Reading material for Stora Tentan at DBB

Everything marked in yellow is mandatory but a general understanding of all topics mentioned below is needed.

1. Resources from Intro course on the web
    a. Biological Databases
        i. http://en.wikipedia.org/wiki/Biological_databases
        ii. http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.0010034
        iii. http://www.uniprot.org/help/about
        iv. http://en.wikipedia.org/wiki/List_of_biological_databases
        v. EMBD http://www.emdatabank.org
        vi. http://en.wikipedia.org/wiki/Entrez
        vii. http://metadatabase.org/wiki/Main_Page
    b. Genes, sequencing and genomes
        i. http://www.yourgenome.org/facts/what-is-a-genome
        ii. http://en.wikipedia.org/wiki/Bioinformatics
        iii. http://en.wikipedia.org/wiki/Genome
        iv. https://en.wikipedia.org/wiki/Gene_prediction
        v. https://en.wikipedia.org/wiki/Synteny
        vi. http://en.wikipedia.org/wiki/Introduction_to_genetics
        vii. http://en.wikipedia.org/wiki/Human_genome
        viii. http://en.wikipedia.org/wiki/Genome_evolution
        ix. http://en.wikipedia.org/wiki/Sequencing
        x. Are there laws of genome evolution ?
        xi. https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna-

    c. Sequence alignment and searches
        i. https://en.wikipedia.org/wiki/Sequence_homology
        ii. http://en.wikipedia.org/wiki/Sequence_alignment
        iii. http://en.wikipedia.org/wiki/Needleman-Wunsch_algorithm
        iv. http://en.wikipedia.org/wiki/Substitution_matrix
        v. http://en.wikipedia.org/wiki/BLOSUM
        vi. http://en.wikipedia.org/wiki/Point_accepted_mutation
        vii. http://bioinfo.se/courses/introduction-to-bioinformatics/lessons/sequence-alignments/
        viii. http://en.wikipedia.org/wiki/Smith%E2%80%93Waterman_algorithm
        ix. http://www.ncbi.nlm.nih.gov/blast/tutorial/
        x. http://en.wikipedia.org/wiki/BLAST
        xi. http://www.youtube.com/watch?v=HXEpBnUbAMo
        xii. http://en.wikipedia.org/wiki/FASTA
        xiii. http://en.wikipedia.org/wiki/Sequence_alignment_software
        xiv. http://www.ncbi.nlm.nih.gov/books/NBK1734/
        xv. http://openwetware.org/wiki/Wikiomics:BLAST_tutorial
    d. Multiple sequence alignments

i. https://en.wikipedia.org/wiki/Multiple_sequence_alignment
ii. https://en.wikipedia.org/wiki/Hidden_Markov_model

e. Phylogeny and evolution
i. http://evolution.berkeley.edu/evolibrary/article/phylogenetics_01
ii. https://en.wikipedia.org/wiki/Phylogenetic_tree
1. https://en.wikipedia.org/wiki/UPGMA
2. https://en.wikipedia.org/wiki/Neighbor_joining
iii. https://en.wikipedia.org/wiki/Phylogenomics
iv. https://en.wikipedia.org/wiki/Phylogenetic_network
v. https://en.wikipedia.org/wiki/List_of_phylogenetics_software
vi. https://en.wikipedia.org/wiki/Phylogenetic_tree_viewers
vii. https://en.wikipedia.org/wiki/Phylogenetics
viii. https://www.ted.com/talks/svante_paeaebo_dna_clues_to_our_inner_neanderthal?language=en

f. Machine learning
i. https://en.wikipedia.org/wiki/Machine_learning
ii. https://en.wikipedia.org/wiki/Deep_learning
iii. http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.0030116

g. Protein structure and folding
i. https://en.wikipedia.org/wiki/Protein_folding
ii. https://en.wikipedia.org/wiki/Homology_modeling
iii. https://en.wikipedia.org/wiki/Protein_structure_prediction
iv. https://en.wikipedia.org/wiki/Membrane_topology

h. Systems biology
i. https://en.wikipedia.org/wiki/Biological_network
ii. https://en.wikipedia.org/wiki/Interactome
iii. https://en.wikipedia.org/wiki/Systems_biology
iv. https://en.wikipedia.org/wiki/Flux_balance_analysis
v. https://en.wikipedia.org/wiki/Metabolic_network_modelling

