

***DNA BARCODING DENGAN ALGORITMA PARTICLE SWARM
OPTIMIZATION MENGGUNAKAN APACHE SPARK SQL***

SKRIPSI

Diajukan untuk Memenuhi Sebagian dari
Syarat Memperoleh Gelar Sarjana Komputer
Program Studi Ilmu Komputer



Oleh
Muhammad Ilham Nurfathiya
NIM 1604564

PROGRAM STUDI ILMU KOMPUTER
DEPARTEMEN PENDIDIKAN ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2019

***DNA BARCODING DENGAN ALGORITMA PARTICLE SWARM
OPTIMIZATION MENGGUNAKAN APACHE SPARK SQL***

Oleh
Muhammad Ilham Nurfathiya
NIM 1604564

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer di Fakultas Pendidikan Matematika dan Ilmu
Pengetahuan Alam

© Muhammad Ilham Nurfathiya 2020
Universitas Pendidikan Indonesia
Mei 2020

Hak Cipta Dilindungi Undang-Undang
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

MUHAMMAD ILHAM NURFATHIYA

1604564

***DNA BARCODING DENGAN ALGORITMA PARTICLE SWARM
OPTIMIZATION MENGGUNAKAN APACHE SPARK SQL***

DISETUJUI DAN DISAHKAN OLEH PEMBIMBING:

Pembimbing I,



Lala Septem Riza, M.T., Ph.D.

NIP. 197809262008121001

Pembimbing II,



Jajang Kusnendar, M.T.

NIP. 197506012008121001

Mengetahui,

Kepala Departemen Pendidikan Ilmu Komputer



Lala Septem Riza, M.T., Ph.D.

NIP. 197809262008121001

PERNYATAAN

Dengan ini penulis menyatakan bahwa skripsi dengan judul “*DNA Barcoding dengan Algoritma Particle Swarm Optimization Menggunakan Apache Spark SQL*” ini beserta seluruh isinya adalah benar-benar karya penulis sendiri. Penulis tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, penulis siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya penulis ini.

Bandung, Mei 2020

Yang Membuat Pernyataan,



Muhammad Ilham Nurfathiya

NIM 1604564

DNA BARCODING DENGAN ALGORITMA PARTICLE SWARM OPTIMIZATION MENGGUNAKAN APACHE SPARK SQL

Oleh

Muhammad Ilham Nurfathiya – ilhamn33@gmail.com

1604564

ABSTRAK

Terdapat salah satu tahap dalam DNA *barcoding* yang masih menggunakan metode manual seperti *similarity check* yang mengakibatkan tahap ini ketelitian dan waktu yang cukup lama. Data sekuens DNA makhluk hidup merupakan data yang sangat banyak pada bidang biologi. Untuk itu penelitian ini membuat sebuah model komputasi untuk mendapatkan DNA *barcode* secara cepat dan efektif dengan mengimplementasikan algoritma *particle swarm optimization* pada *big data platform* yaitu Apache Hadoop dan Apache Spark. Data yang digunakan pada penelitian kali ini adalah data RNA SARS-CoV-2. Hasil dari program yang dibangun berupa DNA *barcode* yang ditemukan dari sampel yang ada beserta waktu yang dibutuhkan untuk menyelesaikan kalkulasi. Dilakukan 2 skenario pengujian, skenario pertama yaitu dengan menggunakan 4 *cores* dan beberapa *worker nodes* dan yang kedua yaitu penggunaan *cluster* dengan 2 *worker nodes* dan beberapa *cores*. Hasil dari penelitian ini membuktikan bahwa model komputasi yang dibangun pada *big data platform* menunjukkan adanya perkembangan fitur dan percepatan terhadap penelitian terdahulu.

