

BIBLIOGRAPHY

Bibliography

- [1] <http://www.genome.gov/10001772>
- [2] <http://www.nlm.nih.gov/medlineplus/magazine/issues/summer13/articles/summer13pg15.html> - Understanding the Human Genome Project – A Fact Sheet
- [3] <http://spectrum.ieee.org/biomedical/devices/the-dna-data-deluge> - The DNA Data Deluge, Michael C. Schatz, Ben Langmead, IEEE Spectrum, Jun 2013
- [4] <http://en.wikipedia.org/wiki/Moore'sLaw>
- [5] Andrew S. Tanenbaum et al., Distributed Systems: Principles and Paradigms, NJ, Prentice-Hall, 2003.
- [6] <http://www.ncbi.nlm.nih.gov>
- [7] <http://www.genome.gov/glossary/index.cfm?id=17>
- [8] <http://www.fhwa.dot.gov/ohim/handbook/chap7.pd> - Travel Time Data Collection Handbook
- [9] <http://www.genome.gov/glossary/index.cfm?id=193>
- [10] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, Wavelet Toolbox User's Guide Version 1, The MathWorks Inc, 1996-97 (<http://www.mathworks.com>)
- [11] Ajay D. Kshemkalyani et al., Distributed Computing: Principles, Algorithms and Systems, UK, Cambridge University Press, 2008.
- [12] http://docs.oracle.com/cd/A57673_01/DOC/server/doc/SD173/ch1.htm
- [13] <http://www.oracle.com>
- [14] Ronaghi Mostafa, Pyrosequencing Sheds Light on DNA Sequencing, Genome Research 11, 2001, 3-11

- [15] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software, Addison Wesley, USA, 1994
- [16] <http://www.biojava.org>
- [17] Guy Steele, Gilad Bracha, Bill Joy, James Gosling, The Java Language Specifications, Sun Microsystems
- [18] Bryon Bergeron, Bioinformatics Computing, Pearson Education, Inc. 2003
- [19] <http://www.genome.gov/glossary/index.cfm?id=532>
- [20] <http://www.genome.gov/glossary/index.cfm?id=90>
- [21] <http://www.genome.gov/glossary/index.cfm?id=33>
- [22] <http://www.genome.gov/glossary/index.cfm?id=70>
- [23] <http://www.genome.gov/glossary/index.cfm?id=48>
- [24] <http://ghr.nlm.nih.gov/glossary=rna>
- [25] http://upload.wikimedia.org/wikipedia/commons/3/37/Difference_DNA_RNA-EN.svg
- [26] <http://ghr.nlm.nih.gov/glossary=protein>
- [27] Christopher Moretti, Michael Olson, Scott Emrich, and Douglas Thain, Highly Scalable Genome Assembly on Campus Grids, Portland, Oregon, USA, MTAGS '09 2009 ACM 978-1-60558-714-1/09/11
- [28] <http://www.genome.gov/glossary/index.cfm?id=39>
- [29] <http://upload.wikimedia.org>
- [30] www.rpi.edu
- [31] <http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>
- [32] Analyzing DNA Sequences and DNA Barcoding, Northwest Association for Biomedical Research 262
- [33] www.454.com
- [34] www.illumina.com
- [35] www.appliedbiosystems.com

- [36] Susanne M. D. Goldberg, Justin Johnson, Dana Busam, Tamara Feldblyum et. al., A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes
- [37] www.helicosbio.com
- [38] www.pacificbiosciences.com
- [39] www.polonator.org
- [40] P. Prandoni , M. Vetterli, Signal Processing For Communication, EPFL Press, 2008
- [41] <http://www.cdeep.iitb.ac.in/nptel>
- [42] <http://nptel.iitm.ac.in/courses/Webcourse-contents/IIT-KANPUR/>
- [43] Lecture Notes: Stephen P. Boyd, www.stanford.edu/~boyd/ee102/signals.pdf
- [44] James S. Walker, A Primer on Wavelets and their Scientific Applications, Taylor & Francis Group, 2008
- [45] Alan v. Oppenheim, Ronald W. Schaffer, John R. Buck, Discrete-Time Signal Processing, Prentice Hall, 2nd Edition, 1998
- [46] Leland Wilkinson, Statistics and Computing - The Grammar of Graphics, Springer, 28-Jan-2006
- [47] Lecture Notes: Professor David Heeger, Signals, Linear Systems, and Convolution, September 26, 2000
- [48] <http://www.r2labs.org/references/Convolution.pdf>
- [49] Lecture Notes : Willard Miller, Introduction to the Mathematics of Wavelets, May 3, 2006
- [50] Robi Polikar, The Wavelet Tutorial, Rowan University, 2001.
- [51] Amara Graps, An Introduction to Wavelets
- [52] <http://www.mathworks.com> [Matlab documentation].
- [53] J.K. Meher, M. R. Panigrahi, G. N. Dash, P. K. Meher, "Wavelet Based Lossless DNA Sequence Compression For Faster Detection Of

