

Bibliography

1. www.genome.gov
2. <http://www.ncbi.nlm.nih.gov/>
3. www.genomesonline.org/
4. <http://camera.calit2.net/index.php/>
5. <http://www.ncbi.nlm.nih.gov/BLAST>
6. www.mysql.com
7. <http://www.sqlobject.org/>
8. www.sqlalchemy.org
9. www.kid-templating.org
10. www.cherrypy.org
11. www.python.org
12. <http://www.bi.up.ac.za/SeqWord/sniffer/>
13. www.cbs.dtu.dk/services/GenomeAtlas/
14. <http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi>
15. SWGB FTP site [<ftp://milliways.bi.up.ac.za/SeqWord/GenomeBrowser/>]
16. www.clcbio.com
17. http://www.genomenetwork.org/resources/sequenced_genomes/genome_guide_p2.shtml

Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T. & Ikemura, T. (2003) Informatics for unveiling hidden genome signatures. *Genome Res* **13**(4), 693-702.

Achilli, A., Perego, U. A., Bravi, C. M., Coble, M. D., Kong, Q.-P., Woodward, S. R., Salas, A., Torroni, A. & Bandelt, H.-J. (2008) The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS ONE* **3**(3), e1764.

Ahmadian, A., Gharizadeh, B., Gustafsson, A. C., Sterky, F., Nyrén, P., Uhlén, M. & Lundberg, J. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. *Anal Biochem* **280**(1), 103-110.

Alland, D., Whittam, T. S., Murray, M. B., Cave, M. D., Hazbon, M. H., Dix, K., Kokoris, M., Duesterhoeft, A., Eisen, J. A., Fraser, C. M. & Fleischmann, R. D. (2003) Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *J Bacteriol* **185**(11), 3392-3399.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) Basic local alignment search tool. *J Mol Biol* **215**(3), 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17), 3389-3402.

Aluru, S. (2006) Handbook of Computational Molecular Biology, Chapman & Hall/CRC Computer & Information Science Series.

Azad, R. K. & Lawrence, J. G. (2007) Detecting laterally transferred genes: use of entropic clustering methods and genome position. *Nucleic Acids Res* **35**(14), 4629-4639.

Azad, R. K. & Lawrence, J. G. (2005) Use of artificial genomes in assessing methods for atypical gene detection. *PLoS Comput Biol* **1**(6), e56.

Badger, J. H. & Olsen, G. J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* **16**(4), 512-524.

Banu, S., Honoré, N., Saint-Joanis, B., Philpott, D., Prévost, M.-C. & Cole, S. T. (2002) Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?, *Mol Microbiol* **44**(1), 9-19.

- Bastien, O., Lespinats, S., Roy, S., Métayer, K., Fertil, B., Codani, J.-J. & Maréchal, E. (2004) Analysis of the compositional biases in *Plasmodium falciparum* genome and proteome using *Arabidopsis thaliana* as a reference. *Gene* **336**(2), 163-173.
- Becq, J., Gutierrez, M. C., Rosas-Magallanes, V., Rauzier, J., Gicquel, B., Neyrolles, O. & Deschavanne, P. (2007) Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* **24**(8), 1861-1871.
- Bergh, S. & Cole, S. T. (1994) MycDB: an integrated mycobacterial database. *Mol Microbiol* **12**(4), 517-534.
- Bohlin, J., Skjerve, E. & Ussery, D. W. (2008) Investigations of oligonucleotide usage variance within and between prokaryotes. *PLoS Comput Biol* **4**(4), e1000057.
- Brosch, R., Philipp, W. J., Stavropoulos, E., Colston, M. J., Cole, S. T. & Gordon, S. V. (1999) Genomic analysis reveals variation between *Mycobacterium tuberculosis* H37Rv and the attenuated *M. tuberculosis* H37Ra strain. *Infect Immun* **67**(11), 5768-5774.
- Brosch, R., Pym, A. S., Gordon, S. V. & Cole, S. T. (2001) The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends Microbiol* **9**(9), 452-458.
- Brosch, S., Häge, A. & Johannsen, H. S. (2002) Prognostic indicators for stuttering: the value of computer-based speech analysis. *Brain Lang* **82**(1), 75-86.
- Bulmer, M. (1998) Galtons law of ancestral heredity. *Heredity* **81**(5), 579-585.
- Camacho, L. R., Ensergueix, D., Perez, E., Gicquel, B. & Guilhot, C. (1999) Identification of a virulence gene cluster of *Mycobacterium tuberculosis* by signature-tagged transposon mutagenesis. *Mol Microbiol* **34**(2), 257-267.
- Cameron, M., Bernstein, Y. & Williams, H. E. (2007) Clustered sequence representation for fast homology search. *J Comput Biol* **14**(5), 594-614.
- Cameron, M. & Williams, H. E. (2007) Comparing compressed sequences for faster nucleotide BLAST searches. *IEEE/ACM Trans Comput Biol Bioinform* **4**(3), 349-364.
- Camus, J.-C., Pryor, M. J., Médigue, C. & Cole, S. T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* **148**(10), 2967-2973.