Kata Kunci: *Big Data*, algoritma *Particle Swarm Optimization*, *similarity check*, *motif discovery*, DNA *barcoding*

**DNA BARCODING WITH PARTICLE SWARM OPTIMIZATION
ALGORITHM USING APACHE SPARK SQL**

Arranged by

Muhammad Ilham Nurfathiya – ilhamn33@gmail.com

1604564

ABSTRACT

There is one stage in DNA barcoding that still uses manual methods such as similarity check which results in this stage of accuracy and quite a long time. DNA sequence data of living things is very much data in the field of biology. For this reason, this research creates a computational model to obtain DNA barcodes quickly and effectively by implementing the particle swarm optimization algorithm on the big data platform, Apache Hadoop, and Apache Spark. The data used in this study is SARS-CoV-2 RNA data. The results of the program that were built consisted of DNA barcodes found from the existing sample of time needed to complete calculations. The results of this study indicate that there is a significant acceleration between standalone and big data platform with 2 experimental scenarios. The first scenario is to use 4 cores and some worker nodes and the second is to use a cluster with 2 worker nodes and several cores. This research proves that the computational model built on the big data platform shows the development of features and acceleration of previous research.

Keywords: Big Data, Particle Swarm Optimization algorithm, similarity check, discovery motive, DNA barcoding

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah swt. karena hanya dengan kehendak, berkat, serta karunia-Nya lah penulis dapat menyelesaikan skripsi yang berjudul “*DNA Barcoding dengan Algoritma Particle Swarm Optimization menggunakan Apache Spark SQL*” ini dapat terselesaikan.

Penyusunan skripsi ini ditunjukkan untuk memenuhi dan melengkapi salah satu syarat untuk penyusunan skripsi yang merupakan syarat untuk mendapatkan gelar sarjana komputer atas jenjang studi S1 pada Program Studi Ilmu Komputer Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam Universitas Pendidikan Indonesia.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan dan keterbatasan yang perlu disempurnakan. Oleh karena itu, penulis sangat mengharapkan saran maupun kritik yang membangun agar tidak terjadi kesalahan yang sama dikemudian hari dan dapat meningkatkan kualitas ke tahap lebih baik.

Bandung, Mei 2020

Penulis

UCAPAN TERIMAKASIH

Alhamdulillahirabilalamin, puji dan syukur kehadiran Allah Subhanahu Wa Ta'ala Yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis diberikan kelancaran dalam menyelesaikan penulisan skripsi ini. Dalam proses menyelesaikan penelitian dan penyusunan skripsi ini, peneliti banyak mendapat bimbingan, dorongan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini peneliti mengucapkan terimakasih serta penghargaan yang setinggi-tingginya, kepada:

1. Abi dan Umi yaitu Johni Purwantoro dan Heni Setiyowati yang selalu memberikan doa dan dukungan baik moral dan materil, serta selalu menjadi penyemangat utama dalam menempuh pendidikan tinggi sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Lala Septem Riza, M.T., Ph.D. selaku pembimbing I dan Kepala Departemen Pendidikan Imlu Komputer FPMIPA Universitas Pendidikan Indonesia atas segala waktu yang dicurahkan untuk membimbing penulis demi terselesaikannya skripsi ini.
3. Bapak Jajang Kusnendar, M.T., selaku pembimbing II atas segala waktu yang dicurahkan untuk membimbing penulis demi terselesaikannya skripsi ini.
4. Ibu Rani Megasari, M.T., selaku Ketua Program Studi Ilmu Komputer.
5. Bapak Topik Hidayat, Ph.D., selaku Dosen Departemen Pendidikan Biologi yang telah membantu penulis salah satunya dalam memahami konsep dasar dari penelitian ini.
6. Bapak Herbert Siregar, M.T., selaku Dosen Pembimbing Akademik yang telah memberikan arahan dan bimbingan selama penulis menjalani perkuliahan.
7. Ibu Rosa Ariani Sukamto, M.T., selaku Dosen Pemrograman Dasar yang telah memberikan motivasi kepada penulis untuk selalu belajar pada saat awal kuliah.
8. Bapak dan Ibu Dosen Departemen Pendidikan Ilmu Komputer yang tidak bisa penulis sebutkan Namanya satu persatu yang telah berbagi ilmu yang sangat bermanfaat kepada penulis.