2. Alignments and sequence searches
a. Needleman SB, Wunsch CD. **A general method applicable to the search for similarities in the amino acid sequence of two proteins.** J Mol Biol. 1970 Mar;48(3):443-53.PubMed
b. Dayhoff et al, jA model of evolutionary change in proteins. Atlas of Protein Sequence and Structure 5 (3): 345–352 (1978)*PDF*
c. Henikoff & Henikoff: Amino acid substitution matrices from protein blocks. PNAS 89 (22): 10915–9
d. Smith TF, Waterman MS. **Identification of common molecular subsequences.** J Mol Biol. 1981 Mar 25;147(1):195-7.PubMed
e. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. **Basic local alignment search tool.** J Mol Biol. 1990 Oct 5;215(3):403-10.
f. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **Domains Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** Nucleic Acids Res. 1997 Sep 1;25(17):3389-402. Review. PubMed
g. Kent WJ. **BLAT--the BLAST-like alignment tool.** Genome Res. 2002 Apr;12(4):656-64. PMID: PubMed

h. Hidden Markov models in computational biology. Applications to protein modeling. Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. J Mol Biol. 1994 Feb 4;235(5):1501-31. pdf
i. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. Eddy SR. PLoS Comput Biol. 2008 May 30;4(5):e1000069. PDF
j. A tutorial on hidden Markov models and selected applications in speech recognition. Lawrence R Rabiner, PDF
k. Söding J. **Protein homology detection by HMM-HMM comparison.**Bioinformatics. 2005 Apr 1;21(7):951-60. Epub 2004 Nov 5. Erratum in: Bioinformatics. 2005 May 1;21(9):2144. PMID: PubMed
l. **Review of Common Sequence Alignment Methods: Clues to Enhance Reliability** Christophe Lambert*, Jean-Marc Van Campenhout, Xavier DeBolle and Eric Depiereux Current Genomics, 2003, 4, 131-146

3. Fold Recognition
   a. Bowie JU, Luthy R, Eisenberg D. **A method to identify protein sequences that fold into a known three-dimensional structure.** Science. 1991 Jul 12;253(5016):164-70, PubMed
   b. Jones DT, Taylor WR, Thornton JM. **A new approach to protein fold recognition.** Nature. 1992 Jul 2;358(6381):86-9, PubMed
   c. Park J, Karplus K, Barret C, Hughey R, Haussler D, Hubbard T, Chothia C. **Sequence Comparisons Using Multiple Sequences Detect Three Times as Many Remote Homologues as Pairwise Methods.** J. Mol. Biol.(1998) 284, 1201-1210. PubMed
   d. Simons, KT, Ruczinski, I, & Kooperberg, C. (1999). Improved recognition of native like protein structures using a combination of sequence-dependent and sequence independent features of proteins. *Proteins: Structure*.
   e. Fischer D, Eisenberg D. **Protein fold recognition using sequence-derived predictions.**Protein Sci. 1996 May;5(5):947-55.PubMed

4. Protein Domains
   a. Holm L, Sander C. **Mapping the protein universe.** Science. 1996 Aug 2;273(5275):595-603. Review. PubMed
   b. Murzin AG, Brenner SE, Hubbard T, Chothia C. Abstract **SCOP: a structural classification of proteins database for the investigation of sequences and structures.** J Mol Biol. 1995 Apr 7;247(4):536-40. PubMed PDF
   c. Sonnhammer EL, Eddy SR, Durbin R. Abstract **Pfam: a comprehensive database of protein domain families based on seed alignments.** Proteins. 1997 Jul;28(3):405-20.PubMed
   d. Gordana Apic, Julian Gough, Sarah A Teichmann, **Domain combinations in archaeal, eubacterial and eukaryotic proteomes,** Journal of Molecular Biology, Volume 310, Issue 2, 6 July 2001, Pages 311-325, ISSN 0022-2836 Link
   e. Gabrielle A. Reeves, Timothy J. Dallman, Oliver C. Redfern, Adrian Akpor, Christine A. Orengo, **Structural Diversity of Domain Superfamilies in the CATH Database** Journal of Molecular Biology, Volume 360, Issue 3, 14 July 2006, Pages 725-741Link
   f. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. **Assembly reflects evolution of protein complexes.** Nature. 2008 Jun 26;453(7199):1262-5. Epub 2008 Jun 18.DOI

g. **Domain Rearrangements in Protein Evolution** Asa K. Bjorklund, Diana Ekman, Sara Light, Johannes Frey-Skott and Arne Elofsson J. Mol. Biol. (2005) 353, 911–923

