

BIOINFORMATIKA: TREND DAN PROSPEK DALAM PENGEMBANGAN KEILMUAN BIOLOGI¹

Arief B. Witarto² dan Sajidan³

PENDAHULUAN

Tulisan ini terutama dimaksudkan untuk pengguna awal bioinformatika, termasuk mereka yang selama ini belum mengenal bioinformatika dan tertarik untuk memulainya⁴. Tapi bagi yang selama ini sudah menggeluti bioinformatika secara otodidak melalui internet terutama, cenderung mengalami kesalahan bila tidak memiliki dasar-dasar yang kuat terhadap biologi molekuler atau menemui kebuntuan untuk melihat permasalahan biologi yang bisa dipecahkan dengan bioinformatika. Maka dari itu dengan penjelasan agak mendalam tentang prinsip biologi molekuler termasuk bagaimana data biologi molekuler itu didapatkan, diharapkan pembaca makalah ini bisa lebih optimal menggunakan bioinformatika khususnya menunjang pengembangan keilmuan biologi di Tanah Air⁵.

BIOINFORMATIKA VS BIOLOGI KOMPUTASI

“Bioinformatika, apakah itu?”. Pertanyaan ini perlu dijelaskan di awal untuk memberikan pengertian yang benar mengenainya. Secara definisi, bioinformatika adalah ilmu yang mengaplikasikan teknologi informatika yaitu teknologi komputasi dan komunikasi terutama internet pada ilmu biologi. Kalau demikian apa bedanya dengan biologi komputasi yang lebih dulu dikenal dan juga menggunakan komputer untuk menganalisa fenome-fenomena biologi? Contoh klasik biologi komputasi yang lazim dilakukan di Indonesia misalnya adalah analisis pertumbuhan bakteri. Bakteri penghasil enzim chitosanase dikultur dalam media dengan berbagai komposisi yang berbeda untuk mencari kondisi terbaik produksi enzim itu⁶. Dari data eksperimen pembiakkan bakteri itu, seluruh parameternya kemudian dianalisis dengan komputer untuk mendapatkan kurva-kurva pertumbuhan bakteri sehingga dapat dipilih komposisi medium terbaik.

¹ Disampaikan pada Seminar yang diselenggarakan oleh Fakultas Keguruan dan Ilmu Pendidikan, UNS, Solo, Sabtu, 31 Juli 2010

² Pendiri Himpunan Bioinformatika Indonesia. Peneliti Rekayasa Protein. Menyelesaikan pendidikan S-1, S-2, S-3 di Departemen Bioteknologi, Fakultas Teknik, Tokyo University of Agriculture and Technology, Tokyo, Jepang. Tinggal di Depok. Alamat kontak, witarto@yahoo.com (email/facebook), witarto.wordpress.com (blog)

³ Dosen Pendidikan Biologi FKIP UNS

⁴ Witarto, A.B. Bioinformatika: Mengawinkan teknologi informasi dengan bioteknologi. Trendnya di dunia dan prospeknya di Indonesia. Disampaikan pada Seminar Seminar Teknologi Informasi diselenggarakan oleh MIFTA, Bogor, 9 Januari 2003. Bisa diunduh dari witarto.wordpress.com.

⁵ Witarto, A.B. Bioinformatics in Indonesia. Disampaikan pada First ASEAN-India Workshop on Bioinformatics di Center for DNA Fingerprinting and Diagnostics, Hyderabad, India, 7-11 November 2005. Bisa diunduh dari witarto.wordpress.com

⁶ Chasanah, E., Suhartono, M.T., Witarto, A.B., Hwang, J.K., and Pyun, Y.R. Effect of Growth Substrate on Production of Thermostable Chitosanase from MB-2 Isolate. *Proceedings of International Seminar on Marine and Fisheries* (Organized by Agency for Marine and Fisheries Research-Indonesian Ministry of Marine Affairs and Fishery, in Jakarta, 15-16 December),



Penelitian ini dilakukan oleh mahasiswa S-3 di IPB yang dibimbing oleh penulis. Jauh sebelum penelitian ini, di tingkat dunia biologi komputasi telah digunakan pada ilmu biologi struktur untuk menjelaskan struktur protein. Dari buku *Introduction to Protein Structure* yang ditulis oleh Carl Branden dan John Tooze yang menjadi buku teks wajib untuk mempelajari ilmu biologi struktur, kita bisa melihat data eksperimen yang melimpah berupa titik-titik hitam difraksi sinar X yang informasinya kemudian diolah oleh komputer menjadi koordinat atom penyusun protein yang kristalnya ditabrak sinar X sehingga menghasilkan pola difraksi tadi. Eksperimen ini pertama kali dilakukan oleh Max Perutz dan mahasiswanya John Kendrew pada tahun 1957. Pada tahun 1962, Prof. G. Hagg dari Panitia Hadiah Nobel dalam penyerahan Hadiah Nobel Kimia kepada Perutz dan Kendrew menjelaskan, “untuk menjelaskan struktur protein myoglobin yang terdiri dari 2.600 atom, diperlukan 110 buah kristal untuk mendapatkan 250.000 pola difraksi sinar X. Perhitungan (koordinat atom) tidak akan mungkin terjadi bila tidak menggunakan komputer yang besar”⁷.