9. Sahabat surabi, Ali, Degitgit, Rafif, Iqdam, Tia, dan Winda yang senantiasa memberikan dukungan, semangat, canda, dan tawa kepada penulis baik selama proses perkuliahan maupun selama proses pengerjaan skripsi ini.
10. Pengurus Inti BEM KEMAKOM 2018-2019 yang telah berjuang bersama sama dan “Berdedikasi dengan Hati” untuk menyebarkan kebaikan.
11. DPMB BEM KEMAKOM 2017/2018 yang telah memberikan pengalaman berorganisasi pertama kali dan mengenalkan dunia baru di perkuliahan.
12. SUPER KOMINFO BEM REMA UPI 2020 yang telah memberikan pengalaman baru dalam berorganisasi ditingkat UPI.
13. Kelas C 2016, yang sama-sama berjuang dari awal perkuliahan hingga ke titik akhir perkuliahan ini.
14. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Semoga semua amal baik yang telah diberikan kepada penulis mendapatkan balasan yang berlipat ganda dari Allah Subhanahu Wa Ta’ala dan dapat menjadi salah satu kebaikan yang dapat membawa dan mempertemukan kita kembali ke surganya Allah Subhanahu Wa Ta’ala. Aamiin.

Bandung, Mei 2020

Penulis

DAFTAR ISI

ABSTRAK	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
UCAPAN TERIMAKASIH	iv
DAFTAR ISI	vi
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	14
1.1 Latar Belakang	14
1.2 Rumusan Masalah	17
1.3 Tujuan Penelitian	18
1.4 Manfaat Penelitian	18
1.5 Batasan Masalah	18
1.6 Sistematika	19
BAB II KAJIAN PUSTAKA	21
2.1 Bioinformatika	21
2.1.1 Tujuan Bioinformatika	23
2.1.2 Ruang Lingkup Bioinformatika	24
2.1.3 Batasan Bioinformatika	26
2.1.4 Bidang Penelitian dalam Bioinformatika	27
2.2 DNA dan DNA <i>Barcode</i>	28
2.2.1 Definisi DNA	28
2.2.2 Struktur dan Fungsi DNA	30
2.2.3 Definisi DNA <i>Barcoding</i>	31
2.2.4 Cara Kerja DNA <i>Barcoding</i>	32

2.3	Motif Discovery.....	34
2.4	Algoritma Particle Swarm Optimization.....	38
2.5	Big Data	42
2.5.1	Definisi Big Data.....	42
2.5.2	Karakteristik <i>Big Data</i>	45
2.5.3	Apache Hadoop	48
2.5.4	Apache Spark	54
2.5.5	Spark SQL.....	62
2.6	Google Cloud Platform	66
2.7	Penelitian Terkait.....	71
BAB III METODE PENELITIAN		74
3.1	Desain Penelitian	74
3.2	Metode Penelitian	76
3.2.1	Metode Pengumpulan Data	76
3.2.2	Metode Pengembangan Perangkat Lunak.....	77
3.3	Alat dan Bahan Penelitian.....	78
3.3.1	Alat Penelitian	78
3.3.2	Bahan Penelitian.....	79
BAB IV HASIL DAN PEMBAHASAN		80
4.1	Pengumpulan Data.....	80
4.1.1	Mengunduh Data dari GISAID	81
4.1.2	Penjelasan Isi File.....	84
4.2	Perancangan Model.....	84
4.2.1	Standalone Python	85
4.2.2	Big Data Analisis.....	97
4.3	Pengembangan Perangkat Lunak Standalone	100

