

# Assembly

Sequencing platform/data type

Illumina paired-end + jump

Illumina paired-end

Illumina + PacBio

Adapter trimming/  
quality control

FastQC + Trim Galore!

Transcriptome  
Assembly

Trinity

Genome  
Assembly

ALLPATHS-LG

MaSuRCA

SPAdes

DISCOVAR denovo

SOAPdenovo

MaSuRCA

SOAPdenovo

MaSuRCA

Product: Assembled Genome or Transcriptome

# Annotation

Quality control/  
Annotation  
preparation

BUSCO

Kraken

RepeatMasker

Reapr

Gene prediction &  
functional annotation

JAMg

AUGUSTUS

Trinotate

Web hosting/  
annotation refinement

Smithsonian  
Genome Collection

WebApollo

Comparing  
multiple  
genomes and  
transcriptomes

LASTZ

PHYLUCE

GATK

OrthoMCL

Tuxedo suite

SAMtools

OMA

BWA

BEDtools

# Phylogenomics

MAFFT

MARE

ExaBayes

RAxML

\*BEAST

TNT

ExaML

PhyloBayes

References for software pictured above:

Aberer, A. J., Kobert, K., Stamatakis, A. 2014. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Mol Biol Evol* 31 (10): 2553-2556. doi: 10.1093/molbev/msu236

Altenhoff, A., Skunca, N., Glover, N., Train, C. M., Sueki, A., Pilizota, I., Gori, K., Tomiczek, B., Muller, S., Redestig, H., Gonnet, G. H., Dessimoz, C. 2015. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucl Acids Res* 43 (D1): D240-D249 (doi:10.1093/nar/gku1158).

Andrews, S.: FastQC: A Quality Control tool for High Throughput Sequence Data  
<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., Pevzner, P. A. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 19(5): 455–477. doi: 10.1089/cmb.2012.0021.

Drummond, A. J., Suchard, M. A., Xie, D., Rambaut, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7 *Mol Biol Evol* 29: 1969-1973.

Faircloth, B. C. 2014. phyluce: phylogenetic estimation from ultraconserved elements. doi:10.6079/J9PHYL.

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., Glenn, T. C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst Biol* 61: 717–726. doi:10.1093/sysbio/SYS004.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 2011 May 15;29(7):644-52. doi: 10.1038/nbt.1883.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., Jaffe, D. B. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *PNAS* 108 (4): 1513–1518, doi: 10.1073/pnas.1017351108

Goloboff, P. A., Farris, J. S., Nixon, K. C. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24(5): 774–786. DOI: 10.1111/j.1096-0031.2008.00217.x

Harris, R.S. 2007. Improved pairwise alignment of genomic DNA. Ph.D. Thesis, The Pennsylvania State University.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., Otto, T. D. 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biol* 14:R47, doi:10.1186/gb-2013-14-5-r47

- Katoh, K., Misawa, K., Kuma, K., Miyata, T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059-3066
- Keller, O., Kollmar, M., Stanke, M., Waack, S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, doi: 10.1093/bioinformatics/btr010
- Kozlov, A. M., Aberer, A. J., Stamatakis, A. 2015. ExaML Version 3: A Tool for Phylogenomic Analyses on Supercomputers. *Bioinformatics* 31 (15): 2577-2579. doi: 10.1093/bioinformatics/btv184
- Krueger, F.: Trim Galore! [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
- Lartillot, N., Lepage, T., Blanquart, S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*. 25(17): 2286–2288 doi:10.1093/bioinformatics/btp368
- Lee, E., Helt, G. A., Reese, J. T., Munoz-Torres, M. C., Childers, C. P., Buels, R. M., Stein, L., Holmes, I. H., Elisk, C. G., Lewis, S. E. 2013. Web Apollo: a web-based genomic annotation editing platform. *Genome Biol* 14:R93.
- Li, H., Durbin, R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics* 26(5):589-95. doi: 10.1093/bioinformatics/btp69
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25, 2078-9.
- Li, L., Stoeckert, Jr., C. J., Roos, D. S. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* 13: 2178-2189.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T-W., Wang, J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 1297-303.
- Misof, B., Meyer, B., von Reumont, B. M., Kück, P., Misof, K., Meusemann, K. 2013. Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. *BMC Bioinformatics* 14:348 doi 10.1186/1471-2105-14-348.
- Papanicolaou, A., Haas, B. JAMg: <http://jamg.sourceforge.net>
- Quinlan, A. R., Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6): 841–842. doi: 10.1093/bioinformatics/btq033

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* doi: 10.1093/bioinformatics/btv351.

Smit, A. F. A., Hubley, R., Green, P. RepeatMasker Open-4.0. 2013-2015. <<http://www.repeatmasker.org>>.

Stamatakis, A. 2014. Raxml version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Trapnell, C, Pachter, L., Salzberg, S. L. 2009 TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-11. doi: 10.1093/bioinformatics/btp120

Weisenfeld, N. I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., Sogoloff, B., Tabbaa, D., Williams, L., Russ, C., Nusbaum, C., Lander, E. S., MacCallum, I., Jaffe, D. B. 2014 Comprehensive variation discovery in single human genomes. *Nature Genetics* 46, 1350–1355 doi:10.1038/ng.3121.

Wood, D. E., Salzberg, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., Yorke, J. A. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29 (21): 2669-2677. doi: 10.1093/bioinformatics/btt476.