Catanho, M., Mascarenhas, D., Degraeve, W. & de Miranda, A. B. (2006) GenoMycDB: a database for comparative analysis of mycobacterial genes and genomes. *Genet Mol Res* **5**(1), 115-126.

Chain, P., Kurtz, S., Ohlebusch, E. & Slezak, T. (2003), An applications-focused review of comparative genomics tools: capabilities, limitations and future challenges. *Brief Bioinform* **4**(2), 105-123.

Chimara, E., Ferrazoli, L. & Leão, S. C. (2004) *Mycobacterium tuberculosis* complex differentiation using gyrB-restriction fragment length polymorphism analysis. *Mem Inst Oswaldo Cruz* **99**(7), 745-748.

Chimara, E., Giampaglia, C. M. S., Martins, M. C., da Silva Telles, M. A., Ueki, S. Y. M. & Ferrazoli, L. (2004) Molecular characterization of *Mycobacterium kansasii* isolates in the State of São Paulo between 1995-1998. *Mem Inst Oswaldo Cruz* **99**(7), 739-743.

Chung, H.-J., Jung, J. D., Park, H.-W., Kim, J.-H., Cha, H. W., Min, S. R., Jeong, W.-J. & Liu, J. R. (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Rep* **25**(12), 1369-1379.

Coenye, T. & Vandamme, P. (2004) Bacterial whole-genome sequences: minimal information and strain availability. *Microbiology* **150**(7), 2017-2018.

Coenye, T. & Vandamme, P. (2004) A genomic perspective on the relationship between the Aquificales and the epsilon-Proteobacteria. *Syst Appl Microbiol* **27**(3), 313-322.

Cole, S. T. (2002) Comparative mycobacterial genomics as a tool for drug target and antigen discovery. *Eur Respir J Suppl* **36**, 78s-86s.

Cole, S. T. (1998) Comparative mycobacterial genomics. *Curr Opin Microbiol* **1**(5), 567-571.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A., Rajandream, M. A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Taylor, K., Whitehead, S. & Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**(6685), 537-544.

- Conlon, E. M., Liu, X. S., Lieb, J. D. & Liu, J. S. (2003), Integrating regulatory motif discovery and genome-wide expression analysis. *Proc Natl Acad Sci U S A* **100**(6), 3339-3344.
- Couronne, O., Poliakov, A., Bray, N., Ishkhanov, T., Ryaboy, D., Rubin, E., Pachter, L. & Dubchak, I. (2003), Strategies and tools for whole-genome alignments. *Genome Res* **13**(1), 73-80.
- Danelishvili, L., Cirillo, S. L. G., Cirillo, J. D. & Bermudez, L. E. (2007) Virulent mycobacteria and the many aspects of macrophage uptake. *Future Microbiol* **2**(5), 461-464.
- Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**(7), 1394-1403.
- Deb, C., Daniel, J., Sirakova, T. D., Abomoelak, B., Dubey, V. S. & Kolattukudy, P. E. (2006) A novel lipase belonging to the hormone-sensitive lipase family induced under starvation to utilize stored triacylglycerol in *Mycobacterium tuberculosis*. *J Biol Chem* **281**(7), 3866-3875.
- Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O. & Salzberg, S. L. (1999) Alignment of whole genomes. *Nucleic Acids Res* **27**(11), 2369-2376.
- Deschavanne, P. J., Giron, A., Vilain, J., Fagot, G. & Fertil, B. (1999) Genomic signature: characterization and classification of species assessed by chaos game representation of sequences. *Mol Biol Evol* **16**(10), 1391-1399.
- Dolja, V. V., Kreuze, J. F. & Valkonen, J. P. T. (2006) Comparative and functional genomics of closteroviruses. *Virus Res* **117**(1), 38-51.
- Domenech, P., Pym, A. S., Cellier, M., Barry, C. E. & Cole, S. T. (2002) Inactivation of the *Mycobacterium tuberculosis* Nramp orthologue (mntH) does not affect virulence in a mouse model of tuberculosis. *FEMS Microbiol Lett* **207**(1), 81-86.
- Domenech, P., Reed, M. B. & Barry, C. E. (2005) Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. *Infect Immun* **73**(6), 3492-3501.
- Dufraigne, C., Fertil, B., Lespinats, S., Giron, A. & Deschavanne, P. (2005) Detection and characterization of horizontal transfers in prokaryotes using genomic signature. *Nucleic Acids Res* **33**(1), e6.

