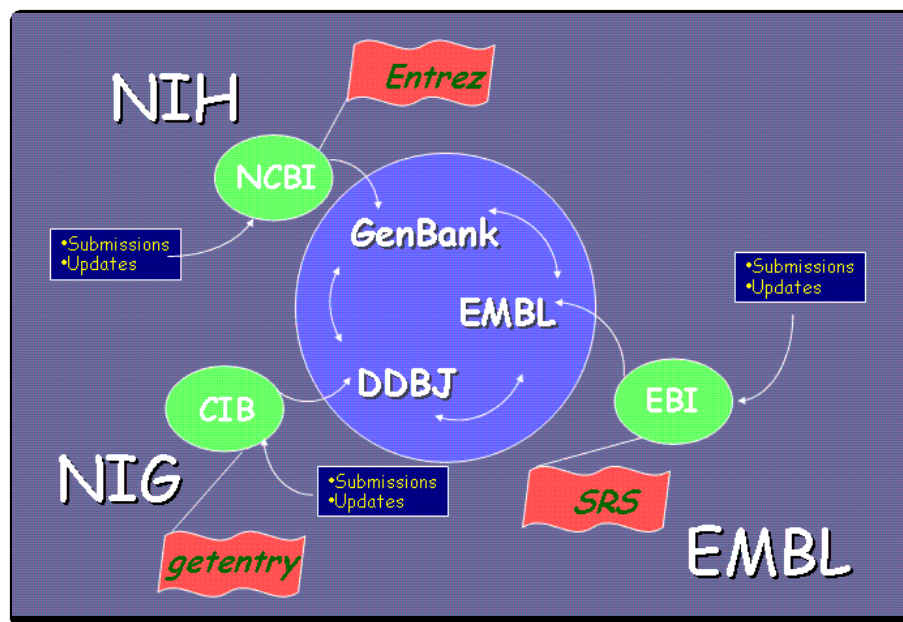


Pemanfaatan basis data sekuen dalam riset bioinformatika pada tanaman perkebunan

Genomika tanaman telah mengalami peningkatan jumlah data sekuen yang sangat pesat, terutama dikarenakan revolusi dalam teknologi sekuensing generasi berikutnya. Revolusi tersebut telah mendorong para peneliti biologi molekuler untuk memanfaatkan data yang “menggungung” tersebut dan memberikan makna lebih dibalik sekuen genomika. Saat ini, ketersediaan basis data besar seperti INSDC dan basis data spesifik dapat dimanfaatkan untuk mendorong riset terkait bioinformatika, terutama pada tanaman perkebunan.

Genomika tanaman telah mengalami peningkatan jumlah data sekuen yang sangat pesat, terutama dikarenakan revolusi dalam teknologi sekuensing generasi berikutnya (*Next Generation Sequencing*, NGS). Dengan keberadaan data sekuen yang “menggungung” tersebut, peneliti biologi molekuler pada masa sekarang mendapatkan tantangan untuk dapat mengidentifikasi fitur dari DNA genomik serta memberikan deskripsi fungsi biologisnya. Tantangan tersebut dijawab oleh Konsorsium Internasional dari Kolaborasi Data Sekuen (*The International Nucleotide Sequence Data Collaboration*, INSDC dalam situs <http://www.insdc.org/>) dengan menciptakan interkoneksi diantara basis-basis data sekuen yang ada di dunia seperti *GenBank* dari *National Center for Biotechnology Information* (NCBI dalam situs <https://www.ncbi.nlm.nih.gov/>), *DNA Data Bank of Japan* (DDBJ dalam situs <http://www.ddbj.nig.ac.jp/>) dan *European Nucleotide Archive* (ENA dalam situs <http://www.ebi.ac.uk/ena>) dari *European Bioinformatics Institute* (EBI) [1-4].



Ilustrasi interkoneksi antara tiga basis data besar dunia yaitu NCBI, DDBJ dan EBI-ENA.
Gambar disadur dari <https://www.ncbi.nlm.nih.gov/>.