5. Secondary Structure Predictions
    a. Chou PY, Fasman GD. **Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins.** Biochemistry. 1974 Jan 15;13(2):211-22. PDF
    b. Chou PY, Fasman GD. **Prediction of protein conformation.** Biochemistry. 1974 Jan 15;13(2):222-45. PubMed PDF
    c. **Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins.** Garnier J, Osguthorpe DJ, Robson B. J Mol Biol. 1978 Mar 25;120(1):97-120. PMID: PubMed
    d. Rost B, Sander C. **Prediction of protein secondary structure at better than 70% accuracy.** J Mol Biol. 1993 Jul 20;232(2):584-99, PubMed

6. Protein structure modeling
    a. Scheres, Sjors HW. A Bayesian View on cryo-EM structure determination, J. Mol. Biol. 415 (2012), 406-18
    b. Brunger, A Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, Nature 355, 472 - 475 (30 January 1992); doi:10.1038/355472a0

7. Protein Structure Prediction
    a. Sali A, Blundell TL. **Comparative protein modelling by satisfaction of spatial restraints.** J Mol Biol. 1993 Dec 5;234(3):779-815.
    b. Bradley P, Misura KM, Baker D. **Toward high-resolution de novo structure prediction for small proteins.** Science. 2005 Sep 16;309(5742):1868-71. PDFSupplement
    c. *E Stat Nonlin Soft Matter Phys* **74**(4 Pt 2), 046110.
    d. Weigt, Martin, White, Robert A, Szurmant, Hendrik, Hoch, James A, & Hwa, Terence. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci* **106**(1), 67–72.
    e. Simons, KT., Kooperberg, C., Huang, E., & Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**(1), 209–225.
    f. Simons, KT, Ruczinski, I, & Kooperberg, C. (1999). Improved recognition of native like protein structures using a combination of sequence-dependent and sequence independent features of proteins. *Proteins: Structure*.
    g. Marks DS, Colwell LJ, Sheridan R, Hopf TA, Pagnani A, Zecchina R, Sander C. **Protein 3D structure computed from evolutionary sequence variation.** PLoS One. 2011;6(12):e28766.
    h. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. Nat Biotechnol. 2012 Nov;30(11):1072-80

8. Motif finding
    a. Lawrence, C. E., Altschul, S. F., Boguski, M. S., Liu, J. S., Neuwald, A. F., & Wootton, J. C. **Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment.** Science, 1993, 262(5131), 208-14.PubMed
    b. Gary D. Stormo, **DNA binding sites: representation and discovery** Bioinformatics 2000 16: 16-23. PDF
9. Subcellular sorting

a. **Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.**Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Protein Eng. 1997 Jan;10(1):1-6. PMID: PubMed PDF
b. **A new method for predicting signal sequence cleavage sites.** von Heijne G. Nucleic Acids Res. 1986 Jun 11;14(11):4683-90. PMID: PubMed PDF

10. Membrane proteins
    a. **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** Krogh A, Larsson B, von Heijne G, Sonnhammer EL. J Mol Biol. 2001 Jan 19;305(3):567-80. PMID: PubMed PDF
    b. **Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule.** von Heijne G. J Mol Biol. 1992 May 20;225(2):487-94. PMID: PubMed PDF
    c. Membrane protein structure: prediction versus reality. **Elofsson** A, **von Heijne** G. Annu Rev Biochem. 2007;76:125-40. Review. PMID: 17579561

11. Molecular Modelling and computational chemistry
    a. Chothia C, Levitt M, Richardson D. **Structure of proteins: Packing of alpha-helices and pleated sheets** Proc Natl Acad Sci U S A. 1977 Oct;74(10):4130-4. PDF
    b. McCammon JA, Gelin BR, Karplus M. **Dynamics of folded proteins.** Nature. 1977 Jun 16;267(5612):585-90. PDF
    c. Levinthal C. **Molecular model-building by computer.** Sci Am. 1966 Jun;214(6):42-52.PDF
    d. Bradley P, Misura KM, Baker D. **Toward high-resolution de novo structure prediction for small proteins.** Science. 2005 Sep 16;309(5742):1868-71. PDF Supplement
    e. Kohlhoff, K. et al, Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways, Nature Chem. 6 (2014), 15-21