Ellsworth, R. E., Jamison, D. C., Touchman, J. W., Chissoe, S. L., Maduro, V. V. B., Bouffard, G. G., Dietrich, N. L., Beckstrom-Sternberg, S. M., Iyer, L. M., Weintraub, L. A., Cotton, M., Courtney, L., Edwards, J., Maupin, R., Ozersky, P., Rohlfing, T., Wohldmann, P., Miner, T., Kemp, K., Kramer, J., Korf, I., Pepin, K., Antonacci-Fulton, L., Fulton, R. S., Minx, P., Hillier, L. W., Wilson, R. K., Waterston, R. H., Miller, W. & Green, E. D. (2000) Comparative genomic sequence analysis of the human and mouse cystic fibrosis transmembrane conductance regulator genes. *Proc Natl Acad Sci U S A* **97**(3), 1172-1177.

Ermolaeva, M. D. (2001) Synonymous codon usage in bacteria. *Curr Issues Mol Biol* **3**(4), 91-97.

Filliol, I., Motiwala, A. S., Cavatore, M., Qi, W., Hazbón, M. H., del Valle, M. B., Fyfe, J., García-García, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M. I., León, C. I., Crabtree, J., Angiuoli, S., Eisenach, K. D., Durmaz, R., Joloba, M. L., Rendón, A., Sifuentes-Osornio, J., de León, A. P., Cave, M. D., Fleischmann, R., Whittam, T. S. & Alland, D. (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. *J Bacteriol* **188**(2), 759-772.

Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Ermolaeva, M., Salzberg, S. L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jr, W. R. J., Venter, J. C. & Fraser, C. M. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184**(19), 5479-5490.

Frangeul, L., Nelson, K. E., Buchrieser, C., Danchin, A., Glaser, P. & Kunst, F. (1999) Cloning and assembly strategies in microbial genome projects. *Microbiology* **145**(10), 2625-2634.

Fu, L. M. & Fu-Liu, C. S. (2007) The gene expression data of *Mycobacterium tuberculosis* based on Affymetrix gene chips provide insight into regulatory and hypothetical genes. *BMC Microbiol* **7**, 37.

Fu, L. M. & Fu-Liu, C. S. (2002) Genome comparison of *Mycobacterium tuberculosis* and other bacteria. *OMICS* **6**(2), 199-206.

Fu, L. M. & Shinnick, T. M. (2007) Genome-Wide Analysis of Intergenic Regions of *Mycobacterium tuberculosis* H37Rv Using Affymetrix GeneChips. *EURASIP J Bioinform Syst Biol*, **23054**.

Gagneux, S., Burgos, M. V., DeRiemer, K., Encisco, A., Muñoz, S., Hopewell, P. C., Small, P. M. & Pym, A. S. (2006) Impact of bacterial genetics on the transmission of isoniazid-resistant *Mycobacterium tuberculosis*. *PLoS Pathog* **2**(6), e61.

Galves, M., Quitzau, J. A. A. & Dias, Z. (2006) New strategy to detect single nucleotide polymorphisms. *Genet Mol Res* **5**(1), 143-153.

Ganesan, H., Rakitianskaia, A. S., Davenport, C. F., Tümmeler, B. & Reva, O. N. (2008) The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. *BMC Bioinformatics* **9**, 333.

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. & Hewinson, R. G. (2003), The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**(13), 7877-7882.

