

# SISMID14 Metagenomics      References, June,2011

Alex Aleksyenko and Susan Holmes ©

## References

- Acinas, S., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M. (2005). Pcr-induced sequence artifacts and bias: insights from comparison of two 16s rrna clone libraries constructed from the same sample. *Applied and Environmental Microbiology*, 71(12):8966.
- Barker, D., Meade, A., and Pagel, M. (2007). Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. *Bioinformatics*, 23(1):14–20.
- Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J., and Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci USA*, 103(3):732–7.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., and Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA*, 99(22):14250–5.
- Caporaso, J., Bittinger, K., Bushman, F., DeSantis, T., Andersen, G., and Knight, R. (2010a). Pynast: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, 26(2):266.
- Caporaso, J., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F., Costello, E., Fierer, N., Peña, A., Goodrich, J., Gordon, J., et al. (2010b). Qiime allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336.
- Cazes, P., Chessel, D., and Dolédec, S. (1988). L'analyse des correspondances internes dun tableau partitionné: son usage en hydrobiologie. *Rev. Stat. Appl.*, 36.

- Chakerian, J. and Holmes, S. (2010a). Computational methods for evaluating phylogenetic trees. arXiv.
- Chakerian, J. and Holmes, S. (2010b). distory:Distances between trees.
- Charnomordic, B. and Holmes, S. (2001). Correspondence analysis for microarrays. *Statistical Graphics and Computing Newsletter*, 12.
- Chessel, D., Dufour, A., and Thioulouse, J. (2004). The ade4 package - i: One-table methods. *R News*, 4(1):5-10.
- Costello, E., Lauber, C., Hamady, M., Fierer, N., Gordon, J., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694.
- Davey, H. and Kell, D. (1996). Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analyses. *Microbiology and Molecular Biology Reviews*, 60(4):641.
- DeSantis, T., Keller, K., Karaoz, U., Alekseyenko, A., Singh, N., Brodie, E., Pei, Z., Andersen, G., and Larsen, N. (2011). Simrank: Rapid and sensitive general-purpose k-mer search tool. *BMC ecology*, 11(1):11.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L. (2006). Green-genes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl. Environ. Microbiol.*, 72(7):5069-5072.
- Diaconis, P. (1988). Group representations in probability and statistics. *Lecture Notes-Monograph Series*, 11.
- Diaconis, P., Goel, S., and Holmes, S. (2007). Horseshoes in multidimensional scaling and kernel methods. *Annals of Applied Statistics*.
- Diaconis, P. and Holmes, S. (1998). Matchings and phylogenetic trees. *Proc. Natl. Acad. Sci. USA*, 95(25):14600-14602 (electronic).
- Dray, S., Chessel, D., and Thioulouse, J. (2003). Procrustean co-inertia analysis for the linking of multivariate datasets. *Ecoscience*.
- Dray, S., Chessel, D., and Thioulouse, J. (2008). Co-inertia analysis and the linking of ecological data tables. *Eco Soc America*.

- Dray, S., Dufour, A., and Chessel, D. The ade4 package—ii: Two-table and k-table methods. *New Functions for Multivariate Analysis*.
- Drummond, A. and Rambaut, A. (2007). Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1):214.
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science*, 308(5728):1635–8.
- Efron, B. and Tibshirani, R. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.
- Escoufier, Y. (1977). Operators related to a data matrix. In Barra, J. e. a., editor, *Recent developments in Statistics.*, pages 125–131. North Holland,.
- Escoufier, Y. (1987). The duality diagram: a means of better practical applications. In Legendre, P. and Legendre, L., editors, *Development in numerical ecology.*, pages 139–156.
- Evans, S. N. and Matsen, F. A. (2010). The phylogenetic kantorovich-rubinstein metric for environmental sequence samples. arXiv, q-bio.PE.
- Faith, D. (1992a). Conservation evaluation and phylogenetic diversity. *Biological Conservation*.
- Faith, D. (1992b). Conservation evaluation and phylogenetic diversity. *Biological Conservation*.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15.
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H., and Zhang, J. (2004a). Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*, 5(10):R80.

