

Review

Pemanfaatan bioinformatika dalam bidang pertanian dan kesehatan

The utilization of bioinformatics in the field of agriculture and health

Arli Aditya PARIKESIT^{1)*}, Dito ANUROGO²⁾ & Riza Arief PUTRANTO³⁾

¹⁾Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences, Jl. Pulomas Barat Kav.88, Jakarta, 13210, Indonesia

²⁾Fakultas Kedokteran Universitas Gadjah Mada, Yogyakarta, Jl. Farmako Sekip Utara, Yogyakarta, 55281, Indonesia

³⁾Pusat Penelitian Bioteknologi dan Bioindustri Indonesia, Jl. Taman Kencana No 1, Bogor 16128, Indonesia

Diterima tgl 5 Januari 2017 / disetujui tgl 4 Agustus 2017

Abstract

Bioinformatics can be used to manage the data storage resulted from in silico molecular biology experiments. Off-network (offline) applications require large computing resources, in which researchers in the bioinformatics field of agriculture and health sectors do not necessarily possess. This review paper addressed examples of affordable and applicable in silico analytical cases in both mentioned sectors. Genome sequence analysis and in silico drug design using (1) a computational method, pharmacokinetic parameter prediction, (2) Computer Aided Design and Drafting (CADD) technology, (3) potential protein action prediction, (4) OMICs application in stem cell biology, and (5) lncRNAs based database computing internet sites is one of examples. In agriculture, bioinformatics-based research has been used in (1) the development of molecular markers; (2) the design of primer for differential gene expression analysis; (3) the development of genetic maps; and (4) gene expression analysis. Further application of bioinformatics also targets the design of applicative products for pest control and the protection of plant varieties in the farm. Through this example, novice researchers in the bioinformatics field of agriculture and health sectors can conduct sophisticated research using standard computer tools, internet networks, and sufficient knowledge about bioinformatics. On the other hand, multidisciplinary collaboration between these scientists can be carried out through social networking. The synergy can be directed to improve computing capabilities and data analysis via procurement of computing resources and use of public information clusters.

[Key words: genome sequences, in silico drug design, online, bioinformatics, health, agriculture.]

^{*} Penulis korespondensi: arli.parikesit@i3l.ac.id

Abstrak

Bioinformatika dapat digunakan dalam manajemen informasi di bidang penyimpanan data *in silico* dari eksperimen biologi molekuler. Aplikasi luar jaringan (*luring*) memerlukan sumber daya komputasi yang besar, yang belum tentu dimiliki oleh para peneliti dalam bidang bioinformatika kesehatan dan pertanian. Kajian ilmiah ini membahas contoh kasus analisis *in silico* yang terjangkau dan aplikatif dalam bidang kesehatan dan pertanian. Contoh kasus tersebut adalah analisis sekuen genom dan desain obat *in silico*, menggunakan pendekatan metode komputasional, prediksi parameter farmakokinetik, teknologi *Computer Aided Design and Drafting* (CADD), prediksi potensial aksi protein, aplikasi OMICs pada biologi sel punca, hingga komputasi basis data *lncRNAs* berbasis situs internet. Pada bidang pertanian, penelitian berbasis bioinformatika telah digunakan dalam (1) pengembangan penanda molekuler; (2) desain primer untuk analisis ekspresi gen diferensial; (3) pengembangan peta genetik; dan (4) analisis ekspresi gen. Pemanfaatan bioinformatika dalam ilmu terapan dibidang pertanian juga menasar desain produk aplikatif untuk pengendalian hama dan perlindungan varietas tanaman. Melalui contoh tersebut, peneliti pemula dibidang bioinformatika kesehatan dan pertanian dapat melakukan penelitian canggih hanya dengan alat komputer standar, jaringan internet, dan pengetahuan mencukupi tentang bioinformatika. Disisi lain, sinergi dan kolaborasi antar peneliti multi-disipliner dapat dilakukan melalui penggunaan jejaring sosial. Sinergi tersebut dapat diarahkan untuk meningkatkan kemampuan komputasi dan analisis data melalui pengadaan sumber daya komputasi dan penggunaan kluster informatika publik.

[Kata kunci: sekuen genom, desain obat *in silico*, daring, bioinformatika, kesehatan, pertanian.]

Pendahuluan

Dalam eksperimen biologi molekuler, data yang dihasilkan perlu diolah dengan pendekatan khusus untuk menjadi informasi yang berguna. Bioinformatika, ilmu berbasis multidisipliner yang menggabungkan pendekatan biologi molekuler dan teknik informatika, dapat digunakan dalam manajemen informasi tersebut (Saeys *et al.*, 2007; Whitfield *et al.*, 2006; Eisenhaber *et al.*, 2009; Parikesit, 2009; Searls, 2012; Goodman & Dekhtyar, 2014). Untuk keperluan tersebut, penelitian bioinformatika membutuhkan sumber daya komputasi dalam jumlah tertentu. Sumber daya komputasi yang besar didefinisikan sebagai minimal sebuah komputer *desktop* kelas *workstation* dan maksimal sebuah komputer super dengan jumlah lebih dari 10.000 prosesor. Di negara maju seperti Amerika Serikat yang berbasis sumber daya komputasi, pemanfaatan ilmu bioinformatika dalam manajemen data masif sudah menjadi bisnis besar (Bourne *et al.*, 2015).

Aplikasi luar jaringan (luring) atau dalam bahasa Inggris disebut *offline* memerlukan sumber daya komputasi yang besar. Di Indonesia, tidak banyak peneliti dalam bidang bioinformatika pertanian dan kesehatan memiliki sumber daya komputasi yang diperlukan sehingga berpotensi membatasi penelitian biologi berbasis komputasi. Keterbatasan tersebut dapat diatasi dengan penggunaan aplikasi bioinformatika secara daring (dalam jaringan) atau dalam bahasa Inggris disebut *online*. Di sisi lain, keberadaan basis data secara daring belum banyak dikenal peneliti untuk dimanfaatkan dalam mempercepat penelitian dalam bidang pertanian dan kesehatan.

Dalam kajian ilmiah ini, pembahasan dilakukan secara umum pada pemanfaatan sekuen genom dan aplikasi bioinformatika dalam bidang pertanian dan kesehatan. Pembahasan lebih spesifik dilakukan dengan studi kasus terkait desain obat secara *in silico*. Materi pembahasan dalam desain obat secara *in silico* meliputi langkah-langkah analisis seperti: (1) metode komputasional pada polifarmakologi, (2) prediksi parameter farmakokinetik, (3) teknologi CADD pada desain obat secara *in silico*, (4) prediksi potensial aksi, (5) aplikasi OMICs di biologi sel punca, dan (6) komputasi basis data *lncRNAs* berbasis situs internet. Kajian ilmiah ini diharapkan dapat memperkenalkan kembali aplikasi daring bioinformatika yang dapat digunakan untuk analisis komputasi yang terjangkau dalam penelitian pertanian dan kesehatan.

