Metabiota
15. MicroBiome Therapeutics
16. NatureMetrics
17. One Codex
18. OpenBiome
19. Phylagen
20. RealTime Genomics
21. ReBiotix
22. Second Genome
23. Seres Therapeutics
24. Siolta Therapeutics
25. Thryve
26. uBiome
27. Vedanta Biosciences
28. Viome
29. WholeBiome

MISC. Other stuff below…

BIOSURVEILLANCE / (Meta)GENOMICS REVIEW ARTICLES

This (growing) list of relevant review articles for the use of metagenomics / genomics in biosurveillance and/or clinical diagnostics. [[***NEEDS TO BE UPDATED!!!]***

1. Lefterova, M. I., Suarez, C. J., Banaei, N. & Pinsky, B. A. Next-Generation Sequencing for Infectious Disease Diagnosis and Management A Report of the Association for Molecular Pathology. *J. Mol. Diagnostics* **17,** (**2015**).
2. Franzosa, E. a. *et al.* Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* **13,** 360–372 (**2015**).
3. Madoff, L.C., and Li, A. (**2014**). *Web-Based Surveillance Systems for Human, Animal and Plant Diseases.* Microbiol. Spectr. *2*, OH–0015–2012.
4. Lipkin, W.I. (**2013**). *The changing face of pathogen discovery and surveillance.* Nat. Rev. Microbiol. *11*, 133–141.
5. Tegos, G.P. (2013). *Biodefense: trends and challenges in combating biological warfare agents.* Virulence *4*, 740–744.
6. Valdivia-Granda, W. a (**2013**). *Biosurveillance enterprise for operational awareness, a genomic-based approach for tracking pathogen virulence.* Virulence *4*, 745–751.
7. Miller, R.R., Montoya, V., Gardy, J.L., Patrick, D.M., and Tang, P. (**2013**). *Metagenomics for pathogen detection in public health.* Genome Med. *5*, 81.
8. Kaydos-Daniels, S.C., Rojas Smith, L., and Farris, T.R. (**2013**). *Biosurveillance in outbreak investigations.* Biosecur. Bioterror. *11*, 20–28.
9. Kman, N.E., and Bachmann, D.J. (**2012**). *Biosurveillance: a review and update.* Adv. Prev. Med. *2012*, 301408.

10. Russell, K.L., Rubenstein, J., Burke, R.L., Vest, K.G., Johns, M.C., Sanchez, J.L., Meyer, W., Fukuda, M.M., and Blazes, D.L. (**2011**). *The Global Emerging Infection Surveillance and Response System (GEIS), a U.S. government tool for improved global biosurveillance: a review of 2009.* BMC Public Health *11 Suppl 2*, S2.

GENOMICS DATA COMPRESSION / STREAMING

When gzip just isn't enough… . Also see: http://omictools.com/data-compression-c383-p1.html

1. Roguski, Ł., & Ribeca, P. (**2015**). CARGO: Effective format-free compressed storage of genomic information. Retrieved from http://arxiv.org/abs/1506.05185
2. Y. Zhang *et al.*, Light-weight reference-based compression of FASTQ data. *BMC Bioinformatics*. **16**, 188 (**2015**).
3. S. Pathak, S. Rajasekaran, LFQC: a lossless compression algorithm for FASTQ files. *Bioinformatics* (**2014**), doi:10.1093/bioinformatics/btu701.
4. J. K. Bonfield, M. V. Mahoney, Compression of FASTQ and SAM format sequencing data. *PLoS One*. **8**, e59190 (**2013**).
5. (REVIEW OF PRIOR METHODS) S. Deorowicz, S. Grabowski, Data compression for sequencing data. *Algorithms Mol. Biol.* **8**, 25 (**2013**).