REFERENCES

[1] R. Li *et al.*, "De novo assembly of human genomes with massively parallel short read sequencing," *Genome Research*, vol. 20, no. 2, pp. 265–272, Feb. 2010, doi: 10.1101/gr.097261.109.

[2] J. Meng, B. Wang, Y. Wei, S. Feng, and P. Balaji, "SWAP-Assembler: scalable and efficient genome assembly towards thousands of cores," *BMC Bioinformatics*, vol. 15, no. Suppl 9, p. S2, 2014, doi: 10.1186/1471-2105-15-S9-S2.

[3] D. R. Kelley, M. C. Schatz, and S. L. Salzberg, "Quake: quality-aware detection and correction of sequencing errors," *Genome Biology*, vol. 11, no. 11, p. R116, 2010, doi: 10.1186/gb-2010-11-11-r116.

[4] Y. Liu, J. Schröder, and B. Schmidt, "Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data," *Bioinformatics*, vol. 29, no. 3, pp. 308–315, Feb. 2013, doi: 10.1093/bioinformatics/bts690.

[5] S. Kurtz, A. Narechania, J. C. Stein, and D. Ware, "A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes," *BMC Genomics*, vol. 9, no. 1, p. 517, 2008, doi: 10.1186/1471-2164-9-517.

[6] A. L. Price, N. C. Jones, and P. A. Pevzner, "De novo identification of repeat families in large genomes," *Bioinformatics*, vol. 21 Suppl 1, pp. i351-358, Jun. 2005, doi: 10.1093/bioinformatics/bti1018.

[7] R. C. Edgar, "MUSCLE: multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res*, vol. 32, no. 5, pp. 1792–1797, 2004, doi: 10.1093/nar/gkh340.

[8] M. Hozza, T. Vinař, and B. Brejová, "How Big is that Genome? Estimating Genome Size and Coverage from k-mer Abundance Spectra," in *String Processing and Information Retrieval*, vol. 9309, C. Iliopoulos, S. Puglisi, and E. Yilmaz, Eds. Cham: Springer International Publishing, 2015, pp. 199–209. doi: 10.1007/978-3-319-23826-5_20.

[9] B. Liu *et al.*, "Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects," *arXiv:1308.2012 [q-bio]*, Feb. 2020, Accessed: Feb. 01, 2021. [Online]. Available: http://arxiv.org/abs/1308.2012

[10] F. S. Collins, "The Human Genome Project: Lessons from Large-Scale Biology," *Science*, vol. 300, no. 5617, pp. 286–290, Apr. 2003, doi: 10.1126/science.1084564.

[11] E. L. van Dijk, H. Auger, Y. Jaszczyszyn, and C. Thermes, "Ten years of next-generation sequencing technology," *Trends in Genetics*, vol. 30, no. 9, pp. 418–426, Sep. 2014, doi: 10.1016/j.tig.2014.07.001.

[12] G. Marçais and C. Kingsford, "A fast, lock-free approach for efficient parallel counting of occurrences of k-mers," *Bioinformatics*, vol. 27, no. 6, pp. 764–770, Mar. 2011, doi: 10.1093/bioinformatics/btr011.

[13] M. Erbert, S. Rechner, and M. Müller-Hannemann, "Gerbil: a fast and memory-efficient k-mer counter with GPU-support," *Algorithms Mol Biol*, vol. 12, no. 1, p. 9, Dec. 2017, doi: 10.1186/s13015-017-0097-9.

[14] G. Rizk, D. Lavenier, and R. Chikhi, "DSK: k-mer counting with very low memory usage," *Bioinformatics*, vol. 29, no. 5, pp. 652–653, Mar. 2013, doi: 10.1093/bioinformatics/btt020.

[15] M. Kokot, M. Długosz, and S. Deorowicz, "KMC 3: counting and manipulating k-mer statistics," *Bioinformatics*, vol. 33, no. 17, pp. 2759–2761, Sep. 2017, doi: 10.1093/bioinformatics/btx304.

[16] J. Wang, S. Chen, L. Dong, and G. Wang, "CHTKC: a robust and efficient k-mer counting algorithm based on a lock-free chaining hash table," *Briefings in Bioinformatics*, p. bbaa063, May 2020, doi: 10.1093/bib/bbaa063.

[17] M. Roberts, W. Hayes, B. R. Hunt, S. M. Mount, and J. A. Yorke, "Reducing storage requirements for biological sequence comparison," *Bioinformatics*, vol. 20, no. 18, pp. 3363–3369, Dec. 2004, doi: 10.1093/bioinformatics/bth408.

[18] Y. Li and XifengYan, "MSPKmerCounter: A Fast and Memory Efficient Approach for K-mer Counting," *arXiv:1505.06550 [cs, q-bio]*, May 2015, Accessed: Feb. 06, 2021. [Online]. Available: http://arxiv.org/abs/1505.06550

[19] S. Deorowicz, M. Kokot, S. Grabowski, and A. Debudaj-Grabysz, "KMC 2: fast and resource-frugal k-mer counting," *Bioinformatics*, vol. 31, no. 10, pp. 1569–1576, May 2015, doi: 10.1093/bioinformatics/btv022.