1. Pemanfaatan sekuen genom secara daring

Saat ini, basis data genom berbagai organisme

sudah dapat diakses secara daring pada situs internet: <http://www.ncbi.nlm.nih.gov/genome/>. Konsorsium internasional dari basis data genom juga telah membuat *National Center for Biotechnology Information* (NCBI) terhubung dengan *European Nucleotide Archive* (ENA) dan *DNA Data Bank of Japan* (DDBJ) sehingga memudahkan akses sekuen genom (NCBI, 2013). Aplikasi daring seperti *The Basic Local Alignment Search Tool* (BLAST), konstruksi pohon filogenetik, dan *Map Viewer* telah tersedia untuk melakukan analisis pada genom-genom tersebut (Tambunan & Parikesit, 2010). Dalam analisis lebih lanjut, aplikasi teknis seperti prediksi gen, anotasi domain protein, dan modifikasi pasca-translasi dapat digunakan untuk memberi makna biologis pada data sekuen (Parikesit *et al.*, 2014; Tambunan *et al.*, 2007).

Dalam bidang pertanian, aplikasi BLAST2GO sangat umum digunakan dalam anotasi sekuen genomik dan proteomik dalam jumlah besar. Metode anotasi menggunakan aplikasi tersebut sering disebut sebagai anotasi otomatis (Conesa *et al.*, 2005). Secara umum, analisis tersebut kemudian dilanjutkan dengan anotasi manual menggunakan aplikasi Geneious® sebagai piranti lunak *front end* (Kearse *et al.*, 2012). Penelitian terkait sekuensing genom tanaman telah dilakukan pada padi dan jagung sejak 19 tahun yang lalu (COE 1998; Goff *et al.*, 2002). Beberapa contoh lain genom lengkap dari spesies tanaman yang sudah disequens dalam kurun waktu 10 tahun terakhir adalah tanaman kentang, tomat, kacang kedelai, rumput, kelapa sawit, karet, kopi dan tebu (Haas *et al.*, 2009; Schmutz *et al.*, 2010; Vogel *et al.*, 2010; Xu *et al.*, 2011; Sato *et al.*, 2012; Singh *et al.*, 2013; Denoeud *et al.*, 2014; de Setta *et al.*, 2014; Lau *et al.*, 2016; Tang *et al.*, 2016). Sebagian basis data terkait genom tanaman tersebut bersifat publik atau disebut *open access* sehingga dapat digunakan dalam penelitian lanjutan oleh para peneliti di seluruh dunia.

Disisi lain, dalam bidang kesehatan, informasi genom banyak digunakan untuk interaksi protein-protein dalam rangka mempelajari basis molekuler dari sebuah penyakit (Parikesit, 2010). Salah satu contoh dalam kasus tersebut adalah identifikasi sifat patogenesis dari protein huntingtin (htt) penyebab *Huntington's disease*, sebuah penyakit degeneratif autosomal yang menyebabkan ketidakseimbangan kognitif dan disfungsi syaraf motorik pada penderitanya. Studi interaksi protein htt dan targetnya telah mengidentifikasi 104 protein terkait dengan penyakit tersebut. Metode terapi kemudian dapat dirancang oleh dokter disesuaikan dari hasil penelitian tersebut.

2. Bioinformatika dalam bidang pertanian

Bioinformatika telah sejak lama digunakan untuk mendukung eksperimen laboratorium pada bidang pertanian. Baru-baru ini, hadirnya teknologi *Next Generation Sequencing* (NGS) telah mengubah paradigma dalam penelitian berbasis bioinformatika. Penelitian berbasis bioinformatika telah digunakan dalam (1) pengembangan penanda molekuler (Priyono & Putranto, 2014); (2) desain primer untuk analisis ekspresi gen diferensial (Budiani *et al.*, 2016); (3) pengembangan peta genetik (Priyono & Putranto, 2016); dan (4) analisis ekspresi gen (Putranto *et al.*, 2015). Kombinasi antara bioinformatika dalam anotasi gen serta desain primer dan analisis ekspresi gen menggunakan Real-Time PCR telah berhasil memetakan ekspresi 35 gen *Ethylene Response Factors* (ERFs) pada tanaman karet (*Hevea brasiliensis*). Tiga gen kemudian dipilih untuk analisis fungsional melalui transgenesis (Putranto *et al.*, 2015). Di sisi lain, pemetaan genetik pada tanaman kopi (*Coffea canephora*) juga telah memanfaatkan bioinformatika. Sekuen genom tanaman kopi dari the *International Coffee Genomics Network* (ICGN) digunakan sebagai referensi untuk mencocokkan marka RFLP, SSR, dan SNP pada peta genetik sepanjang 1.471 cM dengan kepadatan marka 0,5 per cM (Priyono & Putranto, 2016).

Pada skala yang lebih besar, alur kerja (*workflow*) dan skema (*pipeline*) bioinformatika juga telah membantu berbagai analisis transkriptomik dan genomik pada tanaman perkebunan seperti identifikasi famili gen penting (Liu *et al.*, 2016, Zou *et al.*, 2015, Dou *et al.*, 2014, Duan *et al.*, 2013), studi asosiasi gen (Teh *et al.*, 2016; Liu *et al.*, 2016; Zou *et al.*, 2015; Dou *et al.*, 2014; Duan *et al.*, 2013; Ting *et al.*, 2013), identifikasi SNP (Zhou *et al.*, 2016), dan identifikasi QTL (Ting *et al.*, 2013). Skema kerja dari CIRAD, Prancis yang disebut basis data *ESTtik* (<http://esttik.cirad.fr/>) telah berhasil digunakan untuk mengidentifikasi superfamili APETALA2 /ETHYLENE RESPONSE FACTORS (AP2/ERF) pada tanaman karet dari hasil sekuensing dengan teknologi *Whole Genome Sequencing* Roche 454 (Duan *et al.*, 2013). Dari pustaka sekuen (*global library*), sebanyak 173 contig AP2/ERF telah diidentifikasi dengan serangkaian analisis *in silico* berdasarkan sekuen asam amino dari domain AP2. Sementara, tim dari CATAS, Tiongkok melakukan identifikasi famili metacaspase dari tanaman karet menggunakan alur kerja analisis genom komparatif. Perbandingan genom dilakukan menggunakan data genom publik tanaman karet yang tersedia di NCBI disejajarkan dengan genom *Arabidopsis thaliana* di

basis data TAIR (Liu *et al.*, 2016). Runtutan analisis lengkap dengan BLAST, penjajaran multi-sekuen hingga desain primer digunakan untuk mengkonfirmasi identifikasi gen tersebut.

Pemanfaatan bioinformatika dalam ilmu terapan dibidang pertanian juga menyorot desain produk aplikatif untuk mengatasi infeksi virus mosaik tembakau, pengendalian hama terhadap nyamuk *Aedes aegypti*, hama laba-laba, serangga *phytophagous*, serta perlindungan tanaman transgenik (Lawrence & Koundal, 2002; Gatehouse, 2011; Santamaría *et al.*, 2012; Prabahar *et al.*, 2015; Reegan *et al.*, 2016).