Gutacker, M. M., Smoot, J. C., Migliaccio, C. A. L., Ricklefs, S. M., Hua, S., Cousins, D. V., Graviss, E. A., Shashkina, E., Kreiswirth, B. N. & Musser, J. M. (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **162**(4), 1533-1543.

Guyon, F. & Guénochea, A. (2008) Comparing bacterial genomes from linear orders of patterns, *Discrete Applied Mathematics* **156**, 1251-1262.

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W. A., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M. A., Ormond, D., Doggett, J., Trueman, H. E., Mendoza, J., Bidwell, S. L., Rajandream, M.-A., Carucci, D. J., Yates, J. R., Kafatos, F. C., Janse, C. J., Barrell, B., Turner, C. M. R., Waters, A. P. & Sinden, R. E. (2005) A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**(5706), 82-86.

Hallin, P. F., Binnewies, T. T. & Ussery, D. W. (2008) The genome BLASTatlas-a GeneWiz extension for visualization of whole-genome homology. *Mol Biosyst* **4**(5), 363-371.

Harper, K. N., Liu, H., Ocampo, P. S., Steiner, B. M., Martin, A., Levert, K., Wang, D., Sutton, M. & Armelagos, G. J. (2008) The sequence of the acidic repeat protein (arp) gene differentiates venereal from nonvenereal *Treponema pallidum* subspecies, and the gene has evolved under strong

positive selection in the subspecies that causes syphilis. *FEMS Immunol Med Microbiol* **53**(3), 322-332.

Harper, K. N., Ocampo, P. S., Steiner, B. M., George, R. W., Silverman, M. S., Bolotin, S., Pillay, A., Saunders, N. J. & Armelagos, G. J. (2008) On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl Trop Dis* **2**(1), e148.

Hartmans, S., Bont, J. A. M. D. & Stackerbrandt, E. (2006) The Genus *Mycobacterium*—Nonmedical, Prokaryotes **3**, 889-918.

Hasin, Y., Avidan, N., Bercovich, D., Korczyn, A., Silman, I., Beckmann, J. S. & Sussman, J. L. (2004) A paradigm for single nucleotide polymorphism analysis: the case of the acetylcholinesterase gene. *Hum Mutat* **24**(5), 408-416.

Huelsenbeck, J. P. & Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**(8), 754-755.

Jernigan, R. W. & Baran, R. H. (2002) Pervasive properties of the genomic signature. *BMC Genomics* **3**(1), 23.

Jones, A. R., Miller, M., Aebersold, R., Apweiler, R., Ball, C. A., Brazma, A., Degreef, J., Hardy, N., Hermjakob, H., Hubbard, S. J., Hussey, P., Igra, M., Jenkins, H., Julian, R. K., Laursen, K., Oliver, S. G., Paton, N. W., Sansone, S.-A., Sarkans, U., Stoeckert, C. J., Taylor, C. F., Whetzel, P. L., White, J. A., Spellman, P. & Pizarro, A. (2007) The Functional Genomics Experiment model (FuGE): an extensible framework for standards in functional genomics. *Nat Biotechnol* **25**(10), 1127-1133.

Jones, A. R., Pizarro, A., Spellman, P., Miller, M. & Group, F. E. W. (2006) FuGE: Functional Genomics Experiment Object Model. *OMICS* **10**(2), 179-184.

Jones, C. E., Baumann, U. & Brown, A. L. (2005) Automated methods of predicting the function of biological sequences using GO and BLAST. *BMC Bioinformatics* **6**, 272.

Kamvysselis, M., Patterson, N., Birren, B., Berger, B. & Lander, E. (2003), Whole-genome Comparative Annotation and Regulatory Motif Discovery in Multiple Yeast Species, *Recomb*, 10-13.

Karlin, S. (1998) Global dinucleotide signatures and analysis of genomic heterogeneity. *Curr Opin Microbiol* **1**(5), 598-610.