12. Molecular Docking
    a. Kuntz, ID; Blaney, JM; OAtley, SJ; et al. **A geometric approach to macromolecule-ligand interactions** Journal of Molecular BIOLOGY Volume: 161 Issue: 2 Pages: 269-288
    b. Meng, EC; SHoichet, BK; Kuntz, ID **Automated docking with grid-based energy evaluation** JOURNAL OF COMPUTATIONAL CHEMISTRY Volume: 13 Issue: 4 Pages: 505- 524 DOI: 10.1002/jcc.540130412 Published: MAY 1992
    c. BOHM, HJ **The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known 3- dimensional structure** Journal of Computer-Aided Molecular Design Volume: 8 Issue: 3 Pages: 243-256 DOI: 10.1007/BF00126743 Published: JUN 1994
    d. Muegge, I; Martin, YC **A general and fast scoring function for protein-ligand interactions: A simplified potential approach** Journal of Medicinal Chemistry Volume: 42 Issue: 5 Pages: 791-804 DOI: 10.1021/jm980536j Published: MAR 11 1999
    e. Warren, Gregory L.; Andrews, C. Webster; Capelli, Anna-Maria; et al. **A critical assessment of docking programs and scoring functions** Source: JOURNAL OF MEDICINAL CHEMISTRY Volume: 49 Issue: 20 Pages: 5912- 5931 DOI: 10.1021/jm050362n Published: OCT 5 2006

13. Protein Design

a. Dahiyat BI, Mayo SL. **De novo protein design: fully automated sequence selection.**Science. 1997 Oct 3;278(5335):82-7. PubMed
b. Dalal S, Balasubramanian S, Regan L. **Protein alchemy: changing beta-sheet into alpha-helix.** Nat Struct Biol. 1997 Jul;4(7):548-52. PubMed PDF
c. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D. **De novo computational design of retro-aldol enzymes.** Science. 2008 Mar 7;319(5868):1387-91.PubMed
d. Koga N, Tatsumi-Koga R, Liu G, Xiao R, Acton TB, Montelione GT, Baker D,**Principles for designing ideal protein structures**, Nature 491, 222-227 (2012)
e. Yin H, Slusky JS, Berger BW, Walters RS, Vilaire G, Litvinov RI, Lear JD, Caputo GA, Bennett JS, Degrado WF, **Computational design of peptides that target transmembrane helices,** Science 315, 1817-1822 (2007)
f. *Procko E, Berguig GY, Shen BW et al. **A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells.** Cell. 2014 Jun 19;157(7):1644-56, PubMed*
g. *Huang PS, Boyken SE, Baker D. **The coming of age of de novo protein design.** Nature. 2016 Sep 15;537(7620):320-7. PubMed*

14. Protein protein interactions
    a. Tsoka S, Ouzounis CA. **Prediction of protein interactions: metabolic enzymes are frequently involved in gene fusion.** Nat Genet. 2000 Oct;26(2):141-2, PubMed
    b. Marcotte EM, Pellegrini M, Thompson MJ, Yeates TO, Eisenberg D. **A combined algorithm for genome-wide prediction of protein function.** Nature. 1999 Nov 4;402(6757):83-6 PubMed
    c. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D. **Detecting protein function and protein-protein interactions from genome sequences.** Science. 1999 Jul 30;285(5428):751-3. PubMed
    d. Doolittle RF. **Do you dig my groove?** Nature Genetics, 1999, 23, 6-8 PDF
    e. Enright AJ, Ouzounis CA. **Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions** Genome Biology, 2001, 2, 0034.1-0034.7 PDF
    f. Snel B, Bork P, Huynen M. **Genome evolution: Gene fusion versus gene fission** TIG, 2000, 16, 9-11 PDF
    g. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. **Comparative assessment of large-scale data sets of protein-protein interactions** Nature, 2002, 417, 399-403. PDF
    h. **What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? D**iana Ekman, Sara Light, Åsa K Björklund and Arne Elofsson Genome Biology 2006, 7:R45 (doi:10.1186/gb-2006-7-6-r45)
    i. Burger, Lukas, & Nimwegen, Erik. (2008). Accurate prediction of protein-protein interactions from sequence alignments using a bayesian method. *Molecular systems biology* **4**(1), 165.
15. Protein Networks
    a. Ideker TE, Thorsson V, Karp RM. **Discovery of regulatory interactions through perturbation: inference and experimental design.** Pac Symp Biocomput 2000;:305-16. PDF, PubMed

b. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** Science 2001 May 4;292(5518):929-34. PDF, PubMed
c. Ideker T, Ozier O, Schwikowski B, Siegel AF. **Discovering regulatory and signalling circuits in molecular interaction networks.** Bioinformatics 2002 Jul;18 Suppl 1:S233-40 PDF, PubMed
d. Yanai I, DeLisi C. **The society of genes: networks of functional links between genes from comparative genomics** Genome Biology, 2002, 3, 0064.1-0064.12 PDF
e. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. **The large-scale organization of metabolic networks.**Nature 2000 Oct 5;407(6804):651-4 PDF, PubMed
f. Barabasi AL, Albert R. **Emergence of scaling in random networks.** Science 1999 Oct 15;286(5439):509-12PDF, PubMed
g. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, Kalyana-Sundaram S, Ghosh D, Pandey A, Chinnaiyan AM. **Probabilistic model of the human protein-protein interaction network.** Nat Biotechnol. 2005 Aug;23(8):951-9. PubMed
h. Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM. **A single gene network accurately predicts phenotypic effects of gene perturbation in Caenorhabditis elegans.**Nat. Genet. 2008, 40:181-8 PubMed
i. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. **The human disease network.** Proc Natl Acad Sci U S A. 2007 104:8685-90 PubMed

16. Sequencing and genomes
    a. Sanger F, Nicklen S, Coulson AR. **DNA sequencing with chain-terminating inhibitors.**Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.
    b. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. **Fluorescence detection in automated DNA sequence analysis.** Nature. 1986 Jun 12-18;321(6071):674-9
    c. Margulies M et al. **Genome sequencing in microfabricated high-density picolitre reactors.** Nature. 2005 Sep 15;437(7057):376-80 PubMed
    d. Ronaghi M, Uhlén M, Nyrén P. **A sequencing method based on real-time pyrophosphate.** Science. 1998 Jul 17;281(5375):363, 365. pubmed

17. Gene assignment
    a. Lukashin AV, Borodovsky M. **GeneMark.hmm: new solutions for gene finding.**Nucleic Acids Res. 1998 Feb 15;26(4):1107-15.PDF

18. Genome assembly
    a. Anson E & Myers G **Algorithms for whole genome shotgun sequencing** Annual Conference on Research in Computational Molecular Biology, 1999 PDF
    b. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC.**A whole-genome assembly of Drosophila.** Science. 2000 Mar 24;287(5461):2196-204.pubmed
    c. ***The Fragment Assembly string graph.****Myers 2005 PDF*
    d. ***Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.****Zerbino, 2009 PDF*

e. ***Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species.*** *Bradnman and others 2013 PDF*

19. Gene mapping and genotyping
    a. Li, H., & Durbin, R. (2009). **Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14),** 1754–1760. pdf
    b. Yang Li, Hong-Mei Li, Paul Burns, Mark Borodovsky, Gene E. Robinson, Jian Ma **TrueSight: Self-training Algorithm for Splice Junction Detection Using RNA-seq**Research in Computational Molecular Biology Lecture Notes in Computer Science Volume 7262, 2012, pp 163-164 link
    c. Trapnell, C., Williams, B., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M., Salzberg, S., et al. (2010). **Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.** Nature Biotechnology, 28(5), 511–515. Link
    d. *Li, H.* ***A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.*** *Bioinformatics. 2011 Nov 1;27(21):2987-93. PubMed PDF*
    e. *DePristo, MA., Banks, E.* ***A framework for variation discovery and genotyping using next-generation DNA sequencing data.*** *Nat Genet. 2011 May;43(5):491-8. PubMed PDF*
    f. *Li, H., Ruan, J., Durbin R.* ***Mapping short DNA sequencing reads and calling variants using mapping quality scores.*** *Genome Res. 2008 Nov;18(11):1851-8. PubMed PDF*
    g. *Marth, GT., Korf, I., Yandell, MD.* ***A general approach to single-nucleotide polymorphism discovery.*** *Nat Genet. 1999 Dec;23(4):452-6. PubMed PDF*

20. Molecular Evolution
    a. Linus Pauling and Emile Zuckerkandl, **Chemical Paleogenetics. Molecular "Restoration Studies" of Extinct Forms of Life.** Acta Chem. Scand. 17 (1963) Suppl 1, pp 9-16.
    b. Langley, C.H. and Fitch, W.M., **An examination of the constancy of the rate of molecular evolution.** J. Mol. Evol. 3 (1974) 161-177.
    c. Kimura, M. **Evolutionary rate at the molecular level.** Nature 217 (1968) 624-626.
    d. Fitch, W. M. **Homology: a personal view on some of the problems.** Trends in Genetics (2000) 16:227-231. (hardcopy available)
    e. Goodman, M., Cselusniak, J., Moore, G. W., Romero-Herrera, A. E., and Matsuda, G. **Fitting the gene lineage into its species lineage: A parsimony strategy illustrated by cladograms constructed from globin sequences.**Syst. Zool. (1979) 28:132-168. (hardcopy available)
    f. Ohno, S., Wolf, U. and Atkin, N.B. **Evolution from fish to mammals by gene duplication.** Hereditas (1968) 59:169-187. (hardcopy available)