- Eukaryotic Protein Coding Regions" I.J. Image, Graphics And Signal Processing 2012, 47-53
- [54] Ingrid Daubechies, "Ten Lectures On Wavelets", 1992
 - [55] G. H. Watson, 'Application of Mathematical Signal Processing Techniques to Mission Systems", Paper presented at the RTO SC1 Lecture Series on held in Kiiln, Germany, 1-2 November 1999; Paris, France, 4-5 November 1999; Monterey, USA, 9-10 November 1999, and published in RTO EN-7. (Introduction to Wavelet Analysis)
 - [56] D. Lee Fugal, Conceptual Wavelets in Digital Signal Processing, Space & Signals Technologies LLC, 2009 (www.ConceptualWavelets.com)
 - [57] Charu C. Aggarwal, "On The Use Of Wavelet Decomposition For String Classification," Springer - Data Mining And Knowledge Discovery, 10, 117–139, 2005
 - [58] Web Content: Musawir Ali, An Introduction to Wavelets and the Haar Transform (www.cs.ucf.edu/~mali/haar/)
 - [59] http://courses.ae.utexas.edu/ase463q/design_pages/fall02/wavelet/4_wavelet_theory.htm
 - [60] Moharir P.S., "Pattern recognition transforms," New York: Wiley, 1992
 - [61] Radomir S. Stankovic, Bogdan J. Falkowski, "The Haar wavelet transform: its status and achievements," Elsevier
 - [62] Ruiz G, Michell JA, Buron A, "Switch-level fault detection and diagnosis environment for MOS digital circuits using spectral techniques," IEEE Proc Part E, 1992; 139(4):293–307
 - [63] Ingrid Daubechies, "Orthonormal Bases of Compactly Supported Wavelets," Comm. Pure Appl. Math, Vol 41, 1988, pp. 906-966
 - [64] Dimitris Sacharidis, "Constructing Optimal Wavelet Synopses", Proceedings of the 2006 International Conference on Current Trends in Database Technology EBDT '06, Pg 97-104, 2006 10.1007/11896548_10

- [65] D.Taubman, M. Marcellin, Kluwer, JPEG2000: Image Compression Fundamentals, Standards and Practice", Academic Publishers, 2001.
- [66] T. Ebrahimi, F. Pereira, "The MPEG-4 Book", Prentice Hall, 2002.
- [67] http://ai.stanford.edu/~serafim/CS374_2011/presentations/lecture6.pdf - Source: Stein 2010
- [68] Ziv J, Lempel A. A universal algorithm for sequential data compression, IEEE Transaction of Information Theory 1977; 23(3):337-343.
- [69] Burrows M, Wheeler D, A block sorting lossless data compression algorithm, Tech. Rep. 124, Digital Equipment Corporation 1994.
- [70] Huffman D.A., A Method for the Construction of Minimum-Redundancy Codes, Proceedings of the I.R.E., September 1952, pp 1098–1102. Huffman's original article
- [71] Paolo Ferragina, Giovanni Manzini, Veli M`akinen, and Gonzalo Navarro, "An Alphabet-Friendly FM-index?" In International Symposium on String Processing and Information Retrieval (SPIRE), (2004) pages 150–160
- [72] O`guzhan M, Vitter J S, Xu B, Efficient Maximal Repeat Finding Using the Burrows-Wheeler Transform and Wavelet Tree, IEEE/ACM Transactions On Computational Biology And Bioinformatics.
- [73] Congmao Wang, Dabing Zhang, A novel compression tool for efficient storage of genome resequencing data
- [74] Hisahiko Sato, Takashi Yoshioka, Akihiko Konagaya, Tetsuro Toyoda, DNA Data Compression in the Post Genome Era, Genome Informatics 12: 512–514 (2001)
- [75] Ana Balevic, Lars Rockstroh, Marek Wroblewski, and Sven Simon, Using Arithmetic Coding for Reduction of Resulting Simulation Data Size on Massively Parallel GPGPUs
- [76] Sayood, K., Lossless Compression Handbook, Academic Press (2003)