[20] H. Mohamadi, H. Khan, and I. Birol, "ntCard: a streaming algorithm for cardinality estimation in genomics data," *Bioinformatics*, p. btw832, Jan. 2017, doi: 10.1093/bioinformatics/btw832.

[21] S. Behera, S. Gayen, J. S. Deogun, and N. V. Vinodchandran, "KmerEstimate: A Streaming Algorithm for Estimating k-mer Counts with Optimal Space Usage," in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, Washington DC USA, Aug. 2018, pp. 438–447. doi: 10.1145/3233547.3233587.

[22] P. Melsted and B. V. Halldórsson, "KmerStream: streaming algorithms for k -mer abundance estimation," *Bioinformatics*, vol. 30, no. 24, pp. 3541–3547, Dec. 2014, doi: 10.1093/bioinformatics/btu713.

[23] R. Chikhi and P. Medvedev, "Informed and automated k-mer size selection for genome assembly," *Bioinformatics*, vol. 30, no. 1, pp. 31–37, Jan. 2014, doi: 10.1093/bioinformatics/btt310.

[24] P. Pandey, M. A. Bender, R. Johnson, and R. Patro, "Squeakr: an exact and approximate k-mer counting system," *Bioinformatics*, vol. 34, no. 4, pp. 568–575, Feb. 2018, doi: 10.1093/bioinformatics/btx636.

[25] U. Ferraro Petrillo, G. Roscigno, G. Cattaneo, and R. Giancarlo, "FASTdoop: A Versatile and Efficient Library for the Input of FASTA and FASTQ Files for MapReduce Hadoop Bioinformatics Applications," *Bioinformatics*, p. btx010, Jan. 2017, doi: 10.1093/bioinformatics/btx010.

[26] G. Cattaneo, U. F. Petrillo, R. Giancarlo, and G. Roscigno, "An effective extension of the applicability of alignment-free biological sequence comparison algorithms with Hadoop," *J Supercomput*, vol. 73, no. 4, pp. 1467–1483, Apr. 2017, doi: 10.1007/s11227-016-1835-3.

[27] W. Zhou *et al.*, "MetaSpark: a spark-based distributed processing tool to recruit metagenomic reads to reference genomes," *Bioinformatics*, p. btw750, Jan. 2017, doi: 10.1093/bioinformatics/btw750.

[28] T. Gao *et al.*, "Bloomfish: A Highly Scalable Distributed K-mer Counting Framework," in *2017 IEEE 23rd International Conference on Parallel and Distributed Systems (ICPADS)*, Shenzhen, Dec. 2017, pp. 170–179. doi: 10.1109/ICPADS.2017.00033.

[29] T. Pan, P. Flick, C. Jain, Y. Liu, and S. Aluru, "Kmerind: A Flexible Parallel Library for K-mer Indexing of Biological Sequences on Distributed Memory Systems," *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 16, no. 4, pp. 1117–1131, Jul. 2019, doi: 10.1109/TCBB.2017.2760829.

[30] U. Ferraro Petrillo, M. Sorella, G. Cattaneo, R. Giancarlo, and S. E. Rombo, "Analyzing big datasets of genomic sequences: fast and scalable collection of k-mer statistics," *BMC Bioinformatics*, vol. 20, no. S4, p. 138, Apr. 2019, doi: 10.1186/s12859-019-2694-8.

[31] N. Siva, "1000 Genomes project," *Nat Biotechnol*, vol. 26, no. 3, pp. 256–256, Mar. 2008, doi: 10.1038/nbt0308-256b.

[32] H. Li, A. Ramachandran, and D. Chen, "GPU Acceleration of Advanced k-mer Counting for Computational Genomics," in *2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP)*, Milan, Jul. 2018, pp. 1–4. doi: 10.1109/ASAP.2018.8445084.

[33] N. Mcvicar, C.-C. Lin, and S. Hauck, "K-Mer Counting Using Bloom Filters with an FPGA-Attached HMC," in *2017 IEEE 25th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM)*, Napa, CA, USA, Apr. 2017, pp. 203–210. doi: 10.1109/FCCM.2017.23.

[34] J. Meena, S. Sze, U. Chand, and T.-Y. Tseng, "Overview of emerging nonvolatile memory technologies," *Nanoscale Res Lett*, vol. 9, no. 1, p. 526, 2014, doi: 10.1186/1556-276X-9-526.

[35] N. Cadenelli, J. Polo, and D. Carrera, "Accelerating K-mer Frequency Counting with GPU and Non-Volatile Memory," in *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*, Bangkok, Dec. 2017, pp. 434–441. doi: 10.1109/HPCC-SmartCity-DSS.2017.57.

[36] V. Gramoli, "More than you ever wanted to know about synchronization: synchrobench, measuring the impact of the synchronization on concurrent algorithms," in *Proceedings of the 20th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, San Francisco CA USA, Jan. 2015, pp. 1–10. doi: 10.1145/2688500.2688501.

[37] P. Melsted and J. K. Pritchard, "Efficient counting of k-mers in DNA sequences using a bloom filter," *BMC Bioinformatics*, vol. 12, no. 1, p. 333, Dec. 2011, doi: 10.1186/1471-2105-12-333.