4.3.1	Analisa	100
4.3.2	Desain	101
4.3.3	Implementasi	102
4.3.4	Pengujian.....	109
4.4	Pengembangan Perangkat Lunak <i>Big Data Platform</i>	110
4.4.1	Analisa	110
4.4.2	Desain	111
4.4.3	Implementasi	112
4.4.4	Pengujian.....	119
4.5	Rancangan Skenario Eksperimen	119
4.5.1	Skenario data dummy pada <i>standalone</i>	121
4.5.2	Skenario data angrek pada <i>standalone</i>	121
4.5.3	Skenario 4 <i>core</i> dan Beberapa <i>Worker Nodes</i>	122
4.5.4	Skenario Beberapa <i>Core</i> dan 2 <i>Worker Nodes</i>	122
4.6	Langkah Menjalankan Eksperimen	123
4.7	Hasil Eksperimen.....	124
4.7.1	Hasil Eksperimen Data <i>Dummy</i> pada <i>Standalone</i>	124
4.7.2	Hasil Eksperimen Data Anggrek pada <i>Standalone</i>	125
4.7.3	Hasil Eksperimen Data RNA SARS-CoV-2 pada <i>Big Data Platform</i> 128	
4.8	Analisa	135
4.8.1	Pembahasan Eksperimen Data <i>Dummy</i> pada <i>Standalone</i>	136
4.8.2	Pembahasan Eksperimen Data Anggrek pada <i>Standalone</i>	137
4.8.3	Perbandingan Kecepatan Eksperimen dengan 4 <i>core</i> dan beberapa <i>worker nodes</i>	138
4.8.4	Perbandingan Kecepatan Eksperimen dengan 2 <i>worker nodes</i> dan beberapa <i>core</i>	141

4.8.5	Analisis Biologi terhadap Hasil Eksperimen	145
4.8.6	Perbandingan dengan Penelitian Terkait	146
BAB V KESIMPULAN DAN SARAN.....		149
4.9	Kesimpulan.....	149
4.10	Saran	150
DAFTAR PUSTAKA		151
LAMPIRAN		161

DAFTAR PUSTAKA

- Abdurakhmonov, I. Y. (2016). *Bioinformatics : Basics , Development , and Future*. (August). <https://doi.org/10.5772/63817>
- Agrahari, A., Rao, P. D. T. V. D., Sci, C., College, A., & Pradesh, A. (2017). A Review paper on Big Data : Technologies , Tools and Trends. *International Research Journal of Engineering and Technology (IRJET)*, 4(10), 640–649.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (4th editio). New York: Garland Science.
- Aprijani, D. A., & Elfaizi, M. A. (2004). *Bioinformatika : Perkembangan , Disiplin Ilmu dan Penerapannya di Indonesia*.
- Armbrust, M., Xin, R. S., Lian, C., Huai, Y., Liu, D., Bradley, J. K., ... Zaharia, M. (2015). Spark SQL: Relational data processing in spark. *Proceedings of the ACM SIGMOD International Conference on Management of Data, 2015-May*, 1383–1394. <https://doi.org/10.1145/2723372.2742797>
- Arora, S., & Singh, S. (2019). Butterfly optimization algorithm: a novel approach for global optimization. *Soft Computing*, 23(3), 715–734. <https://doi.org/10.1007/s00500-018-3102-4>
- Attwood, T. K., Parry-Smith, D. J., & Phukan, S. (2001). *Introduction to Bioinformatics*. (English). Benjamin-Cummings Publishing Company.
- Bandyopadhyaya, S., Vankayalapati, R., Rajanna, L., & Kulkarni, S. (2013). DNA barcoding and its applications - A critical review. *Cmrim J. Res. and Dev*, 1(1), 77–81. Retrieved from <https://www.researchgate.net/publication/285322603>
- Bayat, A. (2016). *Bioinformatics : Science , medicine , and the future Clinical review*. (May 2002). <https://doi.org/10.1136/bmj.324.7344.1018>