- [77] Amir Said, Introduction to Arithmetic Coding, Theory and Practice, Academic Press, 2004.
- [78] Toshiko Matsumoto Kunihiro Sadakane, Hiroshi Imai, Biological Sequence Compression Algorithms, Genome Informatics 11: 43–52 (2000) 43
- [79] Chen, X., Kwong, S., and Li, M., A compression algorithm for DNA sequences and its applications in genome comparison, Genome Informatics, 10:52–61, 1999
- [80] Grumbach S., Tahi F., A new challenge for compression algorithms, Genetic sequences, Information Processing and Management, 30:875–886, 1994
- [81] Kenny Daily, Paul Rigor, Scott Christley, Xiaohui Xie, Pierre Baldi, Data structures and compression algorithms for high-throughput sequencing technologies, BMC Bioinformatics 2010, 11:514
- [82] Samantha Woodward, A Critical Analysis of DNA Data Compression Methods BIOC 218, Winter 2011-2012
- [83] Bernhard Haubold, Thomas Wiehe, How repetitive are genomes?, BMC Bioinformatics 2006, 7:541 doi : 10.1186/1471-2105-7-541
- [84] Ziv J, Lempel A. A universal algorithm for sequential data compression, IEEE Transaction of Information Theory 1977; 23(3):337-343
- [85] Markus Hsi-Yang Fritz, Rasko Leinonen, Guy Cochrane, and Ewan Birney, Efficient storage of high throughput DNA sequencing data using reference-based compression, Genome Research, 21:734–740_2011 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/11
- [86] W Timothy, J White, Michael D Hendy, Compressing DNA sequence databases with coil, BMC Bioinformatics 2008, 9:242 doi:10.1186/1471-2105-9-242

- [87] Al Junid, S. A. M. Tahir, N.M. Haron, M.A. Abd Majid, Z., Idros, M.F. Osman, F.N., Development and Implementation of Novel Data Compression Technique for Accelerate DNA Sequence Alignment Based on Smith–Waterman Algorithm, IJSSST, Vol. 11, No. 3 ISSN: 34 1473-804x online, 1473-8031
- [88] S. Nair, S. P. Sreenathan, “A Coding Measure Scheme Employing Electron-Ion Interaction Pseudopotential (EIIP),” *Bioinformation*, Vol. 1, No. 6, 2006, pp. 197-202
- [89] Aghili Alireza, A Agrawal, D El Abbadi A, “Sequence Similarity Search Using Discrete Fourier and Wavelet Transformation Techniques,” Postprints, UC Santa Barbara, 2005 <http://www.escholarship.org/uc/item/9w7094b9>
- [90] Sanger, F. et al. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* 1977, 74: 5463–5467
- [91] Ronaghi, M. Pyrosequencing Sheds Light on DNA Sequencing *Genome Res.* 2001, 11: 3-11
- [92] Metzker, M. L. et al. Emerging Technologies in DNA Sequencing. *Genome Res.* 2005, 15: 1767-1776
- [93] Dong, H. et al. Artificial duplicate reads in sequencing data of 454 Genome Sequencer FLX System *Acta Biochim Biophys Sin* 2011, 43: Issue6: 496–500
- [94] Gomez-Alvarez V. et al. Systematic artifacts in metagenomes from complex microbial communities. *ISME J*, 2009, 3 : 1314–1317
- [95] Benson G., “Tandem repeats finder: a program to analyze DNA sequences,” *PMC, PubMed Nucleic Acids Research*, Vol. 27. No. 2 1999; 27:573–80
- [96] Angelika Merkel, Neil Gemmell, “Detecting short tandem repeats from genome data: opening the software black box,” *Briefings in Bioinformatics Advance Access*, July 10, 2008
- [97] Goldstein DB, Schlotterer C , “Microsatellites: Evolution and Applications,” Oxford: Oxford University Press,, 1999

- [98] Pearson CE, Edamura KN, Cleary JD, "Repeat instability: mechanisms of dynamic mutations," *Nat Rev Genet*, 2005; 6: 729–42
- [99] Kashi Y, King DG, "Simple sequence repeats as advantageous mutators in evolution," *Trends in Genetics*, 2006;22: 253–9
- [100] Moxon ER, Wills C., "DNA microsatellites: agents of evolution?" *Sci Am*, 1999; 280:94–9
- [101] Alex Van Belkum, Stewart Scherer, Loek Van Alphen, And Henri Verbrugh , "Short-Sequence DNA Repeats in Prokaryotic Genomes," *American Society for Microbiology-MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS*, June 1998, p.275–293 Vol.62, No. 2
- [102] Jeffreys, A. J., V. Wilson, and S. L. Thein, "Hypervariable "minisatellite" regions in human DNA," *Nature*, 1985 314:67–73
- [103] Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White, "Variable number of tandem repeat (VNTR) markers for human gene mapping ," *Science*, 1987 235:1616–1622
- [104] Thiel T, Michalek W, Varshney RK, "Exploiting EST databases for the development and characterization of gene derived SSR-markers in barley (*Hordeum vulgare* L.)," *TheorAppl Genet*, 2003; 106:411–22
- [105] Smit AFA, Green P, "RepeatMasker" <http://www.repeatmasker.org>, 1996
- [106] Goffeau A, Barrell BG, Bussey H, "Life with 6000 genes," *Science*, 1996;274:546–67
- [107] Surya Saha and Susan Bridges and Zenaida V. Magbanua and Daniel G. Peterson, "Computational Approaches and Tools Used in Identification of Dispersed Repetitive DNA Sequences," Springer, 2008
- [108] Li YC, Korol AB, Fahima T, "Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review," *Mol Ecol*, 2002; 11:2453–65