- Karlin, S. & Burge, C. (1995) Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* **11**(7), 283-290.
- Karlin, S., Campbell, A. M. & Mrázek, J. (1998) Comparative DNA analysis across diverse genomes. *Annu Rev Genet* **32**, 185-225.
- Kato-Maeda, M., Bifani, P. J., Kreiswirth, B. N. & Small, P. M. (2001) The nature and consequence of genetic variability within *Mycobacterium tuberculosis*. *J Clin Invest* **107**(5), 533-537.
- Kato-Maeda, M., Rhee, J. T., Gingeras, T. R., Salamon, H., Drenkow, J., Smittipat, N. & Small, P. M. (2001) Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* **11**(4), 547-554.
- Kellis, M., Patterson, N., Birren, B., Berger, B. & Lander, E. S. (2004) Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. *J Comput Biol* **11**(2-3), 319-355.
- Kiewitz, C., Larbig, K., Klockgether, J., Weinel, C. & Tümmler, B. (2000) Monitoring genome evolution ex vivo: reversible chromosomal integration of a 106 kb plasmid at two tRNA(Lys) gene loci in sequential *Pseudomonas aeruginosa* airway isolates. *Microbiology* **146**(10), 2365-2373.
- Kiewitz, C. & Tümmler, B. (2000) Sequence diversity of *Pseudomonas aeruginosa*: impact on population structure and genome evolution. *J Bacteriol* **182**(11), 3125-3135.
- Klockgether, J., Reva, O., Larbig, K. & Tümmler, B. (2004) Sequence analysis of the mobile genome island pKLC102 of *Pseudomonas aeruginosa* C. *J Bacteriol* **186**(2), 518-534.
- Koski, L. B., Morton, R. A. & Golding, G. B. (2001) Codon bias and base composition are poor indicators of horizontally transferred genes. *Mol Biol Evol* **18**(3), 404-412.
- Kubo, T. & Newton, K. J. (2008) Angiosperm mitochondrial genomes and mutations. *Mitochondrion* **8**(1), 5-14.
- Kumar, S., Nei, M., Dudley, J. & Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9**(4) 299-306.
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L. (2004) Versatile and open software for comparing large genomes. *Genome Biol* **5**(2), R12

Lawrence, J. G., Hendrix, R. W. & Casjens, S. (2001) Where are the pseudogenes in bacterial genomes?, *Trends Microbiol* **9**(11), 535-540.

Lawrence, J. G. & Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J Mol Evol* **44**(4), 383-397.

Li, W. & Godzik, A. (2006) VISSA: a program to visualize structural features from structure sequence alignment. *Bioinformatics* **22**(7), 887-888.

Lyons, E. & Freeling, M. (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* **53**(4), 661-673.

Mallon, A. M., Platzer, M., Bate, R., Gloeckner, G., Botcherby, M. R., Nordsiek, G., Strivens, M. A., Kioschis, P., Dangel, A., Cunningham, D., Straw, R. N., Weston, P., Gilbert, M., Fernando, S., Goodall, K., Hunter, G., Greystrom, J. S., Clarke, D., Kimberley, C., Goerdes, M., Blechschmidt, K., Rump, A., Hinzmann, B., Mundy, C. R., Miller, W., Poustka, A., Herman, G. E., Rhodes, M., Denny, P., Rosenthal, A. & Brown, S. D. (2000) Comparative genome sequence analysis of the Bpa/Str region in mouse and Man. *Genome Res* **10**(6), 758-775.

Marmiesse, M., Brodin, P., Buchrieser, C., Gutierrez, C., Simoes, N., Vincent, V., Glaser, P., Cole, S. T. & Brosch, R. (2004) Macro-array and bioinformatic analyses reveal mycobacterial core genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. *Microbiology* **150**(2), 483-496.

Marri, P. R., Bannantine, J. P. & Golding, G. B. (2006) Comparative genomics of metabolic pathways in Mycobacterium species: gene duplication, gene decay and lateral gene transfer. *FEMS Microbiol Rev* **30**(6), 906-925.

McCarren, J. & Brahamsha, B. (2005) Transposon mutagenesis in a marine synechococcus strain: isolation of swimming motility mutants. *J Bacteriol* **187**(13), 4457-4462.

McDonough, K. A., Kress, Y. & Bloom, B. R. (1993), The interaction of *Mycobacterium tuberculosis* with macrophages: a study of phagolysosome fusion. *Infect Agents Dis* **2**(4), 232-235.

McDonough, K. A., Kress, Y. & Bloom, B. R. (1993), Pathogenesis of tuberculosis: interaction of *Mycobacterium tuberculosis* with macrophages. *Infect Immun* **61**(7), 2763-2773.