- Beakta, R. (2015). *Big Data And Hadoop: A Review Paper*. 2(2), 13–15.
- Bhattacharya, A., & Bhatnagar, S. (2016). Big Data and Apache Spark: A Review. *International Journal of Engineering Research & Science (IJOER) ISSN*, 2(5), 206–210. Retrieved from <https://ijoer.com/Paper-May-2016/IJOER-MAR-2016-9.pdf>
- Bhosale, H. S., & Gadekar, D. P. (2014). A Review Paper on Big Data and Hadoop. *International Journal of Scientific and Research Publications*, 4(1), 2250–3153. Retrieved from www.ijsrp.org
- Butcher, B. A., Smith, M. A., Sharkey, M. J., & Quicke, D. L. J. (2012). A turbo-taxonomic study of Thai Aleiodes (Aleiodes) and Aleiodes (Arcaleiodes) (Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa*, 232, 1–6.
- Buyya, R., Calheiros, R. N., & Dastjerdi, A. V. (2016). Big Data: Principles and Paradigms. In *Big Data: Principles and Paradigms*. <https://doi.org/10.1016/C2015-0-04136-3>
- Chen, J., Hsu, W., Lee, M. L., & Ng, S.-K. (2006). *NeMoFinder*. (November 2016), 106. <https://doi.org/10.1145/1150402.1150418>
- Ciriello, G., & Guerra, C. (2008). A review on models and algorithms for motif discovery in protein-protein interaction networks. *Briefings in Functional Genomics and Proteomics*, 7(2), 147–156. <https://doi.org/10.1093/bfgp/eln015>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. *OSDI 2004 - 6th Symposium on Operating Systems Design and Implementation*, 137–149. <https://doi.org/10.21276/ijre.2018.5.5.4>
- Desalle, R., Egan, M. G., & Siddall, M. (2005). *The unholy trinity : taxonomy , species delimitation and DNA barcoding*. (September), 1905–1916.

<https://doi.org/10.1098/rstb.2005.1722>

- Dessureault, S. (2016). Understanding big data. *CIM Magazine*, 11(1).
- Dhyani, B., & Barthwal, A. (2014). Big Data Analytics using Hadoop. *International Journal of Computer Applications*, 108(12), 1–5. <https://doi.org/10.5120/18960-0288>
- Dittrich, J., & Quiané-Ruiz, J. A. (2012). Efficient big data processing in Hadoop MapReduce. *Proceedings of the VLDB Endowment*, 5(12), 2014–2015. <https://doi.org/10.14778/2367502.2367562>
- Eberhart, R. C., & Shi, Y. (2001). Particle Swarm Optimization: Developments, Applications and Resources. *IEEE*, 81. [https://doi.org/10.1016/s0026-0576\(02\)80337-3](https://doi.org/10.1016/s0026-0576(02)80337-3)
- Faatih, M. (2009). Isolasi dan digesti DNA kromosom. *J Penelitian Sains Dan Teknologi*, 20(1), 61–67.
- García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2017). A comparison on scalability for batch big data processing on Apache Spark and Apache Flink. *Big Data Analytics*, 2(1), 1–11. <https://doi.org/10.1186/s41044-016-0020-2>
- Ge, H., Yu, J., Sun, L., Wang, Z., & Yao, Y. (2019). Discovery of DNA Motif Utilising an Integrated Strategy Based on Random Projection and Particle Swarm Optimization. *Mathematical Problems in Engineering*, 2019. <https://doi.org/10.1155/2019/3854646>
- Ghulam, K., Husain, M., Salman, S., Tufail, M., Sukirno, S., & Aldawood, A. S. (2019). Saudi Journal of Biological Sciences Original article DNA barcoding of the fire ant genus *Solenopsis* Westwood (Hymenoptera : Formicidae) from the Riyadh region , the Kingdom of Saudi Arabia. *Saudi Journal of Biological Sciences*, (xxxx), 0–4. <https://doi.org/10.1016/j.sjbs.2019.06.018>