3. Bioinformatika dalam bidang kesehatan dan desain obat secara *in silico*

Selaras dengan bidang pertanian, penelitian bioinformatika pada bidang kesehatan senantiasa terkait dengan usaha untuk mengatasi penyakit pada manusia. Salah satu contohnya adalah keberadaan basis data FIND Tuberculosis Strain Bank yang membantu peneliti dalam melakukan uji patogenisitas dan desain obat untuk mengatasi penyakit tuberculosis (TBC) (Tessema *et al.*, 2017). Disisi lain, penelitian terkait penyakit tersebut juga maju pesat dikarenakan keberadaan basis data NGS yang membantu dalam anotasi genom patogen penyebab penyakitnya yaitu *Mycobacterium tuberculosis* (Skvortsov *et al.*, 2013; Han *et al.*, 2015; Mokrousov *et al.*, 2016). Keberadaan basis data tersebut sebagian besar digunakan untuk desain obat secara *in silico* dengan memanfaatkan kemudahan sistem daring. Publikasi terkait alur langkah dalam melakukan desain obat secara *in silico* telah banyak ditemukan.

Penggunaan aplikasi *in-house* secara luring seperti MOE, GROMACS dan AUTODOCK untuk penambatan dan dinamika molekul merupakan metode yang umum (Parikesit & Tambunan, 2013; Parikesit *et al.*, 2016). Namun, jika sumber daya komputasi untuk melakukan aplikasi tersebut tidak tersedia, langkah lain dengan menggunakan *tool* berbasis internet dapat juga dilakukan. Beberapa *tool* yang tersedia secara daring adalah DOCK BLASTER untuk prediksi penambatan molekul-molekul (Irwin *et al.*, 2009), MDWEB dan MDMOBY untuk analisis dinamika molekul (Hospital *et al.*, 2012), ADMET dan DRUGBANK untuk pengembangan basis data obat (van de Waterbeemd & Gifford, 2003; Knox *et al.*, 2011), serta PreADME untuk ADMET *tools* (Lee *et al.*, 2003). Sebuah program daring bernama FAFdrug3 (<http://fafdrugs3.mti.univ-paris-diderot.fr/index.html>) juga dapat digunakan untuk menyaring pustaka terkait senyawa obat dalam jumlah besar.

Pustaka tersebut kemudian dapat menjadi acuan untuk eksperimen skrining obat secara *in silico* atau studi pemodelan terkait (Lagorce *et al.*, 2008). Penggunaan piranti-piranti lunak daring tersebut banyak diarahkan untuk pengembangan obat dengue, avian influenza, dan kanker serviks (Wei *et al.*, 2006; Lim *et al.*, 2013; Bakri *et al.*, 2014). Secara lebih detail, beberapa aspek penting dalam desain obat *in silico* meliputi: (1) metode komputasional pada polifarmakologi, (2) prediksi parameter farmakokinetik, (3) teknologi CADD pada desain obat secara *in silico*, (4) prediksi potensial aksi, (5) aplikasi OMICs di biologi sel punca, (6) komputasi basis data *lncRNAs* berbasis internet dibahas di dalam kajian ilmiah ini.

3.1. Metode komputasional pada polifarmakologi

Saat ini, metode komputasional digunakan secara teknis sebagai *tool* untuk identifikasi kandidat untuk reposisi obat (*drug repurposing*). Reposisi obat merupakan usaha penerapan obat dan senyawa yang sudah dikenal untuk mengobati penyakit baru (Anighoro *et al.*, 2014). Reposisi obat memerlukan penentuan kecepatan dan skala analisis yang memiliki dampak terhadap hasil akhir reposisi pada target-target penyakit baru serta analisis teknoekonomi (biaya penerapan obat). Dengan mempertimbangkan keuntungan finansial dari teknik reposisi obat, metode komputasional mampu digunakan untuk mendapatkan hasil reposisi terbaik dengan cara memilah target yang dituju. Sebagai contoh, bioinformatika digunakan dalam analisis ekspresi gen secara diferensial dalam menentukan gen-gen yang terinduksi akibat penyakit Alzheimer (Siavelis *et al.*, 2016). Dari hasil penelitian tersebut, metode komputasi dapat menghemat langkah penentuan strategi reposisi dengan berfokus pada sekelompok gen penting yang dianggap berperan dalam fenotip akibat Alzheimer.

Beberapa basis data baru untuk reposisi obat juga telah bermunculan seperti Pubmed, Online Mendelian Inheritance, in Man (OMIM), Gene Expression Omnibus, Mouse Genome, Informatics (MGI), Kyoto Encyclopedia of Genes and Genomes (KEGG), BioCarta, IUPHAR Database, ChEMBL, and PubChem (Loging *et al.*, 2011). Sebagai tambahan, pendekatan eksperimental non komputasional seperti skrining fenotip juga digunakan sebagai salah satu usaha reposisi obat. Di masa yang akan datang, integrasi dari pendekatan komputasi dan eksperimental akan menjadi teknik yang baik untuk reposisi obat (Anighoro *et al.*, 2014). Reposisi obat tersebut kemudian dapat digunakan dalam polifarmakologi. Pendekatan polifarmakologi bertujuan mendesain obat-obat multitarget melalui desain ligan dengan

polifarmakologi transformasi menjadi kandidat obat (Anighoro *et al.*, 2014).

3.2. Prediksi parameter farmakokinetik

Parameter farmakokinetika adalah besaran yang diturunkan secara matematis dari model yang berdasarkan hasil pengukuran kadar obat utuh atau metabolitnya dalam darah, urin atau cairan hayati lainnya. Parameter farmakokinetik suatu obat dapat digunakan untuk memperoleh gambaran dan mempelajari suatu kinetika absorpsi, distribusi dan eliminasi di dalam tubuh (Darvas *et al.*, 2000). Pendekatan *in silico* untuk memprediksi parameter farmakokinetik (*absorption, distribution, metabolism, excretion* atau yang sering disingkat menjadi ADME) dipelopori oleh Lipinski dan tim (Lipinski *et al.*, 1997). Dalam melakukan prediksi bioavailabilitas oral (absorpsi interal) dari sebuah senyawa untuk uji klinis tahap 2, karakteristik fisikokimiawi lebih dari 2000 obat-obatan dari World Drug Index, London diwajibkan memenuhi “The Rule of Five”.

Senyawa dikatakan *bioavailable* atau dapat dikonsumsi jika donor yang berikatan hidrogen kurang dari 5, akseptor hidrogen kurang dari 10, berat molekul kurang dari 500, serta sifat lipofilisitas (logP) kurang dari 5 (Terstappen & Reggiani, 2001). Saat ini, basis data SwissADME (<http://www.swissadme.ch/>) yang dapat diakses secara publik untuk membantu relatif lebih mudah pemilihan senyawa yang memiliki “The Rule of Five” (Daina *et al.*, 2017). Disisi lain, prediksi parameter farmakokinetik juga dapat dilakukan pada bidang veteriner sebagai tahap awal untuk uji pra-klinis bagi obat-obatan (Mahmood, 2007).