McEvoy, C. R. E., van Helden, P. D., Warren, R. M. & van Pittius, N. C. G. (2009) Evidence

for a rapid rate of molecular evolution at the hypervariable and immunogenic *Mycobacterium tuberculosis* PPE38 gene region. *BMC Evolutionary Biology* **9**, 237-248.

Medrano-Soto, A., Moreno-Hagelsieb, G., Vinuesa, P., Christen, J. A. & Collado-Vides, J. (2004) Successful lateral transfer requires codon usage compatibility between foreign genes and recipient genomes. *Mol Biol Evol* **21**(10), 1884-1894.

Moxon, E. R., Hood, D. W., Saunders, N. J., Schweda, E. K. H. & Richards, J. C. (2002) Functional genomics of pathogenic bacteria. *Philos Trans R Soc Lond B Biol Sci* **357**(1417), 109-116.

Mrázek, J. & Karlin, S. (1999) Detecting alien genes in bacterial genomes. *Ann N Y Acad Sci* **870**, 314-329.

Mulder, N., Rabiou, H., Jamieson, G. & Vuppu, V. (2007) Comparative analysis of microbial genomes to study unique and expanded gene families in *Mycobacterium tuberculosis*. *Infect Genet Evol*.

Muller, J., Oma, Y., Vallar, L., Friederich, E., Poch, O. & Winsor, B. (2005) Sequence and comparative genomic analysis of actin-related proteins. *Mol Biol Cell* **16**(12), 5736-5748.

Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. (2004) Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36**(7), 760-766.

Needleman, S. B. & Wunsch, C. D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**(3), 443-453.

Nouvel, L. X., Kassa-Kelembho, E., Vultos, T. D., Zandanga, G., Rauzier, J., Lafoz, C., Martin, C., Blazquez, J., Talarmin, A. & Gicquel, B. (2006) Multidrug-resistant *Mycobacterium tuberculosis*, Bangui, Central African Republic. *Emerg Infect Dis* **12**(9), 1454-1456.

Okkels, L. M., Brock, I., Follmann, F., Agger, E. M., Arend, S. M., Ottenhoff, T. H. M., Oftung, F., Rosenkrands, I. & Andersen, P. (2003), PPE protein (Rv3873) from DNA segment RD1 of *Mycobacterium tuberculosis*: strong recognition of both specific T-cell epitopes and epitopes conserved within the PPE family. *Infect Immun* **71**(11), 6116-6123.

Ozaki, K., Sato, H., Iida, A., Mizuno, H., Nakamura, T., Miyamoto, Y., Takahashi, A., Tsunoda, T., Ikegawa, S., Kamatani, N., Hori, M., Nakamura, Y. & Tanaka, T. (2006) A functional SNP in PSMA6 confers risk of myocardial infarction in the Japanese population. *Nat Genet* **38**(8), 921-925.

- Pan, X., Stein, L. & Brendel, V. (2005) SynBrowse: a synteny browser for comparative sequence analysis. *Bioinformatics* **21**(17), 3461-3468.
- Philipp, W. J., Gordon, S., Telenti, A. & Cole, S. T. (1998) Pulsed field gel electrophoresis for mycobacteria. *Methods Mol Biol* **101**, 51-63.
- Philipp, W. J., Nair, S., Guglielmi, G., Lagranderie, M., Gicquel, B. & Cole, S. T. (1996) Physical mapping of *Mycobacterium bovis* BCG pasteur reveals differences from the genome map of *Mycobacterium tuberculosis* H37Rv and from *M. bovis*. *Microbiology* **142**(11), 3135-3145.
- Philipp, W. J., Schwartz, D. C., Telenti, A. & Cole, S. T. (1998) Mycobacterial genome structure. *Electrophoresis* **19**(4), 573-576.
- van Pittius, N. C. G., Sampson, L. S., Lee, H., Kim Y., van Helden, P. D. & Warren, R. M. (2006) Evolution and expansion of the Mycobacterium tuberculosis PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. *BMC Evolutionary Biology* **6**, 95-126.
- Pizarro, A., Jones, A., Spellman, P., Miller, M., Whetzel, P. & working group, F. (2006) Extensible Framework for Standards in Functional Genomics.
- Pride, D. T. & Blaser, M. J. (2002) Concerted evolution between duplicated genetic elements in *Helicobacter pylori*. *J Mol Biol* **316**(3), 629-642.
- Pride, D. T., Meinersmann, R. J., Wassenaar, T. M. & Blaser, M. J. (2003), Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Res* **13**(2), 145-158.
- Puigbò, P., Bravo, I. G. & Garcia-Vallve, S. (2008) CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct* **3**, 38.
- Puigbò, P., Bravo, I. G. & Garcia-Vallvé, S. (2008) E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI). *BMC Bioinformatics* **9**, 65.
- Pym, A. S., Brodin, P., Brosch, R., Huerre, M. & Cole, S. T. (2002) Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. *Mol Microbiol* **46**(3), 709-717.
- Pym, A. S., Saint-Joanis, B. & Cole, S. T. (2002) Effect of katG mutations on the virulence

