Bibliography

[1] Gwas central, http://www.gwascentral.org/.

[2] Python for beginners, http://www.python.org/doc/Intros.html.

[3] The python language website, http://www.python.org/.

[4] Python tutor, visualization of code execution, http://pythontutor.com/.

[5] The python tutorial, https://docs.python.org/3/tutorial/.

[6] Computational Methods in Molecular Biology, Elsevier Science, 1998.

[7] Biopython: freely available python tools for computational molecular biology and bioinformatics, Bioinformatics 25 (11) (2009) 1422–1423.

[8] Biopython tutorial and cookbook, http://biopython.org/DIST/docs/tutorial/Tutorial.html. (Last update Nov 23, 2016).

[9] Alfred V. Aho, John E. Hopcroft, The Design and Analysis of Computer Algorithms, 1st edition, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1974.

[10] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 4th edition, Garland Science, New York, USA, 2002.

[11] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, David J. Lipman, Basic local alignment search tool, Journal of Molecular Biology 215 (3) (1990) 403–410.

[12] Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, David J. Lipman, Gapped blast and psi-blast: a new generation of protein database search programs, Nucleic Acids Research 25 (17) (1997) 3389–3402.

[13] R. Andersson, et al., An atlas of active enhancers across human cell types and tissues, Nature 507 (7493) (Mar 2014) 455–461.

[14] K. Asai, S. Hayamizu, K. Handa, Prediction of protein secondary structure by the hidden Markov model, Computer Applications in the Biosciences 9 (2) (Apr 1993) 141–146.

[15] A. Auton, et al., A global reference for human genetic variation, Nature 526 (7571) (Oct 2015) 68–74.

[16] Gary D. Bader, Christopher W.V. Hogue, Analyzing yeast protein-protein interaction data obtained from different sources, Nature Biotechnology 20 (10) (2002) 991–997.

[17] T.L. Bailey, Discovering sequence motifs, Methods in Molecular Biology 452 (2008) 231–251.

[18] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proceedings. International Conference on Intelligent Systems for Molecular Biology 2 (1994) 28–36.

[19] Pierre Baldi, Søren Brunak, Bioinformatics: The Machine Learning Approach, 2nd edition, MIT Press, Cambridge, MA, USA, 2001.

[20] Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, et al., Spades: a new genome assembly algorithm and its applications to single-cell sequencing, Journal of Computational Biology 19 (5) (2012) 455–477.

[21] Albert-László Barabási, Réka Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.

[22] Albert-László Barabási, Natali Gulbahce, Joseph Loscalzo, Network medicine: a network-based approach to human disease, Nature Reviews Genetics 12 (1) (2011) 56–68.

[23] Albert-Laszlo Barabási, Zoltan N. Oltvai, Network biology: understanding the cell's functional organization, Nature Reviews Genetics 5 (2) (2004) 101–113.

[24] N.L. Barbosa-Morais, M. Irimia, Q. Pan, H.Y. Xiong, S. Gueroussov, L.J. Lee, V. Slobodeniuc, C. Kutter, S. Watt, R. Colak, T. Kim, C.M. Misquitta-Ali, M.D. Wilson, P.M. Kim, D.T. Odom, B.J. Frey, B.J. Blencowe, The evolutionary landscape of alternative splicing in vertebrate species, Science 338 (6114) (Dec 2012) 1587–1593.

[25] Sebastian Bassi, Python for Bioinformatics, CRC Press, 2016.

[26] T. Beck, R.K. Hastings, S. Gollapudi, R.C. Free, A.J. Brookes, GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies, European Journal of Human Genetics 22 (7) (Jul 2014) 949–952.

[27] E. Birney, et al., Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project, Nature 447 (7146) (Jun 2007) 799–816.

[28] Hans-Joachim Böckenhauer, Dirk Bongartz, Algorithmic Aspects of Bioinformatics, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[29] Robert S. Boyer, J. Strother Moore, A fast string searching algorithm, Communications of the ACM 20 (10) (October 1977) 762–772.