3.3. Teknologi CADD pada desain obat secara in silico

Secara umum, peralatan *Computer Aided Design and Drafting* (CADD) digunakan untuk merancang bagian demi bagian hingga menjadi produk secara keseluruhan. CADD merupakan *tool* yang sangat berguna di industri pertanian, kesehatan hingga robotika (Kapetanovic, 2008). Pada bidang kesehatan, CADD pada awalnya dikembangkan untuk optimisasi obat (melalui *quantitative structure-activity relationship* yang disingkat QSAR), yang kemudian meluas menuju ke penemuan obat baru (seperti melalui skrining virtual). Sekarang peralatan CADD digunakan baik ke arah penelitian hulu maupun hilir dalam penemuan obat.

Di arah hulu, bioinformatika dan metode *reverse docking* biasanya digunakan untuk mengidentifikasi target dari senyawa obat. Setelah target teridentifikasi, metode *in silico* juga dikembangkan untuk memprediksi struktur 3D sebelum masuk ke

tahapan eksperimental. Berbagai metode komputasional CADD dapat digunakan untuk memprediksi “*target druggability*” sebelum melakukan eksperimen di laboratorium. Target yang “*druggable*” merupakan protein, peptide atau asam nukleat dengan aktivitas yang dapat dimodulasi oleh senyawa obat. Biasanya dapat terdiri dari senyawa kimia molekul kecil (SMOL) atau biologis (BIOL) seperti antibodi atau protein rekombinan (Gashaw *et al.*, 2011). Konsep dari analisis bioinformatika tersebut berbasis pada “*hits*” seperti yang biasa dilakukan dalam BLAST.

Di arah hilir, CADD dapat digunakan untuk memprediksi ADMET secara *in silico* dan simulasi farmakokinetik berbasis fisiologis sebagai model uji praklinis (Tang *et al.*, 2006). Teknik CADD seperti skrining virtual (*virtual screening*) dan desain pustaka (*library design*) juga dapat digunakan untuk mendesain molekul kecil (*probes*) dalam rangka memperjelas mekanisme molekuler yang mendasari berbagai proses biologis. Parameter yang diukur adalah identifikasi atau modifikasi fungsi protein-protein target oleh inhibisi atau aktivasi fungsi-fungsi normalnya (Tang *et al.*, 2006).

3.4. Prediksi potensial aksi

Pendekatan bioinformatika juga dapat digunakan untuk memprediksi potensial aksi dari obat yang didesain (Williams & Mirams, 2015). Sebagai contoh, portal internet Action Potential (<https://chaste.cs.ox.ac.uk/ActionPotential>) dapat digunakan untuk memilih model elektrofisiologi secara matematis dalam menjalankan simulasi elektrofisiologi jantung terhadap senyawa obat. Tool daring tersebut dapat memperkirakan efek rerata persenyawaan yang menghambat IKr, ICaL, INa, IKs, IK1 dan Ito. Piranti lunak *open source* ini menyediakan *interface* yang sederhana. Pengguna dapat memasukkan hal-hal detail tentang persenyawaan untuk afinitas saluran-saluran ion untuk nilai IC50 atau pIC50, menjalankan simulasi, menyimpan hasil untuk pencarian, melihat ringkasan grafik-grafik dari beragam hasil, dan mengeksplor data menjadi format *spreadsheet* (Williams & Mirams, 2015). Komputasi potensial aksi juga dapat dilakukan pada tanaman, seperti mengukur flux ion, sehingga dapat diaplikasikan pada bidang pertanian (Sukhov *et al.*, 2011).

3.5. Aplikasi OMICs di biologi sel punca

Peralatan komputasi berperan penting didalam riset biologi perkembangan sejak tahun 1950-an. Sejak keberhasilan sekuensing genom manusia sekitar 15 tahun yang lalu, penggunaan peralatan komputasi di dalam biologi sel punca dan perkembangannya bergeser dari pemodelan (*modelling*) ke pemrosesan sekumpulan data

molekuler berkapasitas besar (Bian & Cahan, 2016). Aplikasi OMICs pada biologi sel punca hampir selalu diikuti dengan deskripsi awal teknologi baru. Definisi dari OMICs mengacu pada bidang studi biologi yang berakhir -omics, seperti genomika, proteomika, dan metabolomika. Perkembangan OMICs menghasilkan data komprehensif secara bioinformatika dari masing-masing kajian molekuler tertentu. Sebagai contoh, proteomika mempelajari dan mengukur proteom, genomika fungsional mempelajari dan mengkuantifikasi ekspresi gen, serta metabolomika yang mempelajari dan mengukur konsentrasi produk-produk reaksi metabolik (Baharvand *et al.*, 2006). Kombinasi OMICs dan target analitik memungkinkan untuk melakukan observasi dalam biologi sel punca yang berpotensi besar di masa yang akan datang. Biologi sel punca komputasional telah muncul sebagai sub-disiplin tersendiri yang fokus kepada sintesis pemodelan (*modeling synthesis*) dari berbagai aspek di tingkat sistem sel punca dengan data molekuler skala besar (Baharvand *et al.*, 2006; Macaulay & Voet, 2014; Bian & Cahan, 2016).

Sejak isolasi *embryonic stem cells* (ESCs) pada tahun 1981, mekanisme molekuler dari *pluripotent stem cells* (PSCs) yang dapat mempertahankan potensi multi-lineage secara tak terbatas terus menjadi aspek penelitian (Cho *et al.*, 2012). Teknologi PSCs sudah diuji coba pada hewan dan terbukti aman (Jj *et al.*, 2011). Teknik OMICs berperan penting dalam mengungkap mekanisme molekuler dalam regulasi pluripotensi. Beberapa teknik OMICs tersebut adalah:

- **ChIP-chip**
ChIP-chip merupakan teknik *Chromatin ImmunoPrecipitation* (ChIP) dikombinasikan dengan *DNA microarray-based profiling* (chip) untuk karakterisasi interaksi protein dengan DNA genomik.
- **ChIP-Seq**
ChIP-Seq adalah *genome-wide profiling* dari berbagai interaksi protein spesifik dengan DNA genomik oleh kombinasi ChIP dengan NGS (*Next Generation Sequencing*).
- **CLIP-Seq/HITS-CLIP**
CLIP-Seq/HITS-CLIP merupakan kombinasi UV *crosslinking* dan imunopresipitasi dengan *high-throughput sequencing* untuk mengidentifikasi *binding sites* dari RNA-binding proteins, seperti: LIN28A, Argonaute (Ago), and Puf5p (Wang *et al.*, 2007; Leung *et al.*, 2011; Cho *et al.*, 2012).