- Hardin, C. T., & Rouchka, E. C. (2005). DNA motif detection using particle swarm optimization and expectation-maximization. *Proceedings - 2005 IEEE Swarm Intelligence Symposium, SIS 2005, 2005*, 187–190. <https://doi.org/10.1109/SIS.2005.1501620>
- Hashim, F. A., Mabrouk, M. S., & Al-Atabany, W. (2019). Review of Different Sequence Motif Finding Algorithms. *Avicenna Journal of Medical Biotechnology*, *11*(2), 130–148. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/31057715><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6490410>
- Hidayat, T. (2017). *DNA Barcoding in Plants*. (April), 12–15.
- Hubert, N., Delrieu-Trottin, E., Irisson, J. O., Meyer, C., & Planes, S. (2010). Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae. *Molecular Phylogenetics and Evolution*, *55*(3), 1195–1203. <https://doi.org/10.1016/j.ympev.2010.02.023>
- Hubert, N., & Hanner, R. (2016). DNA Barcoding, species delineation and taxonomy: a historical perspective. *DNA Barcodes*, *3*(1), 44–58. <https://doi.org/10.1515/dna-2015-0006>
- Ilman, M. N. (2019). *Algoritma Optimasi Spiral Dynamics Menggunakan Bahasa Pemrograman R untuk DNA Barcoding*. Universitas Pendidikan Indonesia.
- Imran, M., Hashim, R., & Khalid, N. E. A. (2013). An overview of particle swarm optimization variants. *Procedia Engineering*, *53*(1), 491–496. <https://doi.org/10.1016/j.proeng.2013.02.063>
- Imtiaz, A., Mohd Nor, S. A., & Md. Naim, D. (2017). Progress and potential of DNA barcoding for species identification of fish species. *Biodiversitas*, *18*(4), 1394–1405. <https://doi.org/10.13057/biodiv/d180415>
- Iverson, C. (2009). *AMA Manual of Style*.

<https://doi.org/10.1093/jama/9780195176339.001.0001>

Jiang, R., Zhang, X., & Zhang, M. Q. (2016). *Basics of Bioinformatics*.
<https://doi.org/10.1007/978-3-642-38951-1>

Karaboga, D., Basturk, B., & Ozturk, C. (2007). Artificial beecolony (ABC) optimization algorithm for training feed-forward neural networks. *Model. Decis. Artif. Intell.*, 3, 318–319.

Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). *Learning Spark*. O'Reilly Media, Inc.

Kashtan, N., Itzkovitz, S., Milo, R., & Alon, U. (2004). Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics*, 20(11), 1746–1758.
<https://doi.org/10.1093/bioinformatics/bth163>

Kaur, A., Singh, V. P., & Singh Gill, S. (2019). The Future of Cloud Computing: Opportunities, Challenges and Research Trends. *Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018*, (February), 213–219. <https://doi.org/10.1109/I-SMAC.2018.8653731>

Keith, J. M. (Ed.). (n.d.). *Bioinformatics* (2nd ed.). <https://doi.org/10.1007/978-1-4939-6622-6>

Kennedy, J., & Eberhart, R. (1995). *Particle swarm optimization*.

Kerr, K. C. R., Stoeckle, M. Y., Dove, C. J., Weigt, L. A., Francis, C. M., & Hebert, P. D. N. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7(4), 535–543.
<https://doi.org/10.1111/j.1471-8286.2007.01670.x>

Kress, W., & Erickson, D. (2012). DNA Barcodes: Methods and Protocols. *Methods in Molecular Biology (Clifton, N.J.)*, 858, 3–8.

https://doi.org/10.1007/978-1-61779-591-6_1

Kumar Pandey, B., Kumar Pandey, S., & Pandey, D. (2011). A Survey of Bioinformatics Applications on Parallel Architectures. *International Journal of Computer Applications*, 23(4), 21–25. <https://doi.org/10.5120/2877-3744>

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6(70), 1.