[30] J. Buhler, M. Tompa, Finding motifs using random projections, Journal of Computational Biology 9 (2) (2002) 225–242.

[31] C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA, Journal of Molecular Biology 268 (1) (Apr 1997) 78–94.

[32] Michael Burrows, David J. Wheeler, A Block-Sorting Lossless Data Compression Algorithm, 1994.

[33] P. Carninci, et al., The transcriptional landscape of the mammalian genome, Science 309 (5740) (Sep 2005) 1559–1563.

[34] Humberto Carrillo, David Lipman, The multiple sequence alignment problem in biology, SIAM Journal on Applied Mathematics 48 (5) (1988) 1073–1082.

[35] Phillip Compeau, Pavel Pevzner, Bioinformatics Algorithms: An Active Learning Approach, Active Learning Publishers, 2015.

[36] Community Content Contributions, Boundless biology, https://www.boundless.com/biology/textbooks/boundless-biology-textbook/, February 2017.

[37] G.M. Cooper, The Cell: A Molecular Approach, 2nd edition, Sinauer Associates, Sunderland, MA, USA, 2000.

[38] Thomas H. Cormen, Clifford Stein, Ronald L. Rivest, Charles E. Leiserson, Introduction to Algorithms, 2nd edition, McGraw-Hill Higher Education, 2001.

[39] Maxime Crochemore, Christophe Hancart, Thierry Lecroq, Algorithms on Strings, Cambridge University Press, 2007.

[40] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Research 14 (6) (Jun 2004) 1188–1190.

[41] M.K. Das, H.K. Dai, A survey of DNA motif finding algorithms, BMC Bioinformatics 8 (Suppl 7) (Nov 2007) S21.

[42] Sanjoy Dasgupta, Christos H. Papadimitriou, Umesh Vazirani, Algorithms, McGraw-Hill, Inc., 2006.

[43] Margaret O. Dayhoff, Atlas of Protein Sequence and Structure, 1965.

[44] N. de Bruijn, A combinatorial problem, Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 49 (7) (1946) 758–764.

[45] Rene De La Briandais, File searching using variable length keys, in: Papers Presented at the March 3–5, 1959, Western Joint Computer Conference, ACM, 1959, pp. 295–298.