Teknik OMICs tersebut umumnya diikuti oleh analisis data proteomik. Analisis tersebut dapat menggunakan beberapa basis data sebagai sumber analisis *pathway/network* tertentu dari sebuah organisme sebagaimana ditampilkan pada Tabel 1 (Wu *et al.*, 2014).

Tabel 1: Informasi sumber basis data untuk jalur metabolisme dari organisme hidup.

Table 1. Information of database sources for metabolic pathways of living organisms.

No	Nama basis data/ Database name	Tautan internet/ Web link
1	Universal Protein Resource	http://www.uniprot.org/
2	Swiss_Model	http://swissmodel.expasy.org//SWISS-MODEL.html
3	National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
4	Protein Data Bank	http://www.rcsb.org/pdb/
5	TarFisDock web server	http://www.dddc.ac.cn/tarfisdock/
6	Virtual Computational Chemistry Laboratory	http://www.virtuallaboratory.org/
7	DrugBank	http://redpoll.pharmacy.ualberta.ca/drugbank/
8	PDBbind Database	http://www.pdbbind.org/
9	The Binding Database	http://www.bindingdb.org/bind/index.jsp
10	ZINC (non-komersial)	http://blaster.docking.org/zinc/

3.6. Komputasi basis data lncRNA berbasis internet

Long non-coding RNAs atau yang dikenal sebagai *lncRNAs* merupakan RNA berukuran panjang (100-200 pb) yang tidak menyandi protein namun diketahui memiliki peran dalam mekanisme epigenetik, ekspresi gen dan ketahanan terhadap penyakit (Boerner & McGinnis, 2012; Rinn & Chang, 2012; Cheetham *et al.*, 2013; Zhang & Zhu 2014). Studi dan pengembangan basis data *lncRNAs* menjadi sangat penting dalam desain obat secara *in silico* maupun penelitian lain pada bidang pertanian seperti studi regulasi ekspresi gen dan marka penyakit (Amaral *et al.*, 2011). Sejumlah *lncRNAs* disimpan dalam basis data yang bervariasi dari 2000 hingga lebih dari 70.000 transkrip (Fritah *et al.*, 2014). Sebagai contoh, basis data terbesar yaitu NONCODE (<http://www.noncode.org/>) menyimpan lebih dari 73.000 transkrip dari berbagai spesies seperti manusia, tikus, sapi, ayam, lalat buah, kuda laut, dan khamir (Bu *et al.*, 2011; Zhao *et al.*, 2016). Meskipun demikian, tidak semua basis data menyediakan informasi yang memadai terkait jumlah total dan asal mula transkrip. Basis data lain seperti DIANA-LncBase (www.microrna.gr/LncBase) memuat sebanyak 2.958 transkrip *lncRNAs* yang telah diverifikasi secara eksperimental (Paraskevopoulou *et al.*, 2016).

Basis data *lncRNAs* yang diketahui hingga kini ditampilkan pada Tabel 2. Basis data lncRNABase merupakan kumpulan dari 108 data CLIP-Seq terutama pada mamalia dan cacing (Li *et al.*, 2014). Basis data ChiPBase merupakan anotasi dari profil ekspresi *lncRNAs* dan *ncRNAs* dari 10.200 ChiP-seq data (Yang *et al.*, 2015). LNCipedia merupakan basis data anotasi dari 21.448 transkrip *lncRNAs* pada manusia (Volders *et al.*, 2013). Di sisi lain, lncRNAdb merangkum secara komprehensif anotasi dari *lncRNAs* dari berbagai spesies seperti manusia hingga *Medicago* (Amaral *et al.*, 2011). Sementara

itu, lncRNAWiki merupakan kumpulan data publik pada manusia yang menggunakan sistem terbuka ala Wiki. User dapat memperbaharui informasi yang terdapat di dalamnya (Ma *et al.*, 2015). Kedua basis data NRED dan C-It-Loci merupakan basis data spesifik. NRED merupakan basis data yang merangkum ekspresi dari *lncRNAs* khusus pada mamalia (Dinger *et al.*, 2009). Sementara, C-It-Loci mengumpulkan data ekspresi RNA dan *lncRNAs* spesifik jaringan pada berbagai spesies (Weirick *et al.*, 2015).

4. Pemerataan sumber daya komputasi secara daring

Saat ini dan di masa depan, penelitian bioinformatika dapat dilakukan oleh siapapun yang memiliki akses internet, tanpa harus membangun sumber daya komputasi yang besar. Sebuah laboratorium bioinformatika tidak mutlak diperlukan untuk komputasi yang sederhana dengan hanya sebuah komputer standar. Secara prinsip, penelitian dasar bioinformatika dapat dilakukan secara mudah selama tersedia akses internet. Kolaborasi dengan peneliti lain dapat dilakukan secara daring dan dipermudah dengan bantuan *tools* media sosial. Optimalisasi penelitian *via* daring bukan berarti menghilangkan tatap muka atau kegiatan luring. Tatap muka tetap diperlukan untuk mengklarifikasi hal-hal yang tidak dipahami selama penelitian berjalan.

Penelitian bioinformatika secara daring biasanya dibatasi oleh besarnya basis data (jumlah sekuen DNA dan residu asam amino), namun kreativitas seorang peneliti dapat mengatasi pembatasan tersebut dengan ide-ide yang memiliki tingkat kebaruan yang tinggi. Pada masa sekarang dimana basis data sekuen terus bertambah, pengalaman dalam menentukan jenis sekuen yang digunakan dalam analisis mutlak diperlukan oleh seorang peneliti.

Tabel 2. Basis data *lncRNAs* umum dan spesifik dari berbagai organisme hidup.Table 2. General and specific *lncRNAs* databases from different living organisms.

No	Nama basis data/ Database name	Deskripsi/ Description	Tautan internet/ Web link
1	lncRNABase	Basis data miRNA-lncRNA dari 108 data set CLIP-Seq	http://starbase.sysu.edu.cn/mirLncRNA.php
2	ChIPBase	Basis data anotasi dari profil ekspresi lncRNAs dan ncRNAs	http://rna.sysu.edu.cn/chipbase/
3	LNCipedia	Basis data anotasi transkrip lncRNAs dari manusia	https://Lncipedia.org/
4	lncRNAdb	Basis data anotasi komprehensif lncRNAs	http://lncrnadb.com/
5	LncRNAWiki	Basis data publik lncRNAs menggunakan sistem terbuka Wiki	http://lncrna.big.ac.cn/
6	NONCODE	Basis data integratif untuk anotasi lncRNAs	http://www.noncode.org/
7	NRED	Basis data ekspresi dari lncRNAs untuk mamalia	http://nred.matticklab.com/cgi-bin/ncrnadb.pl
8	DIANA-LncBase	Basis data lncRNAs berbasis eksperimen	www.microna.gr/LncBase
9	C-It-Loci	Basis data ekspresi RNA dan lncRNAs lintas spesies spesifik jaringan	http://c-it-loci.uni-frankfurt.de/

Tanpa latar belakang biologis yang memadai, seorang peneliti dapat salah menerjemahkan makna biologis dari sekuen genom yang diperoleh dari basis data. Di sisi lain, teknik analisis daring memerlukan proses verifikasi berulang dari peneliti yang menggunakan data sekuen tersebut. Proses verifikasi tersebut dilakukan dalam rangka memastikan kebenaran dari sekuen yang dianalisis.