- Gentleman, R. C., Carey, V. J., Bates, D. M., et al. (2004b). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, 5:R80.
- Grafen, A. (1989). The phylogenetic regression. *Philos Trans R Soc Lond, B, Biol Sci*, 326(1233):119–57.
- Gray, R., Tatem, A., Johnson, J., Alekseyenko, A., Pybus, O., Suchard, M., and Salemi, M. (2011). Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant staphylococcus aureus st239 genome-wide data within a bayesian framework. *Molecular Biology and Evolution*, 28(5):1593.
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press, London.
- Haas, B., Gevers, D., Earl, A., Feldgarden, M., Ward, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S., Sodergren, E., et al. (2011). Chimeric 16s rrna sequence formation and detection in sanger and 454-pyrosequenced pcr amplicons. *Genome Research*, 21(3):494.
- Hamady, M., Lozupone, C., and Knight, R. (2009). Fast unifracc: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and phylochip data. *The ISME Journal*.
- Holmes, S. (2003). Statistics for phylogenetic trees. *Theoretical Population Biology*, 63:17–32.
- Holmes, S. (2006). Multivariate analysis: The French way. In Nolan, D. and Speed, T. P., editors, *Probability and Statistics: Essays in Honor of David A. Freedman*, volume 56 of *IMS Lecture Notes–Monograph Series*. IMS, Beachwood, OH.
- Holmes, S., Alekseyenko, A., Timme, A., Nelson, T., Pasricha, P., and Spormann, A. (2011). Visualization and statistical comparisons of microbial communities using R packages on Phylochip data. In *Pacific Symposium on Biocomputing*. Pacific Symposium on Biocomputing, page 142.

- Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl 1:S96–104.
- Huson, D. H., Auch, A. F., Qi, J., and Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Res*, 17(3):377–86.
- Ideker, T., Ozier, O., Schwikowski, B., and Siegel, A. F. (2002). Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*, 18 Suppl 1:S233–40.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3):299–314.
- Jaccard, P. (1912). The distribution of flora in the alpine zone. *The New Phytologist*, 11:37–50.
- Johnson, P. L. F. and Slatkin, M. (2006). Inference of population genetic parameters in metagenomics: a clean look at messy data. *Genome Res*, 16(10):1320–7.
- Kapelner, A., Lee, P., and Holmes, S. (2007). An interactive statistical image segmentation and visualization system. *Medivis*, 00:81–86.
- Kassinen, A., Krogius-Kurikka, L., and Mäkivuokko, H. (2007). The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology*.
- Kembel, S., Cowan, P., and Helmus, M. (2010a). Picante: R tools for integrating phylogenies and ecology. ....
- Kembel, S., Cowan, P., Helmus, M., Cornwell, W., Morlon, H., Ackerly, D., Blomberg, S., and Webb, C. (2010b). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, 26(11):1463–1464.
- King, T., Elia, M., and Hunter, J. (1998). Abnormal colonic fermentation in irritable bowel syndrome. *The Lancet*.

- Ley, R., Lozupone, C., Hamady, M., Knight, R., and Gordon, J. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, 6(10):776–788.
- Lozupone, C. and Knight, R. (2005). Unifrac: a new phylogenetic method for comparing microbial communities. *Applied and environmental microbiology*, 71(12):8228.
- Malinen, E., Rinttilä, T., Kajander, K., and Mättö, J. (2005). Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time pcr. *The American journal of Gastroenterology*, 100(2):373–382.
- Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. Academic Press, NY.
- Martins, E. P. and Hansen, T. F. (1997). Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4):646–667.
- McMurdie, P. J., Behrens, S. F., Holmes, S., and Spormann, A. M. (2007a). Unusual codon bias in vinyl chloride reductase genes of *dehalococcoides* species. *Appl Environ Microbiol*, 73(8):2744–7.
- McMurdie, P. J., Spormann, A. M., and Holmes, S. (2007b). Multivariate analysis of codon usage in reductive dehalogenase genes: An example of decomposing the chisquare of a contingency table effectively. submitted.
- Moberg, J., Zubieta, I., Piceno, Y., and Andersen, G. (2007). High-density universal 16s rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. *Microbial ecology*.
- Nacu, S., Critchley-Thorne, R., Lee, P., and Holmes, S. (2007). Gene expression network analysis and applications to immunology. *Bioinformatics*, 23(7):850–8.
- Nelson, T., Holmes, S., Alekseyenko, A., Shenoy, M., DeSantis, T., Wu, C., Andersen, G., Winston, J., Sonnenburg, J., Pasricha, P., and Spormann,