[46] T. Derrien, R. Johnson, G. Bussotti, A. Tanzer, S. Djebali, H. Tilgner, G. Guernec, D. Martin, A. Merkel, D.G. Knowles, J. Lagarde, L. Veeravalli, X. Ruan, Y. Ruan, T. Lassmann, P. Carninci, J.B. Brown, L. Lipovich, J.M. Gonzalez, M. Thomas, C.A. Davis, R. Shiekhattar, T.R. Gingeras, T.J. Hubbard, C. Notredame, J. Harrow, R. Guigo, The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression, Genome Research 22 (9) (Sep 2012) 1775–1789.
[47] P. D'haeseleer, What are DNA sequence motifs? Nature Biotechnology 24 (4) (Apr 2006) 423–425.
[48] Reinhard Diestel, Graph Theory, 3rd edition, Graduate Texts in Mathematics, vol. 173, Springer, 2005.
[49] I. Dunham, et al., An integrated encyclopedia of DNA elements in the human genome, Nature 489 (7414) (Sep 2012) 57–74.
[50] R. Durbin, S. Eddy, A. Krogh, G. Mitchison, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge University Press, 1998.
[51] J.R. Ecker, W.A. Bickmore, I. Barroso, J.K. Pritchard, Y. Gilad, E. Segal, Genomics: ENCODE explained, Nature 489 (7414) (Sep 2012) 52–55.
[52] S.R. Eddy, Multiple alignment using hidden Markov models, Proceedings. International Conference on Intelligent Systems for Molecular Biology 3 (1995) 114–120.
[53] S.R. Eddy, Hidden Markov models, Current Opinion in Structural Biology 6 (3) (Jun 1996) 361–365.
[54] S.R. Eddy, What is a hidden Markov model? Nature Biotechnology 22 (10) (Oct 2004) 1315–1316.
[55] S.R. Eddy, A probabilistic model of local sequence alignment that simplifies statistical significance estimation, PLoS Computational Biology 4 (5) (May 2008) e1000069.
[56] S.R. Eddy, A new generation of homology search tools based on probabilistic inference, Genome Informatics 23 (1) (Oct 2009) 205–211.
[57] S.R. Eddy, Accelerated profile HMM searches, PLoS Computational Biology 7 (10) (Oct 2011) e1002195.
[58] Robert C. Edgar, Muscle: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32 (5) (2004) 1792–1797.
[59] E. Eskin, P.A. Pevzner, Finding composite regulatory patterns in DNA sequences, Bioinformatics 18 (Suppl 1) (2002) S354–S363.
[60] Leonhard Euler, Solutio problematis ad geometriam situs pertinentis, Commentarii Academiae Scientiarum Petropolitanae 8 (1741) 128–140.
[61] Even Shimon, Graph Algorithms, 2nd edition, Cambridge University Press, New York, NY, USA, 2011.
[62] Joseph Felsenstein, Inferring Phylogenies, vol. 2, Sinauer Associates, Sunderland, MA, 2004.
[63] Da-Fei Feng, Russell F. Doolittle, Progressive sequence alignment as a prerequisitetto correct phylogenetic trees, Journal of Molecular Evolution 25 (4) (1987) 351–360.
[64] Paolo Ferragina, Giovanni Manzini, Opportunistic data structures with applications, in: Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on, IEEE, 2000, pp. 390–398.
[65] P.G. Ferreira, P.J. Azevedo, Evaluating deterministic motif significance measures in protein databases, Algorithms for Molecular Biology 2 (Dec 2007) 16.
[66] Jeffrey E.F. Friedl, Mastering Regular Expressions, 2nd edition, O'Reilly & Associates, Inc., Sebastopol, CA, USA, 2002.
[67] Emanuel Gonçalves, Joachim Bucher, Anke Ryll, Jens Niklas, Klaus Mauch, Steffen Klamt, Miguel Rocha, Julio Saez-Rodriguez, Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models, Molecular BioSystems 9 (7) (2013) 1576–1583.
[68] M. Gribskov, S. Veretnik, Identification of sequence pattern with profile analysis, Methods in Enzymology 266 (1996) 198–212.
[69] Stéphane Guindon, Jean-François Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, Olivier Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Systematic Biology 59 (3) (2010) 307–321.
[70] Dan Gusfield, Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology, 1st edition, Cambridge University Press, May 1997.
[71] Frank Harary, Graph Theory, Addison-Wesley, 1972.

[72] J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo, T.J. Hubbard, GENCODE: the reference human genome annotation for The ENCODE Project, Genome Research 22 (9) (Sep 2012) 1760–1774.

[73] D. Haussler, A. Krogh, I.S. Mian, K. Sjolander, Protein modeling using hidden Markov models: analysis of globins, in: Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences, IEEE, IEEE, 1993, pp. 792–802.

[74] Steven Henikoff, Jorja G. Henikoff, Amino acid substitution matrices from protein blocks, Proceedings of the National Academy of Sciences 89 (22) (1992) 10915–10919.

[75] G.Z. Hertz, G.W. Hartzell, G.D. Stormo, Identification of consensus patterns in unaligned DNA sequences known to be functionally related, Computer Applications in the Biosciences 6 (2) (Apr 1990) 81–92.

[76] Carl Hierholzer, Über die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechung zu umfahren, Mathematische Annalen 6 (1) (1873) 30–32.