Kolaborasi INSDC merupakan basis data yang tergolong “umum”, “besar” dan banyak digunakan oleh peneliti-peneliti di seluruh dunia. Saat ini, penambahan basis data sekuen terus terjadi dan akan terus bertambah dalam tahun-tahun mendatang, seiring dengan semakin turunnya biaya untuk melakukan sekuensing genomika, transkriptomika, epigenomika, serta

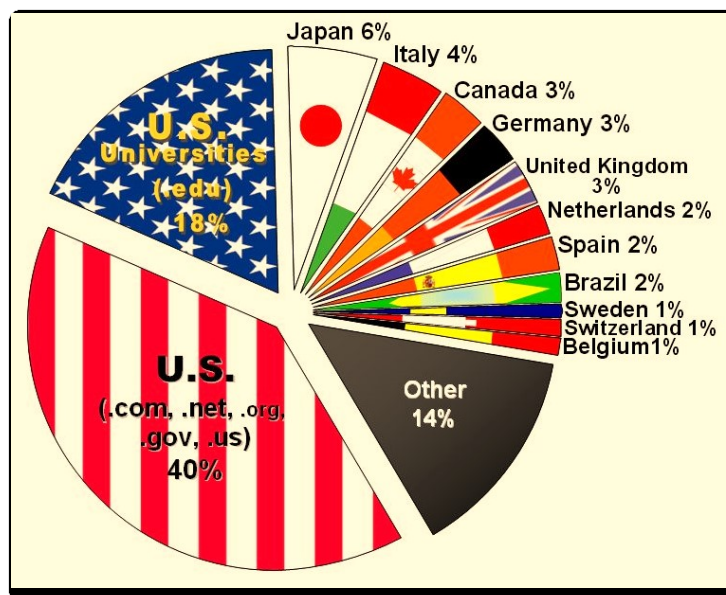
proteomika. Perkembangan riset yang spesifik pada tanaman tertentu juga memunculkan tipe basis data spesifik dimana informasi genomik yang dianotasi berasal dari 1 jenis kingdom, famili atau bahkan genus. Beberapa contoh basis data tersebut adalah *Phytozome* untuk basis data khusus tanaman (dalam situs www.phytozome.net/) [5], *TropGeneDB* untuk basis data khusus tanaman tropis (dalam situs <https://tropgenedb.cirad.fr/>) [6], *TAIR* untuk basis data tanaman model biologi molekuler *Arabidopsis thaliana* (dalam situs <https://www.arabidopsis.org/>) [7], *Banana Genome Hub* untuk basis data khusus tanaman pisang (dalam situs <http://banana-genome-hub.southgreen.fr/>) [8], *PeanutBase* untuk basis data tanaman kacang (dalam situs <https://peanutbase.org/>) [9] dan *MaizeDB* untuk basis data khusus tanaman jagung (dalam situs <https://www.maizegdb.org/>) [10].

Lebih jauh dari itu, basis data spesifik dapat lebih disempitkan lagi luasan data yang dirangkum seperti mengutamakan tema riset tertentu. Tema riset ini umumnya terkonservasi pada beberapa tanaman seperti basis data terkait faktor transkripsi, faktor pemicu epigenomika, atau protein terkonservasi yang memiliki peran penting dalam sel. Beberapa contoh basis data semacam ini adalah *TransportDB* untuk basis data informasi terkait transporter membran dalam sekuen genomika (dalam situs www.membranetransport.org/) [11], *miRNA Digger* untuk basis data spesifik RNA mikro (dalam situs http://www.bioinfoclab.cn/miRNA_Digger/index.html) [12], *PlantTFDB* untuk basis data spesifik yang menyimpan faktor transkripsi pada tanaman (dalam situs <http://plantfdb.cbi.pku.edu.cn/>) [13], *lncRNADB* untuk basis data spesifik RNA panjang yang tidak menyandi protein (*long non coding RNAs*, *lncRNAs* dalam situs <http://www.lncrnadb.org/>) [14], dan lain sebagainya.