Saat ini, Pusat Informatika LIPI membuka jasa penggunaan komputer kluster (GRID LIPI, <http://grid.lipi.go.id/>) bagi peneliti di Indonesia (Akbar & Handoko, 2008). Layanan server publik dari luar negeri yaitu Galaxy (<https://galaxyproject.org>) juga dapat diakses secara bebas (Blankenberg *et al.*, 2010). GRID LIPI maupun Galaxy menyediakan *tools* bioinformatika *pre-installed* pada server masing-masing, sehingga analisis dapat langsung dilakukan. Namun demikian, GRID LIPI menggunakan *command line interface* (CLI) LINUX, sementara Galaxy mengutamakan *graphical line interface* (GUI) yang berbasis grafis. Preferensi menggunakan CLI dan GUI sangat tergantung pada penguasaan teknik komputasi. Peneliti pemula pada umumnya akan memilih berbasis GUI, sedangkan yang sudah mahir akan memilih CLI.

5. Perspektif

Perkembangan teknologi yang begitu pesat mendorong kinerja penelitian menjadi lebih cepat pula. Kecepatan dan ketepatan dalam penelitian biologi molekuler dapat didukung oleh keberadaan

sistem basis data bioinformatika secara daring. Keberadaan beberapa basis data sekuen *existing* belum sepenuhnya dimanfaatkan oleh para peneliti untuk mendukung penelitian mereka di tingkat pendahuluan atau pemula dalam bidang pertanian dan kesehatan. Meskipun demikian, tren penelitian berbasis bioinformatika diperkirakan akan meningkat di Indonesia. Seperti telah dijelaskan dalam kajian ilmiah ini, penelitian berbasis bioinformatika tergolong “mudah” dan mampu menghasilkan data prediktif yang mendekati kondisi nyata di laboratorium. Meskipun demikian, peneliti perlu memiliki keahlian khusus untuk melakukan validasi sekuen untuk memberikan makna biologis dari sekuen tersebut.

Dalam studi kasus desain obat secara *in silico*, berbagai pendekatan bioinformatika memanfaatkan basis data *existing* juga telah dibahas. Dalam bidang pertanian, bioinformatika menjadi bagian yang tidak terpisahkan dalam penelitian terkait, seperti analisis sekuen genom yang mengarah kepada proteksi dan pemuliaan tanaman. Basis data tersebut akan terus berkembang di masa yang akan datang dan tentunya seharusnya dapat dimanfaatkan oleh para peneliti dalam penelitian terkait.

Ucapan Terimakasih

Terima kasih kepada Direktur RIN Victor de Vries dan Ketua LPPM I3L Iwan Surjawan atas dukungannya terhadap penulisan artikel ini. Terima kasih juga diucapkan kepada I Wayan Aditya Swardiana dari Pusat Informatika LIPI atas diskusi mengenai komputasi kluster.

Daftar Pustaka

- Akbar Z & LT Handoko (2008). GRID architecture through a public cluster. In *Proc: International Conference on Computer and Communication Engineering 2008, ICCCE08*. Kuala Lumpur, Malaysia. 13-14 Mei, 2008 p, 1016–1018.
- Amaral PP, MB Clark, DK Gascoigne, ME Dinger, & JS Mattick (2011). LncRNAdb: A reference database for long noncoding RNAs. *Nucleic Acids Res* 39, D146-D151.
- Anighoro A, J Bajorath & G Rastelli (2014). Polypharmacology: Challenges and opportunities in drug discovery. *J Med Chem* 57(19), 7874–7887.
- Baharvand H, M Hajheidari, SK Ashtiani & GH Salekdeh (2006). Proteomic signature of human embryonic stem cells. *Proteomics* 6(12), 3544–3549.
- Bakri R, AA Parikesit, CP Satriyanto, D Kerami & USF Tambunan (2014). Utilization of boron compounds for the modification of suberoyl anilide hydroxamic acid as inhibitor of histone deacetylase class II *Homo sapiens*. *Adv Bioinformatics* 3, 1–10.
- Bian Q & P Cahan (2016). Computational tools for stem cell biology. *Trends Biotechnol* 34(12), 993-1009.
- Blankenberg D, G Kuster, N Von, Coraor, G Ananda, R Lazarus, M Mangan, A Nekrutenko & J Taylor (2010). Galaxy: A web-based genome analysis tool for experimentalists. SUPPL.89. *Curr Protoc Mol Biol* 19(89), 1-21.
- Boerner S & KM McGinnis (2012). Computational identification and functional predictions of long noncoding RNA in *Zea mays*. *PLoS One* 7(8), 1-16.
- Bourne PE, JR Lorsch & ED Green (2015). Perspective: Sustaining the big-data ecosystem. *Nature* 527(7576) S16-7.
- Budiani A, S Woelan, H Miharsih, H Nuhaimi & RA Putranto (2016). Evaluasi 18 primer SSR untuk pengembangan sidikjari DNA tanaman karet (*Hevea brasiliensis* Muell. Arg.) *Menara Perkebunan* 82(2), 81-94.
- Bu D, K Yu, S Sun, C Xie, G Skogerbø, R Miao, H Xiao, Q Liao, H Luo, G Zhao, H Zhao, Z Liu, C Liu, R Chen & Y Zhao (2012). NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Research*, 40, D210-D215.
- Cheetham SW, F Gruhl, JS Mattick & ME Dinger (2013). Long noncoding RNAs and the genetics of cancer. *Br J Cancer* 108(12), 2419–25.
- Cho J, H Chang, SC Kwon, B Kim, Y Kim, J Choe, M Ha, YK Kim & VN Kim (2012). LIN28A is a suppressor of ER-associated translation in embryonic stem cells. *Cell* 151(4), 765–777.
- COE EH (1998). Potentials of the National Corn Genome Initiative *PNAS* 95(5), 2029–2032.
- Conesa A, S Götz, JM García-Gómez, J Terol & M Talon (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18), 3674–3676.
- Daina A, O Michielin & V Zoete (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* 7, 42717.
- Darvas F, G Dormán & A Papp (2000). Diversity measures for enhancing ADME admissibility of combinatorial libraries. *J Chem Inf Comput Sci* 40(2), 314–322.
- de Setta N, CB Monteiro-Vitorello & CJ Metcalfe (2014). Building the sugarcane genome for biotechnology and identifying evolutionary trends. *BMC Genomics* 15(1), 540-557.
- Denoeud F, L Carretero-Paulet, A Dereeper, G Droc, R Guyot & M Pietrella (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis *Science* 345(6201), 1181–1184.
- Dinger ME, KC Pang, TR Mercer, ML Crowe, SM Grimmond & JS Mattick (2009). NRED: a database of long noncoding RNA expression. *Nucleic Acids Research* 37, D122-D126.
- Dou L, X Zhang, C Pang, M Song, H Wei, S Fan & S Yun (2014). Genome-wide analysis of the WRKY gene family in cotton. *Mol Genet Genomics* 289(6), 1103–1121.
- Duan C, X Argout, V Gébelin, A Champion & P Montoro (2013). Identification of the *Hevea brasiliensis* AP2/ERF superfamily by RNA sequencing. *BMC Genomics* 14(1), 30-42.
- Eisenhaber F, C-K Kwoh, S-K Ng, M Summo, & Dufayard JF (2009). Brief Overview of Bioinformatics Activities in Singapore P. E. Bourne, ed. *PLoS Comput Biol* 5(9), 1-4.
- Fritah S, SP Niclou & F Azuaje (2014). Databases for lncRNAs: a comparative evaluation of emerging tools. *RNA* 20(11), 1655–1665.
- Gashaw I, P Ellinghaus, A Sommer & K Asadullah (2011). What makes a good drug target? *Drug Discovery Today* 16, 1037-1043.