[77] Desmond G. Higgins, Paul M. Sharp, Clustal: a package for performing multiple sequence alignment on a microcomputer, Gene 73 (1) (1988) 237–244.

[78] John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman, Introduction to Automata Theory, Languages, and Computation, 3rd edition, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2006.

[79] Michael Huerta, Gregory Downing, Florence Haseltine, Belinda Seto, Yuan Liu, Nih Working Definition of Bioinformatics and Computational Biology, US National Institute of Health, 2000.

[80] Ramana M. Idury, Michael S. Waterman, A new algorithm for DNA sequence assembly, Journal of Computational Biology 2 (2) (1995) 291–306.

[81] Jan Ihmels, Gilgi Friedlander, Sven Bergmann, Ofer Sarig, Yaniv Ziv, Naama Barkai, Revealing modular organization in the yeast transcriptional network, Nature Genetics 31 (4) (2002) 370.

[82] Anantharaman Narayana Iyer, pyhmm, https://github.com/ananthpn/pyhmm. (Retrieved October 2017).

[83] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks, Nature 407 (6804) (2000) 651–654.

[84] N.C. Jones, P. Pevzner, An Introduction to Bioinformatics Algorithms, A Bradford book, London, 2004.

[85] Daniel Jurafsky, James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 1st edition, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2000.

[86] K. Karplus, K. Sjolander, C. Barrett, M. Cline, D. Haussler, R. Hughey, L. Holm, C. Sander, Predicting protein structure using hidden Markov models, Proteins 29 (Suppl 1) (1997) 134–139.

[87] Kazutaka Katoh, Kazuharu Misawa, Kei-ichi Kuma, Takashi Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform, Nucleic Acids Research 30 (14) (2002) 3059–3066.

[88] U. Keich, P.A. Pevzner, Subtle motifs: defining the limits of motif finding algorithms, Bioinformatics 18 (10) (Oct 2002) 1382–1390.

[89] H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and function, Nature Reviews Genetics 11 (5) (May 2010) 345–355.

[90] A. Krogh, Two methods for improving performance of an HMM and their application for gene finding, Proceedings. International Conference on Intelligent Systems for Molecular Biology 5 (1997) 179–186.

[91] A. Krogh, M. Brown, I.S. Mian, K. Sjolander, D. Haussler, Hidden Markov models in computational biology. Applications to protein modeling, Journal of Molecular Biology 235 (5) (Feb 1994) 1501–1531.

[92] Eric S. Lander, Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, et al., Initial sequencing and analysis of the human genome, Nature 409 (6822) (2001) 860–921.

[93] Langmead Ben, Steven L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nature Methods 9 (4) (2012) 357–359.

[94] C.E. Lawrence, S.F. Altschul, M.S. Boguski, J.S. Liu, A.F. Neuwald, J.C. Wootton, Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment, Science 262 (5131) (Oct 1993) 208–214.

[95] M. Lek, et al., Analysis of protein-coding genetic variation in 60,706 humans, Nature 536 (7616) (08, 2016) 285–291.

[96] H.C. Leung, F.Y. Chin, Algorithms for challenging motif problems, Journal of Bioinformatics and Computational Biology 4 (1) (Feb 2006) 43–58.

[97] Vladimir I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals, Soviet Physics Doklady 10 (1966) 707–710.

[98] S. Levy, G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, N. Axelrod, J. Huang, E.F. Kirkness, G. Denisov, Y. Lin, J.R. MacDonald, A.W. Pang, M. Shago, T.B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S.A. Kravitz, D.A. Busam, K.Y. Beeson, T.C. McIntosh, K.A. Remington, J.F. Abril, J. Gill, J. Borman, Y.H. Rogers, M.E. Frazier, S.W. Scherer, R.L. Strausberg, J.C. Venter, The diploid genome sequence of an individual human, PLoS Biology 5 (10) (Sep 2007) e254.

[99] Dinghua Li, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, Bioinformatics 31 (10) (2015) 1674–1676.