Lesk, A. M. (2019). Bioinformatics. Retrieved October 2, 2019, from Encyclopædia Britannica, inc. website: <https://www.britannica.com/science/bioinformatics>

Li, X., Yang, F., Henry, R., Rossetto, M., Wang, Y., & Chen, S. (2014). Plant DNA barcoding: From gene to genome. *Biological Reviews of the Cambridge Philosophical Society*, 90. <https://doi.org/10.1111/brv.12104>

Luscombe, N. M., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? An introduction and overview. In *Yearbook of Medical Informatics* (Vol. 10). <https://doi.org/10.1055/s-0038-1638103>

Miao, L. I. U., Xi-wen, L. I., Bao-sheng, L., Lu, L. U. O., & Yue-ying, R. E. N. (2019). Species identification of poisonous medicinal plant using DNA barcoding. *Chinese Journal of Natural Medicines*, 17(8), 585–590. [https://doi.org/10.1016/S1875-5364\(19\)30060-3](https://doi.org/10.1016/S1875-5364(19)30060-3)

Miller, W. (2006). An Introduction to Bioinformatics Algorithms. In *Journal of the American Statistical Association* (Vol. 101). <https://doi.org/10.1198/jasa.2006.s110>

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594), 824–827. <https://doi.org/10.1126/science.298.5594.824>

Omar, H. K., & Jumaa, A. K. (2019). Big Data Analysis Using Apache Spark MLlib and Hadoop HDFS with Scala and Java. *Kurdistan Journal of Applied*

- Research*, 4(1), 7–14. <https://doi.org/10.24017/science.2019.1.2>
- Parida, L. (2007). Discovering topological motifs using a compact notation. *Journal of Computational Biology*, 14(3), 300–323. <https://doi.org/10.1089/cmb.2006.0142>
- Pavesi, G., Mereghetti, P., Mauri, G., & Pesole, G. (2004). Weeder web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Research*, 32(WEB SERVER ISS.). <https://doi.org/10.1093/nar/gkh465>
- Poli, R., Kennedy, J., & Blackwell, T. (2007). Particle swarm optimization: An overview. *Swarm Intelligence*, 1(1), 33–57. <https://doi.org/10.1007/s11721-007-0002-0>
- Purnomo, Sudjino, Trijoko, & Hadisusanto, S. (2009). *Biologi*. Jakarta: Pusat Perbukuan, Departemen Pendidikan Nasional.
- Qader, N., & Al-Khafaji, H. K. (2014). Motif Discovery and Data Mining in Bioinformatics. *International Journal of Computers & Technology*, 13(1), 4082–4095. <https://doi.org/10.24297/ijct.v13i1.2932>
- Ramadhan, H. A., & Putri, D. A. (2018). *Big Data, Kecerdasan Buatan, Blockchain, dan Teknologi Finansial di Indonesia*. 1–66. Retrieved from <https://aptika.kominfo.go.id/wp-content/uploads/2018/12/Kajian-Kominfo-CIPG-compressed.pdf>
- Rana, J. (2014). *Introduction To Bioinformatics*. (December 2012).
- Reinsel, D., Gantz, J., & Rydning, J. (2018). Data Age 2025: The Digitization of the World From Edge to Core. *International Data Corporation*, (November), 28.
- Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47.

<https://doi.org/10.1109/CTS.2013.6567202>

Shao, L., & Chen, Y. (2009). Bacterial foraging optimization algorithm integrating tabu search for motif discovery. *2009 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2009*, 415–418. <https://doi.org/10.1109/BIBM.2009.12>

Shi, Y., & Eberhart, R. C. (1999). Empirical study of particle swarm optimization. *Proceedings of the 1999 Congress on Evolutionary Computation, CEC 1999*, 3, 1945–1950. <https://doi.org/10.1109/CEC.1999.785511>

Shoro, A. G., & Soomro, T. R. (2015). *Big Data Analysis: Apache Spark Perspective*. (JANUARY).

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop distributed file system. *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies, MSST2010*, 1–10. <https://doi.org/10.1109/MSST.2010.5496972>

Sinden, R. R. (2012). DNA Structure and Function. *The Febs Journal*, 282, 1–398. <https://doi.org/10.1111/febs.13307>

Sommerville, I. (2011). *Software Engineering*. Addison-Wesley.

Srivastava, P., & Khan, R. (2018). A Review Paper on Cloud Computing. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(6), 17. <https://doi.org/10.23956/ijarcsse.v8i6.711>

Su, X. (n.d.). *Introduction to Big Data*.