- Gatehouse JA (2011). Prospects for using proteinase inhibitors to protect transgenic plants against attack by herbivorous insects. *Curr Protein Pept Sci* 12(5), 409–416
- Goff SA, D Ricke, T-H Lan, G Presting, M Dunn, & C Rice Genome (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* 296(5565), 92–100.
- Goodman AL & A Dekhtyar (2014). Teaching bioinformatics in concert J A Fox, ed *PLoS Comput Biol* 10(11), e1003896.
- Haas BJ, S Kamoun, MC Zody, RHY Jiang, S Raffaele, & A Avrova (2009). Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. *Nature* 461(7262), 393–398.
- Han SJ, T Song, Y-J Cho, JS Kim & SY Choi (2015). Complete genome sequence of *Mycobacterium tuberculosis* K from a Korean high school outbreak, belonging to the Beijing family. *Stand Genomic Sci* 10(1), 78-89.
- Hospital A, P Andrio, C Fenollosa, D Cicin-Sain, & JL Gelpi (2012). MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. *Bioinformatics* 28(9), 1278–1279.
- Irwin JJ, BK Shoichet, MM Mysinger, N Huang, Y Cao, & P Wassam (2009). Automated docking screens: a feasibility study. *J Med Chem* 52(18), 5712–5720.
- Miguel JJ, A Pereira, P Bartholomew & GP Lasala (2011). The intrathecal infusion of mesenchymal stem cells into healthy rabbits is safe and devoid of neurological or clinical complications. *J Stem Cell Res Ther* 1(2), 1–5.
- Kapetanovic IM (2008). Computer-Aided Drug Discovery and Development (CADD): in silico-chemico-biological approach. *Chemico-biological interactions* 171, 165-176.
- Kearse M, R Moir, A Wilson, M Cheung & B Ashton (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12), 1647–1649.
- Knox C, V Law, T Jewison, P Liu, S Ly & C Mak (2011). DrugBank 3.0: a comprehensive resource for “omics” research on drugs. *Nucleic Acids Res* 39, 1035–1041.
- Lagorce D, O Sperandio, H Galons, MA Miteva, & BO Villoutreix (2008). FAF-Drugs2: free ADME/tox filtering tool to assist drug discovery and chemical biology projects. *BMC Bioinformatics* 9, 396-405.
- Lau N, Y Makita, M Kawashima, S Kondo & M Matsui (2016). The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Nat Publ Gr* 6(1),1–14.
- Lawrence PK & KR Koundal (2002). Plant protease inhibitors in control of phytophagous insects. *Electron J Biotechnol* 5(1), 93–109.
- Lee SK, IH Lee, HJ Kim, GS Chang, JE Chung & KT No (2003). The PreADME Approach: Web-based program for rapid prediction of physico-chemical, drug absorption and drug-like properties. *EuroQSAR 2002 Des Drugs Crop Prot Process Probl Solut* 2003, 418–420.
- Leung AK, AG Young, A Bhutkar, GX Zheng, AD Bosson & PA Sharp (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* 18(2), 237–244.
- Li J-H, S Liu, H Zhou, L-H Qu & J-H Yang (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, 42, D92-D97.
- Lim SP, Q-Y Wang, CG Noble, S Nilar & P Smith (2013). Ten years of dengue drug discovery: progress and prospects. *Antiviral Res* 100(2),500-519
- Lipinski CA, F Lombardo, BW Dominy & PJ Feeney (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 64(SUPPL) 4–17.
- Liu H, Z Deng, J Chen, L Hao & D Li (2016). Genome-wide identification and expression analysis of the metacaspase gene family in *Hevea brasiliensis*. *Plant Physiol Biochem PPB* 105, 90–101.
- Loging W, R Rodriguez-Esteban, J Hill, T Freeman & J Miglietta (2011). Cheminformatic /bioinformatic analysis of large corporate databases: Application to drug repurposing. *Drug Discovery Today: Therapeutic Strategies* 8, 109-116.
- Ma L, A Li, D Zou, X Xu, L Xia, J Yu, VB Bajic & Z Zhang (2015). LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Research* 43, 187-192.
- Macaulay IC & T Voet (2014). Single cell genomics: Advances and future perspectives. *PLoS Genet* 10(1), 1-9.
- Mahmood I (2007). Application of allometric