[100] Heng Li, Richard Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics 25 (14) (2009) 1754–1760.

[101] N. Li, M. Tompa, Analysis of computational approaches for motif discovery, Algorithms for Molecular Biology 1 (May 2006) 8.

[102] Ruiqiang Li, Yingrui Li, Karsten Kristiansen, Jun Wang, SOAP: short oligonucleotide alignment program, Bioinformatics 24 (5) (2008) 713–714.

[103] David J. Lipman, Stephen F. Altschul, John D. Kececioglu, A tool for multiple sequence alignment, Proceedings of the National Academy of Sciences 86 (12) (1989) 4412–4415.

[104] Ruibang Luo, Binghang Liu, Yinlong Xie, Zhenyu Li, Weihua Huang, Jianying Yuan, Guangzhu He, Yanxiang Chen, Qi Pan, Yunjie Liu, et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, Gigascience 1 (1) (2012) 18.

[105] Nicholas M. Luscombe, Dov Greenbaum, Mark Gerstein, et al., What is bioinformatics? A proposed definition and overview of the field, Methods of Information in Medicine 40 (4) (2001) 346–358.

[106] Daniel Machado, Rafael S. Costa, Miguel Rocha, Eugénio C. Ferreira, Bruce Tidor, Isabel Rocha, Modeling formalisms in systems biology, AMB Express 1 (1) (2011) 45.

[107] S. Marco-Sola, M. Sammeth, R. Guigo, P. Ribeca, The GEM mapper: fast, accurate and versatile alignment by filtration, Nature Methods 9 (12) (Dec 2012) 1185–1188.

[108] Florian Markowetz, All biology is computational biology, PLoS Biology 15 (3) (2017) e2002050.

[109] T. Marschall, S. Rahmann, Efficient exact motif discovery, Bioinformatics 25 (12) (Jun 2009) i356–i364.

[110] Paul Medvedev, Son Pham, Mark Chaisson, Glenn Tesler, Pavel Pevzner, Paired de Bruijn graphs: a novel approach for incorporating mate pair information into genome assemblers, Journal of Computational Biology 18 (11) (2011) 1625–1634.

[111] J. Merkin, C. Russell, P. Chen, C.B. Burge, Evolutionary dynamics of gene and isoform regulation in Mammalian tissues, Science 338 (6114) (Dec 2012) 1593–1599.

[112] F. Mignone, C. Gissi, S. Liuni, G. Pesole, Untranslated regions of mRNAs, Genome Biology 3 (3) (2002), REVIEWS0004.

[113] R.E. Mills, et al., Mapping copy number variation by population-scale genome sequencing, Nature 470 (7332) (Feb 2011) 59–65.

[114] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, Uri Alon, Network motifs: simple building blocks of complex networks, Science 298 (5594) (2002) 824–827.