- A. (2010a). PhyloChip microarray analysis reveals altered gastrointestinal microbial communities in a rat model of colonic hypersensitivity. *Neurogastroenterology & Motility*.
- Nelson, T., Pasricha, P., Holmes, S., and Spormann, A. (2010b). Shifts in luminal and mucosal microbial communities associated with an experimental model of irritable bowel syndrome. *Gastroenterology*.
- Nelson, T. A., Holmes, S., Alekseyenko, A. V., Shenoy, M., DeSantis, T., Wu, C., Anderson, G. L., Sonnenburg, J., Pasricha, P. J., and Spormann, A. Phylochip microarray analysis reveals altered gastrointestinal microbial communities in a rat model of colonic hypersensitivity. Submitted.
- Noguchi, H., Park, J., and Takagi, T. (2006). Metagene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res*, 34(19):5623–30.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., O’Hara, R. G., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2010). *vegan: Community Ecology Package*. R package version 1.17-0.
- Palmer, C., Bik, E., DiGiulio, D., Relman, D., and Brown, P. (2007). Development of the human infant intestinal microbiota. *PLoS biology*, 5(7):e177.
- Paradis, E. (2006). *Ape (analysis of phylogenetics and evolution) v1.8-2*. <http://cran.r-project.org/doc/packages/ape.pdf>.
- Paradis, E., Claude, J., and Strimmer, K. (2004). *Ape: analyses of phylogenetics and evolution in R language*. *Bioinformatics*, 20:289–290.
- Pavoine, S., Dufour, A., and Chessel, D. (2004). From dissimilarities among species to dissimilarities among communities: a double principal coordinate analysis. *Journal of theoretical biology*, 228(4):523–537.
- Pavoine, S., Ollier, S., and Pontier, D. (2005). Measuring diversity from dissimilarities with rao’s quadratic entropy: are any dissimilarities suitable? *Theoretical population biology*, 67(4):231–9.
- Pollard, K. S., Gilbert, H. N., Ge, Y., Taylor, S., and Dudoit, S. (2010). *multtest: Resampling-based multiple hypothesis testing*. R package version 2.4.0.

- Price, M., Dehal, P., and Arkin, A. (2010). Fasttree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, 5(3):e9490.
- Purdom, E. (2006). *Multivariate Kernel Methods in the Analysis of Graphical Structures*. PhD thesis, Stanford University.
- Purdom, E. (2010). Analysis of a data matrix and a graph: Metagenomic data and the phylogenetic tree. *Annals of Applied Statistics*.
- Qin, J., Li, R., Raes, J., Arumugam, M., Ber Solvsten Burgdorf, K., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Song Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Orben Hansen, T., Paslier, D. L., Linneberg, A., Nielsen, H. B., Tier, E. P., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Song Yu, C., Li, S., Jian, M., Zhou, Y., Li, Y., Zhang, X., Song Li, S., Qin, N., Yang, H., Wang, J., Brunak, S., and Francisco Guarner, J. D., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Consortium, M., Bork, P., Ehrlich, S. D., and Wang, J. (2010a). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.
- Qin, J., Li, R., Raes, J., and et al., M. A. (2010b). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.
- Quince, C., Lanzen, A., Davenport, R., and Turnbaugh, P. (2011). Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, 12(1):38.
- Rao, C. R. (1964). The use and interpretation of Principal Component Analysis in applied research. *Sankhya A*, 26:329–359.
- Ricotta, C. and Szeidl, L. (2006). Towards a unifying approach to diversity measures: bridging the gap between the Shannon entropy and Rao's quadratic index. *Theoretical population biology*, 70(3):237–43.
- Rivals, I., Personnaz, L., Taing, L., and Potier, M.-C. (2007). Enrichment or depletion of a go category within a class of genes: which test? *Bioinformatics*, 23(4):401–7.
- Sarkar, D. (2009). *lattice: Lattice Graphics*. R package version 0.17-26.



- Schloss, P. and Handelsman, J. (2008). A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinformatics*.
- ter Braak, C. (1985). Correspondence analysis of incidence and abundance data: Properties in terms of a unimodal response .... *Biometrics*.
- Thioulouse, J. (2010). Simultaneous analysis of a sequence of paired ecological tables: A comparison of several methods. *Annals of Applied Statistics*. (accepted for publication).
- Thioulouse, J., Simier, M., and Chessel, D. (2004). Simultaneous analysis of a sequence of paired ecological tables. *Ecology*.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, 304(5667):66–74.
- Wang, Q., Garrity, G., Tiedje, J., and Cole, J. (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and environmental microbiology*, 73(16):5261.
- Wilson, K. H., Wilson, W. J., Radosevich, J. L., DeSantis, T. Z., Viswanathan, V. S., Kuzmarski, T. A., and Andersen, G. L. (2002). High-Density Microarray of Small-Subunit Ribosomal DNA Probes. *Appl. Environ. Microbiol.*, 68(5):2535–2541.