21. Phylogeny (need to add papers on NJ and UPGMA)
    a. Steel, M. and Penny, D., **Parsimony, likelihood and the role of models in molecular phylogenetics.** Molecular Biology and Evolution 17(6) 839-850. PDF
    b. Fitch, W.M., **Toward defining the course of evolution: minimum change for a specific tree topology.**Systematic Zoology 20 (1971) 406-416.
    c. Felsenstein, J. **Evolutionary Trees from DNA-Sequences - a Maximum-Likelihood Approach.** J. Mol. Evol. (1981) 17:368-376. PDF

d. Huelsenbeck, John P, F Ronquist, R Nielsen, and J P Bollback. 2001. **Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology** Science (New York, NY) 294 (5550) (December 14): 2310–2314. doi:10.1126/science.1065889.
e. Wong, Karen M, Marc A Suchard, and John P Huelsenbeck. 2008. **Alignment Uncertainty and Genomic Analysis Science** (New York, NY) 319 (5862) (January 25): 473–476. doi:10.1126/science.1151532. PubMed
f. Redelings, Benjamin, and Marc Suchard. 2005. **Joint Bayesian Estimation of Alignment and Phylogeny**. Systematic Biology 54 (3) (June 1): 401–418. doi:10.1080/10635150590947041. PubMed
g. **The Roots of Bioinformatics in Protein Evolution** Russell F. Doolittle PlosCB July 2010 | Volume 6 | Issue 7 | e1000875

22. Comparative Genomics
    a. Tatusov RL, Koonin EV, Lipman DJ. **A genomic perspective on protein families.** Science. 1997 Oct 24;278(5338):631-7. PubMed
    b. Maido Remm, Christian E. V. Storm, and Erik L. L. Sonnhammer **Automatic clustering of orthologs and in-paralogs from pairwise species comparisons** J. Mol. Biol. 314:1041-1052 (2001) PubMed
    c.

23. Expression analysis
    a. Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. **Light-directed, spatially addressable parallel chemical synthesis** Science. 1991 Feb 15;251(4995):767-73. PubMed
    b. Schena M, Shalon D, Davis RW, Brown PO. **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** Science. 1995 Oct 20;270(5235):467-70. PubMed
    c. Eisen MB, Spellman PT, Brown PO, Botstein D. **Cluster analysis and display of genome-wide expression patterns.** Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14863-8. PubMed
    d. Quackenbush J. **Computational analysis of microarray data.** Nat Rev Genet. 2001 Jun;2(6):418-27. PubMed
    e. A gene atlas of the mouse and human protein-encoding transcriptomes. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. Proc Natl Acad Sci U S A. 2004 Apr 20;101(16):6062-7. Epub 2004 Apr 9. PMID: PubMed
    f. *Mortazavi et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth (2008) vol. 5 (7) pp. 621-628 NCBI*
    g. *Oshlack et al. From RNA-seq reads to differential expression results. Genome Biology (2010), 11:220 NCBI*
    h. RNAseq http://www.nature.com/nrg/journal/v10/n1/abs/nrg2484.html
24. Statistics
    a. ***Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.*** *Yoav Benjamini, and Yosef Hochberg. Journal of the Royal Statistical Society Series B Methodological 57(1):289-300 1995 PDF*
    b. ***Statistical significance for genomewide studies.*** *John D. Storey, 9440–9445, doi:10.1073/pnas.1530509100 2003 PDF*
    c. ***Empirical bayes methods and false discovery rates for microarrays.*** *Efron B, Tibshirani R. Genet Epidemiol. Jun;23(1):70-86 2002 PDF*
25. Machine learning

a. **Deep learning** Yann LeCun, Yoshua Bengio & Geoffrey Hinton Nature 521, 436–444 (28 May 2015) doi:10.1038/nature14539

Links to many PDFs of papers:

https://sites.google.com/a/scilifelab.se/elofsson/teaching/classical-papers