of *Mycobacterium tuberculosis* and the implication for transmission in humans. *Infect Immun* **70**(9), 4955-4960.

Raman, S., Hazra, R., Dascher, C. C. & Husson, R. N. (2004) Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. *J Bacteriol* **186**(19), 6605-6616.

Raman, S., Puyang, X., Cheng, T.-Y., Young, D. C., Moody, D. B. & Husson, R. N. (2006) *Mycobacterium tuberculosis* SigM positively regulates Esx secreted protein and nonribosomal peptide synthetase genes and down regulates virulence-associated surface lipid synthesis. *J Bacteriol* **188**(24), 8460-8468.

Ranjan, S., Gundu, R. K. & Ranjan, A. (2006) MycoPeronDB: a database of computationally identified operons and transcriptional units in Mycobacteria. *BMC Bioinformatics* **7** Suppl 5, S9.

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M. & Bier, E. (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* **11**(6), 1114-1125.

Reva, O. & Tümmler, B. (2008) Think big-giant genes in bacteria. *Environ Microbiol* **10**(3), 768-777.

Reva, O. N., Hallin, P. F., Willenbrock, H., Sicheritz-Ponten, T., Tümmler, B. & Ussery, D. W. (2008) Global features of the *Alcanivorax borkumensis* SK2 genome. *Environ Microbiol* **10**(3), 614-625.

Reva, O. N. & Tümmler, B. (2005) Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* **6**, 251.

Reva, O. N. & Tümmler, B. (2004) Global features of sequences of bacterial chromosomes, plasmids and phages revealed by analysis of oligonucleotide usage patterns. *BMC Bioinformatics* **5**, 90.

Reva, O. N., Weinel, C., Weinel, M., Böhm, K., Stjepandic, D., Hoheisel, J. D. & Tümmler, B. (2006) Functional genomics of stress response in *Pseudomonas putida* KT2440. *J Bacteriol* **188**(11), 4079-4092.

Rocha, E. P., Viari, A. & Danchin, A. (1998) Oligonucleotide bias in *Bacillus subtilis*: general trends and taxonomic comparisons. *Nucleic Acids Res* **26**(12), 2971-2980.

- Rodríguez, J. C., Jennings, P. A. & Melacini, G. (2002) Effect of chemical exchange on radiation damping in aqueous solutions of the osmolyte glycine. *J Am Chem Soc* **124**(22), 6240-6241.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**(12), 5463-5467.
- Saski, C., Lee, S.-B., Daniell, H., Wood, T. C., Tomkins, J., Kim, H.-G. & Jansen, R. K. (2005) Complete chloroplast genome sequence of *Gycine max* and comparative analyses with other legume genomes. *Plant Mol Biol* **59**(2), 309-322.
- Schwartz, S., Kent, W. J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R. C., Haussler, D. & Miller, W. (2003), Human-mouse alignments with BLASTZ. *Genome Res* **13**(1), 103-107.
- Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R. & Miller, W. (2000) PipMaker-a web server for aligning two genomic DNA sequences. *Genome Res* **10**(4), 577-586.
- See, D. R., Brooks, S., Nelson, J. C., Brown-Guedira, G., Friebe, B. & Gill, B. S. (2006) Gene evolution at the ends of wheat chromosomes. *Proc Natl Acad Sci U S A* **103**(11), 4162-4167.
- Sharp, P. M. & Li, W. H. (1987) The codon Adaptation Index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**(3), 1281-1295.
- Soderlund, C., Nelson, W., Shoemaker, A. & Paterson, A. (2006) SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res* **16**(9), 1159-1168.
- Srinivasachary; Dida, M. M., Gale, M. D. & Devos, K. M. (2007) Comparative analyses reveal high levels of conserved colinearity between the finger millet and rice genomes. *Theor Appl Genet* **115**(4), 489-499.
- Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. & Lewis, S. (2002) The generic genome browser: a building block for a model organism system database. *Genome Res* **12**(10), 1599-1610.
- Teeling, H., Meyerdierks, A., Bauer, M., Amann, R. & Glöckner, F. O. (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6**(9), 938-947.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput Appl Biosci* **10**(1), 19-29.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**(22), 4673-4680.