Keberadaan basis-basis data tersebut juga menunjukkan bahwa riset *in silico* atau riset laboratorium kering sedang menjadi tren serta mampu menjadi alat konfirmasi dan validasi untuk hasil riset “basah” yang dilakukan di laboratorium. Riset *in silico* dapat pula digunakan untuk mempercepat prediksi dan menajamkan riset “basah”, seperti desain marka untuk seleksi pemuliaan tanaman. Hal tersebut diperkuat dengan fakta bahwa perkembangan teknologi komputasi yang pesat sudah memberikan kemudahan analisis bioinformatika skala kecil pada komputer *desktop* atau *laptop*. Dalam riset pada tanaman perkebunan di dunia internasional, pemanfaatan basis data bioinformatika telah banyak digunakan dalam komparasi genomika seperti pada tanaman karet [15-17], tanaman kelapa sawit [18, 19], tanaman kurma [20], tanaman gula [21], tanaman kakao [22, 23], tanaman kopi [24, 25], dan tanaman teh [26, 27].

Komparasi famili gen telah didemonstrasikan sangat efektif menggunakan pendekatan bioinformatika pada tanaman karet. Aktor-aktor molekuler penting dalam respon terhadap etilen serta produksi tanaman karet telah berhasil diidentifikasi dan dikarakterisasi menggunakan metode tersebut [28]. Gen *SHELL* yang berperan vital menentukan ketebalan kulit mesokarp sawit diidentifikasi menggunakan pendekatan bioinformatika (perbandingan sekuen dari tiap varietas sawit) dan dikombinasikan dengan validasi pada tingkat laboratorium “basah” [29]. Seleksi dari ratusan ribu sekuen DNA pada tanaman teh telah berhasil menemukan banyak senyawa penting yang menyusun aroma teh [26]. Seleksi yang sama dilakukan pada tanaman kopi dan ditemukan satu famili protein N-metiltransferase (NMT) yang berperan spesifik dalam biosintesis kafein [25]. Pada tanaman gula, riset bioinformatika telah membantu mengidentifikasi hubungan kekerabatan yang sangat dekat antara tanaman

sorgo dan gula namun ekspresi gen-gen ortolognya (sejenis) sangat berbeda [21]. Tanaman gula telah berevolusi secara genomika untuk meningkatkan efektivitas dalam menyimpan sukrosa pada batang tebu yang akhirnya dimanfaatkan manusia untuk kebutuhannya [30].



Ilustrasi jumlah kunjungan, akses dan pemanfaatan situs NCBI untuk riset bioinformatika berdasarkan negara pengakses hingga tahun 2015. Gambar disadur dari <https://www.ncbi.nlm.nih.gov/>.

Menilik dari beberapa kasus di atas, riset bioinformatika aplikatif sangat menjanjikan dan telah memberikan kontribusi dalam percepatan riset pada bidang perkebunan. Meskipun demikian, di Indonesia, laju akses *traffic* ke dalam NCBI masih dibawah 1% dari total *traffic* per tahunnya memperlihatkan masih minimnya penggunaan basis data seperti NCBI. Hal tersebut menunjukkan masih belum tingginya minat peneliti-peneliti di Indonesia untuk melaksanakan riset terkait bioinformatika. Satu fakta yang menarik bahwa INDSC mengajak negara-negara dengan akses di atas 6% untuk berkolaborasi dalam riset unggulan. Dengan kekayaan hayati yang dimiliki Indonesia, terutama pada tanaman pertanian dan perkebunan, potensi kolaborasi tersebut tentu saja masih terbuka. Dalam beberapa tahun ke depan, sudah saatnya Indonesia terlibat lebih banyak lagi dalam riset-riset bioinformatika berskala internasional.

Referensi

1. van Dijk A-J. Plant Genomics Databases. 1 ed. Dijk A-Jv, editor. New York: Humana Press; 2017.
2. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Research. 2017;45.
3. Toribio AL, Alako B, Amid C, Cerdeño-Tarrága A, Clarke L, Cleland I, et al. European Nucleotide Archive in 2016. Nucleic Acids Research. 2017;45.
4. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 2016;44.

5. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*. 2012;40.
6. Ruiz M, Sempere G, Hamelin C. Using TropGeneDB: a database containing data on molecular markers, QTLs, maps, genotypes, and phenotypes for tropical crops. In: Van Dijk Aalt DJ, editor. *Plant genomics databases: methods and protocols*. Methods in Molecular Biology. New York, USA: Springer; 2017. p. 161-72.
7. Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, et al. The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*. 2012;40.
8. Droc G, Larivière D, Guignon V, Yahiaoui N, This D, Garsmeur O, et al. The Banana Genome Hub. Database: The Journal of Biological Databases and Curation. 2013;1-14.
9. Dash S, Cannon EKS, Kalberer SR, Farmer AD, Cannon SB. Chapter 8 - PeanutBase and Other Bioinformatic Resources for Peanut. In: Wilson RF, editor. *Peanuts*: AOCS Press; 2016. p. 241-52.
10. Andorf CM, Cannon EK, Portwood JL, Gardiner JM, Harper LC, Schaeffer ML, et al. MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Research*. 2016;44.
11. Elbourne Liam DH, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Research*. 2017;45.
12. Yu L, Shao C, Ye X, Meng Y, Zhou Y, Chen M. miRNA Digger: a comprehensive pipeline for genome-wide novel miRNA mining. *Scientific Reports*. 2016;6:18901.
13. Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, et al. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res*. 2011;39.
14. Amaral PP, Clark MB, Gascoigne DK, Dinger ME, Mattick JS. lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Research*. 2011;39.
15. Tang C, Yang M, Fang Y, Luo Y, Gao S, Xiao X, et al. The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants*. 2016:16073.
16. Lau N-S, Makita Y, Kawashima M, Taylor TD, Kondo S, Othman AS, et al. The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific Reports*. 2016;6:28594.
17. Rahman AYA, Usharraj A, Misra B, Thottathil G, Jayasekaran K, Feng Y, et al. Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics*. 2013;14(1):75.
18. Singh R, Ong-Abdullah M, Low E-TL, Manaf MAA, Rosli R, Nookiah R, et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*. 2013;500(7462):335-9.
19. Jin J, Lee M, Bai B, Sun Y, Qu J, Rahmadsyah, et al. Draft genome sequence of an elite Dura palm and whole-genome patterns of DNA variation in oil palm. *DNA Research*. 2016.
20. Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, et al. Genome sequence of the date palm *Phoenix dactylifera* L. *Nat Commun*. 2013;4.
21. de Setta N, Monteiro-Vitorello C, Metcalfe C, Cruz GM, Del Bem L, Vicentini R, et al. Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics*. 2014;15(1):540.

22. Motamayor JC, Mockaitis K, Schmutz J, Haiminen N, III DL, Cornejo O, et al. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biology*. 2013;14(6):r53.
23. Argout X, Salse J, Aury J-M, Guiltinan MJ, Droc G, Gouzy J, et al. The genome of *Theobroma cacao*. *Nat Genet*. 2011;43(2):101-8.
24. Mueller L, Strickler S, Domingues D, Pereira L, Andrade A, Marraccini P, et al. Towards a better understanding of the *Coffea arabica* genome structure. Proceedings of the 25th international conference on coffee science: ASIC; 2015. p. p 42-5.
25. Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, et al. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*. 2014;345(6201):1181-4.
26. Shi C-Y, Yang H, Wei C-L, Yu O, Zhang Z-Z, Jiang C-J, et al. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics*. 2011;12(1):131.
27. Xia E-H, Zhang H-B, Sheng J, Li K, Zhang Q-J, Kim C, et al. The Tea Tree Genome Provides Insights into Tea Flavor and Independent Evolution of Caffeine Biosynthesis. *Molecular Plant*. 2017.
28. Putranto RA, Montoro P. The *Hevea brasiliensis* AP2/ERF superfamily: From ethylene signalling to latex harvesting and physiological disease response. *Menara Perkebunan*. 2016;84(1):49-62.
29. Singh R, Low E-TL, Ooi LC-L, Ong-Abdullah M, Ting N-C, Nagappan J, et al. The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature*. 2013;500(7462):340-4.
30. McCormick AJ, Cramer MD, Watt DA. Differential Expression of Genes in the Leaves of Sugarcane in Response to Sugar Accumulation. *Tropical Plant Biology*. 2008;1(2):142-58.