Sun, P., Lu, X., Xu, C., Sun, W., & Pan, B. (2020). Understanding of COVID-19 based on current evidence. *Journal of Medical Virology*, 92(6), 548–551. <https://doi.org/10.1002/jmv.25722>

Sunaryo, W. (2016). Review : Aplikasi DNA barcoding untuk analisis keragaman

genetik lai-durian (*Durio zibethinus* x *kutejensis*) asal Kalimantan Timur. *PROS SEM NAS MASY BIODIV INDON*, 1(September 2015). <https://doi.org/10.13057/psnmbi/m010602>

Talbi, E.-G. (2009). *Metaheuristics: from design to implementation* (Vol. 74; H. Wiley, Ed.).

Thieman, W. J., & Palladino, M. A. (2013). *Introduction to Biotechnology* (4th ed.). USA: Pearson Education.

Trabuco, L., & Villa, E. (n.d.). *Case Study : DNA*. 1–35.

U.S. National Library of Medicine. (2019). What is DNA? Retrieved October 9, 2019, from <https://ghr.nlm.nih.gov/primer/basics/dna>

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losick, R. (2003). *Molecular Biology of the Gene* (5th ed.). Cold Spring Harbor Laboratory Press.

Wernicke, S. (2006). Efficient Detection of Network Motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics / IEEE, ACM*, 3, 347–359. <https://doi.org/10.1109/TCBB.2006.51>

Wilson, J.-J., Sing, K.-W., & Jaturas, N. (2019). DNA Barcoding: Bioinformatics Workflows for Beginners. *Encyclopedia of Bioinformatics and Computational Biology*, (2012), 985–995. <https://doi.org/10.1016/b978-0-12-809633-8.20468-8>

Xiong, J. (2006). *Essential Bioinformatics*. Cambridge University Press.

Yang, C.-H., Yu-Tang, L., & Chuang, L.-Y. (2011). DNA Motif Discovery Based on Ant Colony Optimization and Expectation Maximization. *Lecture Notes in Engineering and Computer Science*, 1.

Yang, C. H., Liu, Y. T., & Chuang, L. Y. (2011). DNA motif discovery based on

- ant colony optimization and expectation maximization. *IMECS 2011 - International MultiConference of Engineers and Computer Scientists 2011*, 1(March), 169–174.
- Yang, D., & Wang, J. (2008). UPNT : Uniform Projection and Neighbourhood Thresholding method for motif discovery. *International J. Bioinformatics Research and Applications*, 4(1), 96–106.
- Yang, X. (2017). *Optimization Algorithms Optimization and Metaheuristic Algorithms in Engineering*. (March). <https://doi.org/10.1007/978-3-642-20859-1>
- Yu, Q., Huo, H., Chen, X., Guo, H., Vitter, J. S., & Huan, J. (2015). An Efficient Algorithm for Discovering Motifs in Large DNA Data Sets. *IEEE Transactions on Nanobioscience*, 14(5), 535–544. <https://doi.org/10.1109/TNB.2015.2421340>
- Zaharia, M., Xin, R. S., Wendell, P., Das, T., Armbrust, M., Dave, A., ... Gonzalez, J. (2016). Apache spark: A Unified Apache Spark: Engine for Big Data Processing. *Communications of the Acm*, 59(11). <https://doi.org/10.1145/2934664>
- Zhong, Y., Lin, J., Wang, L., & Zhang, H. (2018). Discrete comprehensive learning particle swarm optimization algorithm with Metropolis acceptance criterion for traveling salesman problem. *Swarm and Evolutionary Computation*, 42, 77–88. <https://doi.org/10.1016/j.swevo.2018.02.017>
- Ziv, E., Koytcheff, R., Middendorf, M., & Wiggins, C. (2005). Systematic identification of statistically significant network measures. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, 71, 16110. <https://doi.org/10.1103/PhysRevE.71.016110>