Treangen, T. J. & Messeguer, X. (2006) M-GCAT: interactively and efficiently constructing large-scale multiple genome comparison frameworks in closely related species. *BMC Bioinformatics* **7**, 433.

Tsolaki, A. G., Gagneux, S., Pym, A. S., de la Salmoniere, Y.-O. L. G., Kreiswirth, B. N., Soolingen, D. V. & Small, P. M. (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. *J Clin Microbiol* **43**(7), 3185-3191.

Ureta-Vidal, A., Ettwiller, L. & Birney, E. (2003), Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet* **4**(4), 251-262.

Vishnoi, A., Roy, R. & Bhattacharya, A. (2007) Comparative analysis of bacterial genomes: identification of divergent regions in mycobacterial strains using an anchor-based approach. *Nucleic Acids Res* **35**(11), 3654-3667.

Vishnoi, A., Srivastava, A., Roy, R. & Bhattacharya, A. (2008) MGDD: Mycobacterium tuberculosis genome divergence database. *BMC Genomics* **9**, 373-377.

Voskuil, M. I. (2004) *Mycobacterium tuberculosis* gene expression during environmental conditions associated with latency. *Tuberculosis (Edinb)* **84**(3-4), 138-143.

Wang, B. (2001) Limitations of compositional approach to identifying horizontally transferred genes. *J Mol Evol* **53**(3), 244-250.

Waterhouse, R. M., Kriventseva, E. V., Meister, S., Xi, Z., Alvarez, K. S., Bartholomay, L. C., Barillas-Mury, C., Bian, G., Blandin, S., Christensen, B. M., Dong, Y., Jiang, H., Kanost, M. R., Koutsos, A. C., Levashina, E. A., Li, J., Ligoxygakis, P., Maccallum, R. M., Mayhew, G. F., Mendes, A., Michel, K., Osta, M. A., Paskewitz, S., Shin, S. W., Vlachou, D., Wang, L., Wei, W., Zheng, L., Zou, Z., Severson, D. W., Raikhel, A. S., Kafatos, F. C., Dimopoulos, G., Zdobnov, E. M. & Christophides, G. K. (2007) Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**(5832), 1738-1743.

Waterman, M. S. (1984) Efficient sequence alignment algorithms. *J Theor Biol* **108**(3), 333-337.

Waterman, M. S., Arratia, R. & Galas, D. J. (1984) Pattern recognition in several sequences: consensus and alignment. *Bull Math Biol* **46**(4), 515-527.

Waterston, R. H., Lander, E. S. & Sulston, J. E. (2002) On the sequencing of the human genome. *Proc Natl Acad Sci U S A* **99**(6), 3712-3716.

Weinel, C., Nelson, K. E. & Tümmler, B. (2002) Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environ Microbiol* **4**(12), 809-818.

White, P. S., Kwok, P.-Y., Oefner, P. & Brookes, A. J. (2001) 3rd International Meeting on Single Nucleotide Polymorphism and Complex Genome Analyses: SNPs: Some Notable Progress, *European Journal of Human Genetics* **9**, 316-318.

Zhang, H., Duan, X., Yuan, Z., Li, W., Zhou, G., Zhou, Q., Bing, L., Min, F., Li, X. & Xie, Y. (2006) Chromosomal aberrations induced by (12)C6+ ions and 60Co gamma-rays in mouse immature oocytes. *Mutat Res* **595**(1-2), 37-41.

Zheng, H., Lu, L., Wang, B., Pu, S., Zhang, X., Zhu, G., Shi, W., Zhang, L., Wang, H., Wang, S., Zhao, G. & Zhang, Y. (2008) Genetic Basis of Virulence Attenuation Revealed by Comparative Genomic Analysis of *Mycobacterium tuberculosis* Strain H37Ra versus H37Rv, *PLoS One* **3**, 2375.