[115] Mitchell L. Model, Bioinformatics Programming Using Python: Practical Programming for Biological Data, 1st edition, OReilly Media, Inc., 2009.
[116] Edward F. Moore, The shortest path through a maze, in: Proc. Int. Symp. Switching Theory, 1959, 1959, pp. 285–292.
[117] David W. Mount, Bioinformatics: Sequence and Genome Analysis, 2nd edition, Cold Spring Harbor Laboratory Press, 2004.
[118] Saul B. Needleman, Christian D. Wunsch, A general method applicable to the search for similarities in the amino acid sequence of two proteins, Journal of Molecular Biology 48 (3) (1970) 443–453.
[119] Cédric Notredame, Desmond G. Higgins, Jaap Heringa, T-coffee: a novel method for fast and accurate multiple sequence alignment, Journal of Molecular Biology 302 (1) (2000) 205–217.
[120] C.M. O'Connor, J.U. Adams, Essentials of Cell Biology, NPG Education, Cambridge, MA, USA, 2010.
[121] Smithsonian's National Museum of Natural History and the National Institutes of Health's National Human Genome Research Institute, Unlocking life's code, https://unlockinglifescode.org/, February 2017.
[122] Q. Pan, O. Shai, L.J. Lee, B.J. Frey, B.J. Blencowe, Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing, Nature Genetics 40 (12) (Dec 2008) 1413–1415.
[123] William R. Pearson, David J. Lipman, Improved tools for biological sequence comparison, Proceedings of the National Academy of Sciences 85 (8) (1988) 2444–2448.
[124] Yu Peng, Henry C.M. Leung, Siu-Ming Yiu, Francis Y.L. Chin, IDBA—a practical iterative de Bruijn graph de novo assembler, in: Annual International Conference on Research in Computational Molecular Biology, Springer, 2010, pp. 426–440.
[125] Jonathan Pevsner, Bioinformatics and Functional Genomics, 3rd edition, John Wiley & Sons, 2015.
[126] P.A. Pevzner, S.H. Sze, Combinatorial approaches to finding subtle signals in DNA sequences, Proceedings. International Conference on Intelligent Systems for Molecular Biology 8 (2000) 269–278.
[127] Pavel A. Pevzner, Haixu Tang, Michael S. Waterman, An Eulerian path approach to DNA fragment assembly, Proceedings of the National Academy of Sciences 98 (17) (2001) 9748–9753.
[128] Dusty Phillips, Python 3 Object Oriented Programming, Packt Publishing Ltd, 2010.
[129] R. Phillips, R. Milo, Cell Biology by the Numbers, Garland Science, Cambridge, MA, USA, 2015.
[130] Lawrence R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, in: Proceedings of the IEEE, 1989, pp. 257–286.
[131] A. Ralston, Operons and prokaryotic gene regulation, in: Nature Education, vol. 1, Nature Publishing Group, 2008, p. 216.
[132] Erzsébet Ravasz, Anna Lisa Somera, Dale A. Mongru, Zoltán N. Oltvai, A.-L. Barabási, Hierarchical organization of modularity in metabolic networks, Science 297 (5586) (2002) 1551–1555.
[133] Jennifer L. Reed, Thuy D. Vo, Christophe H. Schilling, Bernhard O. Palsson, An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR), Genome Biology 4 (9) (2003) R54.
[134] Marie-France Sagot, Spelling approximate repeated or common motifs using a suffix tree, in: Proc. of the 3rd Latin American Symposium on Theoretical Informatics, LATIN'98, Campinas, Brazil, 1998, pp. 374–390.
[135] Naruya Saitou, Masatoshi Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Molecular Biology and Evolution 4 (4) (1987) 406–425.
[136] G.K. Sandve, F. Drabløs, A survey of motif discovery methods in an integrated framework, Biology Direct 1 (Apr 2006) 11.
[137] T.D. Schneider, R.M. Stephens, Sequence logos: a new way to display consensus sequences, Nucleic Acids Research 18 (20) (Oct 1990) 6097–6100.
[138] Robert Sedgewick, Kevin Wayne, Algorithms, Addison-Wesley Professional, 2011.
[139] Fabian Sievers, Andreas Wilm, David Dineen, Toby J. Gibson, Kevin Karplus, Weizhong Li, Rodrigo Lopez, Hamish McWilliam, Michael Remmert, Johannes Söding, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Molecular Systems Biology 7 (1) (2011) 539.
[140] H.O. Smith, K.W. Wilcox, A restriction enzyme from Hemophilus influenzae. I. Purification and general properties, Journal of Molecular Biology 51 (2) (Jul 1970) 379–391.