- principles for the prediction of pharmacokinetics in human and veterinary drug development. *Adv Drug Deliv Rev* 59(11), 1177–1192.
- Mokrousov I, E Chernyaeva, A Vyazovaya, V Sinkov & ON Narvskaya (2016). Next-generation sequencing of *Mycobacterium tuberculosis*. *Emerg Infect Dis* 22(6), 1127–1129.
- NCBI (2013). *The NCBI handbook*. Diunduh dari: <http://www.ncbi.nlm.nih.gov/books/NBK21101/> [1 Agustus, 2013].
- Paraskevopoulou MD, IS Vlachos, D Karagkouni, G Georgakilas, I Kanellos, T Vergoulis, K Zagganas, P Tsanakas, E Floros, T Dalamagas & AG Hatzigeorgiou (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Research*, 44, D231–D238.
- Parikesit A (2010). The role of bioinformatics in protein-protein interaction study. *Sigma J Sci Technol* 13(1), 1–7.
- Parikesit A, K Kinanty & USF Tambunan (2013). Screening of commercial cyclic peptides as inhibitor Envelope Protein Dengue Virus (DENV) through molecular docking and molecular dynamics. *Pakistan J Biol Sci* 16(24), 1836–1848.
- Parikesit AA (2009). The role of bioinformatics as auxiliary tools for molecular biology. In *Proc: World-Wide Indonesian Student Association Scientific Writing Olympic*. Paris, 10-11 Oktober, 2009 p, 23–29.
- Parikesit AA, B Ardiansah, DM Handayani, USF Tambunan, D Kerami (2016). Virtual screening of Indonesian flavonoid as neuraminidase inhibitor of influenza a subtype H5N1. *IOP Conf Ser Mater Sci Eng* 107(1), 12053-12062.
- Parikesit AA, L Steiner, PF Stadler & SJ Prohaska (2014). Pitfalls of ascertainment biases in genome annotations—computing comparable protein domain distributions in eukarya. *Malaysian J Fundam Appl Sci* 10(2), 65–75.
- Prabahar A, S Swaminathan, A Loganathan & R Jegadeesan (2015). Identification of novel inhibitors for tobacco mosaic virus infection in Solanaceae plants. *Adv Bioinformatics* 8(1), 1–9.
- Priyono P & RA Putranto (2014). Molecular markers and their application for DNA fingerprinting and genetic diversity studies in *Coffea* species. *Menara Perkebunan* 82(1), 39–50.
- Priyono P & RA Putranto (2016). Studi peta genetik pada *Coffea sp.* menggunakan metode penanda molekuler. *Menara Perkebunan* 83(2), 95-104.
- Putranto R-A, C Duan, Kuswanhadi, J Pirello & F Dessailly (2015). Ethylene response factors are controlled by multiple harvesting stresses in *Hevea brasiliensis*. *PLoS One* 10(4),1-26.
- Reegan AD, A Stalin, & MG Paulraj (2016). In silico molecular docking of niloticin with acetylcholinesterase 1 (AChE1) of *Aedes aegypti* L. (Diptera: Culicidae): a promising molecular target. *Med Chem Res* 25(7), 1411–1419.
- Rinn JL & HY Chang (2012). Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81, 145–66.
- Saews Y, I Inza & P Larrañaga (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* 23(19), 2507–2517.
- Santamaría M, P Hernández-Crespo, F Ortego, I Diaz & M Martinez (2012). Cysteine peptidases and their inhibitors in *Tetranychus urticae*: a comparative genomic approach. *BMC Genomics* 13(1), 307-320.
- Sato S, S Tabata & H Hirakawa (2012). The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485(7400), 635–641.
- Schmutz J, SB Cannon, J Schlueter & SG Consortium (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463(7278), 178–83.
- Searls DB (2012). An online bioinformatics curriculum. *PLoS Comput Biol* 8(9), 1-20.
- Siavelis JC, MM Bourdakou, EI Athanasiadis, GM Spyrou & KS Nikita (2016). Bioinformatics methods in drug repurposing for Alzheimer’s disease. *Briefings in Bioinformatics* 17, 322-335.
- Singh R, M Ong-Abdullah, E-TL Low, RS Fulton & JM Ordway (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* 500(7462), 335–339.
- Skvortsov TA, DV Ignatov, & KB Majorov (2013). *Mycobacterium tuberculosis* Transcriptome Profiling in Mice with Genetically Different Susceptibility to Tuberculosis. *Acta Naturae* 5(2), 62–69.
- Sukhov V, V Nerush, L Orlova & V Vodeneev (2011). Simulation of action potential propagation in plants. *J Theor Biol* 291(1), 47–55.
- Tambunan USF & AA Parikesit (2010). Cracking the genetic code of human virus by using open

- source bioinformatics tools. *Malaysian J Fundam Appl Sci* 6(1), 42-50.
- Tambunan USF, AA Parikesit, D Sugiono & TA Tochary (2007). Computational study of post translation modification in chimeric virus like particles vaccine of human papilloma virus with virion capsid L1. *Makara Seri Sains* 11(2), 56-62.
- Tang C, M Yang, Y Fang & RTG Consortium (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nat Plants* 2(6), 16073-16083.
- Tang Y, W Zhu, K Chen & H Jiang (2006). New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. *Drug Discov Today Technol* 3(3), 307-313.
- Teh C-K, A-L Ong, Q-B Kwong, M Mohammed & D Appleton (2016). Genome-wide association study identifies three key loci for high mesocarp oil content in perennial crop oil palm. *Sci Rep* 6, 19075-19082.
- Terstappen GC & A Reggiani (2001). In silico research in drug discovery. *Trends Pharmacol Sci* 22(1), 23-26.
- Tessema B, P Nabeta, E Valli, E Vailli & NH Lan (2017). FIND tuberculosis strain bank: a resource for researchers and developers working on tests to detect mycobacterium tuberculosis and related drug resistance. *J Clin Microbiol* 55(4), 1066-1073.
- Ting N-C, J Jansen, J Nagappan, SG Tan & R Sigh (2013). Identification of QTLs associated with callogenesis and embryogenesis in oil palm using genetic linkage maps improved with SSR markers. *PLoS One* 8(1), 1-16.
- Vogel JP, DF Garvin, TC Mockler, K Mayer, J Zhai, P Cao, Y Cui & G Consortium (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463(7282), 763-768.
- Volders P-J, K Helsens, X Wang, B Menten, L Martens, K Gevaert, J Vandesompele & P Mestdagh (2013). LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Research*, 41, D246-D251.
- Wang M, K Lamberth, M Harndahl, S Buus & O Lund (2007). CTL epitopes for influenza A including the H5N1 bird flu; genome-, pathogen- and HLA-wide screening. *Vaccine* 25(15), 2823-2831.
- van de Waterbeemd H & E Gifford (2003). ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2(3), 192-204.
- Wei D-Q, Q-S Du, H Sun & KC Chou (2006). Insights from modeling the 3D structure of H5N1 influenza virus neuraminidase and its binding interactions with ligands. *Biochem Biophys Res Commun* 344(3), 1048-1055.
- Weirick T, D John, S Dimmeler & S Uchida (2015). C-It-Loci: a knowledge database for tissue-enriched loci. *Bioinformatics*, 31, 3537-3543.
- Whitfield EJ, M Pruess & R Apweiler (2006). Bioinformatics database infrastructure for biotechnology research. *J Biotechnol* 124(4), 629-639.
- Williams G & GR Mirams (2015). A web portal for in-silico action potential predictions. *J Pharmacol Toxicol Methods* 75, 10-16.
- Wu X, M Al-Hasan & JY Chen (2014). Pathway and network analysis in proteomics. *J Theor Biol*, 362, 44-52.
- Xu X, S Pan, S Cheng, O Ponce, Y Yang, J Gawor, CP Consortium (2011). Genome sequence and analysis of the tuber crop potato. *Nature* 475(7355), 189-195.
- Yang J-H, J-H Li, S Jiang, H Zhou & L-H Qu (2013). ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Research*, 41, D177-D187.
- Zhang H & J-K Zhu (2014). Emerging roles of RNA processing factors in regulating long non-coding RNAs. *RNA Biol* 11(7), 793-797.
- Zhao Y, H Li, S Fang, Y Kang, W Wu, Y Hao, Z Li, D Bu, N Sun, MQ Zhang & R Chen (2016). NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Research* 44, D203-D208.
- Zou Z, J Gong, F An, G Xie, J Wang, Y Mo & L Yang (2015). Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. *BMC Genomics* 16(1), 1001-1020.