- [109] Morgante M, Hanafey M, Powell W, "Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes," *Nat Genet*, 2002; 30:194–200
- [110] Lim S, Notley-McRobb L, Lim M, "A comparison of the nature and abundance of microsatellites in 14 fungal genomes," *Fungal Genet Biol*, 2004; 41:1025–36
- [111] Stefan Kurtz, Jomuna V. Choudhuri, Enno Ohlebusch, "REPuter: the manifold applications of repeat analysis on a genomic scale," *Nucleic Acids Research*, 2001 Vol 29 No. 22 pg 4633 – 4642
- [112] Chris Abajian, "Sputnik - DNA microsatellite repeat search utility," <http://www.espressosoftware.com/sputnik/>
- [113] Surya Saha and Susan Bridges and Zenaida V. Magbanua and Daniel G. Peterson, "Computational Approaches and Tools Used in Identification of Dispersed Repetitive DNA Sequences," Springer, 2008
- [114] Schmidt J. P., "All Highest Scoring Paths in Weighted Grid Graphs and Their Application to Finding All Approximate Repeats in Strings," *Siam J. Comput*, 1998, Vol. 27 Issue 4, 972-992
- [115] Trevor J Pemberton, Conner Sandefur, Mattias Jakobsson and Noah A Rosenberg, Sequence determinants of human microsatellite variability, *BMC Genomics* 2009, 10:612 doi:10.1186/1471-2164-10-612
- [116] Jennifer Kwan, Benjamin Kwan, Hon Kwan, Spectral Analysis of Numerical Exon and Intron Sequences, *IEEE International Conference on Bioinformatics and Biomedicine Workshop*, 978-1-4244-8302-0/10/\$26.00 ©2010 IEEE pp 876-877
- [117] M. Akhtar, J. Epps, and E. Ambikairajah, "Signal processing in sequence analysis: Advances in eukaryotic gene prediction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 310-321, June 2008.

- [118] P. D. Cristea, "Genetic signal representation and analysis," in Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, vol. 4623, January 2002, pp. 77-84.
- [119] N. Chakravarthy, A. Spanias, L. D. Lasemidis, and K. Tsakalis, "Autoregressive modeling and feature analysis of DNA sequences," EURASIP Journal of Genomic Signal Processing, vol. 1, pp. 13-28, January 2004.
- [120] P. D. Cristea, "Conversion of nucleotides sequences into genomic signals," J. Cell. Mol. Med., vol. 6, pp. 279-303, April-June 2002.
- [121] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, "New approaches to genome sequence analysis based on digital signal processing," in Proc. of IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS), October 2002, pp. 1-4.
- [122] H. E. Stanley, S. V. Buldyrev, A. L. Goldberger, Z. D. Goldberger, S. Havlin, S. M. Ossadnik, C.-K. Peng, and M. Simmons, "Statistical mechanics in biology: How ubiquitous are long-range correlations?" Physica A, vol. 205, pp. 214-253, April 1994
- [123] T. Holden, R. Subramaniam, R. Sullivan, E. Cheng, C. Sneider, G. Tremberger, Jr. A. Flamholz, D. H. Leiberman, and T. D. Cheung, "ATCG nucleotide fluctuation of Deinococcus radiodurans radiation genes," in Proc. of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, vol. 6694, August 2007, pp. 669417-1 to 669417-10.
- [124] Muneer Ahmad, Azween Abdullah, Khalid Buragga, A Novel Optimized Approach for Gene Identification in DNA Sequences
- [125] Dimitris Anastassiou, Genomic Signal Processing, IEEE SIGNAL PROCESSING MAGAZINE, vol 18, pp 8-20 1053-5888/01/2001 IEEE
- [126] R. Ranawana, V. Palade, A Neural network based multi-classifier system for gene identification in DNA sequence. Neural Computing and Applications, 2005, 14(2):122-131

[127] B. Demeler, G. W. Zhou, Neural network optimization for E.coli promoter prediction, *Nucleic Acids Res.*, 1991, 19(7):1539-1599.