[141] Temple F. Smith, Michael S. Waterman, Identification of common molecular subsequences, Journal of Molecular Biology 147 (1) (1981) 195–197.
[142] R. Sokal, C. Michener, A statistical method for evaluating systematic relationships, University of Kansas Science Bulletin 28 (1958) 1409–1438.
[143] E.L. Sonnhammer, S.R. Eddy, E. Birney, A. Bateman, R. Durbin, Pfam: multiple sequence alignments and HMM-profiles of protein domains, Nucleic Acids Research 26 (1) (Jan 1998) 320–322.
[144] G.D. Stormo, DNA binding sites: representation and discovery, Bioinformatics 16 (1) (Jan 2000) 16–23.
[145] G.D. Stormo, G.W. Hartzell, Identifying protein-binding sites from unaligned DNA fragments, Proceedings of the National Academy of Sciences of the United States of America 86 (4) (Feb 1989) 1183–1187.
[146] Eric Talevich, Brandon M. Invergo, Peter J.A. Cock, Brad A. Chapman, Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython, BMC Bioinformatics 13 (1) (2012) 209.
[147] Julie D. Thompson, Desmond G. Higgins, Toby J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Research 22 (22) (1994) 4673–4680.
[148] M. Tompa, N. Li, T.L. Bailey, G.M. Church, B. De Moor, E. Eskin, A.V. Favorov, M.C. Frith, Y. Fu, W.J. Kent, V.J. Makeev, A.A. Mironov, W.S. Noble, G. Pavesi, G. Pesole, M. Regnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. Weng, C. Workman, C. Ye, Z. Zhu, Assessing computational tools for the discovery of transcription factor binding sites, Nature Biotechnology 23 (1) (Jan 2005) 137–144.
[149] Guido van Rossum, Personal home page, http://legacy.python.org/~guido/.
[150] J. Craig Venter, Mark D. Adams, Eugene W. Myers, Peter W. Li, Richard J. Mural, Granger G. Sutton, Hamilton O. Smith, Mark Yandell, Cheryl A. Evans, Robert A. Holt, et al., The sequence of the human genome, Science 291 (5507) (2001) 1304–1351.
[151] E.T. Wang, R. Sandberg, S. Luo, I. Khrebtukova, L. Zhang, C. Mayr, S.F. Kingsmore, G.P. Schroth, C.B. Burge, Alternative isoform regulation in human tissue transcriptomes, Nature 456 (7221) (Nov 2008) 470–476.
[152] K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S.F. Grant, H. Hakonarson, M. Bucan, PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data, Genome Research 17 (11) (Nov 2007) 1665–1674.
[153] W. Wei, X.D. Yu, Comparative analysis of regulatory motif discovery tools for transcription factor binding sites, Genomics, Proteomics & Bioinformatics 5 (2) (May 2007) 131–142.
[154] Peter Weiner, Linear pattern matching algorithms, in: Switching and Automata Theory, 1973. SWAT'08. IEEE Conference Record of 14th Annual Symposium on, IEEE, 1973, pp. 1–11.
[155] Niklaus Wirth, Algorithms + Data Structures = Programs, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1978.
[156] D.J. Witherspoon, S. Wooding, A.R. Rogers, E.E. Marchani, W.S. Watkins, M.A. Batzer, L.B. Jorde, Genetic similarities within and between human populations, Genetics 176 (1) (May 2007) 351–359.
[157] Chi-En Wu, PythonHMM, https://github.com/jason2506/PythonHMM. (Retrieved October 2017).
[158] Daniel R. Zerbino, Ewan Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Research 18 (5) (2008) 821–829.
[159] L. Zhang, S. Kasif, C.R. Cantor, N.E. Broude, GC/AT-content spikes as genomic punctuation marks, Proceedings of the National Academy of Sciences of the United States of America 101 (48) (Nov 2004) 16855–16860.
[160] W. Zhang, J. Chen, Y. Yang, Y. Tang, J. Shang, B. Shen, A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies, PLoS ONE 6 (3) (Mar 2011) e17915.
[161] Konrad Zuse, Der Plankalkül. Number 63, Gesellschaft für Mathematik und Datenverarbeitung, 1972.