Dari dua contoh kisah di Indonesia akhir-akhir ini dan di dunia yang mengawali sejarah biologi komputasi, kita bisa memahami bahwa biologi komputasi telah digunakan sejak lama untuk mengurai data-data eksperimen biologi yang besar. Dalam perkembangan ilmu, kita mengetahui bahwa pada tahun 1990-an muncullah teknologi informasi dan komunikasi (TIK) seperti lahirnya internet, email, browser, gopher, ftp, dan sebagainya. Pada era yang hampir sama, di bidang ilmu biologi juga memasuki era revolusi “om” seperti genom dan proteom sehingga terjadilah ledakan informasi biologi. Revolusi “om” pada ilmu biologi antara lain dipicu oleh munculnya teknologi DNA rekombinan atau yang sering disebut juga bioteknologi modern pada tahun 1970-an akibat kemajuan ilmu biologi molekuler. Ledakan informasi biologi molekuler dapat dilihat dari pertumbuhan data di pusat data DNA, GenBank. Pertumbuhan eksponensial terjadi sejak tahun 1998 ketika untuk pertama kalinya genom bakteri berhasil dibaca. Pada tahun yang sama, pertumbuhan data koordinat protein dan makromolekul lainnya di pusat data Protein Data Bank juga menunjukkan pertumbuhan yang eksponensial.

PRINSIP BIOLOGI MOLEKULER

Sebelum lebih jauh membahas bentuk data yang melonjak drastis itu, kita perlu memahami posisi biologi molekuler dalam ilmu biologi secara keseluruhan. Hirarki dalam biologi dimulai dari ekosistem dengan contoh lingkungan di muka bumi. Dalam ekosistem terdapat populasi, misalnya populasi manusia Indonesia. Kemudian terdapat individu, contohnya satu individu manusia, penulis sendiri di Depok. Individu manusia ini terbentuk dari bagian-bagian yang disebut organ seperti paru-paru dan otak. Bila organ adalah kumpulan beberapa jenis sel yang membentuk satu fungsi, maka bentuk yang lebih sederhana dalam struktur maupun jumlah selnya, disebut jaringan seperti jaringan kulit dan jaringan tulang rawan. Seperti sudah disebutkan, organ atau jaringan tadi dibentuk oleh sel sebagai unit terkecil kehidupan. Bila “dibedah” lebih dalam, sel tadi terbentuk dari organela seperti inti sel, mitokondria dan sitoplasma. Dan organela-organela itu

⁷ Bisa diunduh dari www.nobel.se pada bagian Hadiah Nobel Kimia 1962.



menjadi tempat beraktivitas 4 jenis makromolekul utama dalam sel yaitu DNA/RNA, protein, karbohidrat dan lipid. Hirarki terbawah dalam biologi adalah molekul-molekul berukuran kecil seperti air, vitamin, oksigen dan lain-lain. Bioinformatika yang berkembang saat ini terutama “bermain” pada hirarki makromolekul. Komputer tercepat di dunia atau superkomputer yang dibuat IBM yaitu BlueGene misalnya, digunakan untuk memprediksi pelipatan protein⁸. Perakitan potongan-potongan 1.000-an pasang basa (pb) sekuen DNA dari genom manusia, dirakit menjadi 23 rantai utuh genom manusia sepanjang 3.000.000.000 (3 milyar) menggunakan komputer yang disponsori Compaq di perusahaan pembaca sekuen genom Celera, Amerika Serikat. Karena kompleksitas informasi yang ada, kemampuan komputasi yang tersedia baru sampai pada hirarki molekul biologi. Selain itu informasi dari eksperimen biologi juga yang paling banyak tersedia baru sampai tingkat molekul biologi. Tapi sesuai hirarki dalam biologi itu, lambat laun bioinformatika akan menuju ke hirarki yang lebih tinggi yaitu organela, sel, jaringan/organ dan individu.

Seperti kita membuat slide presentasi atau menulis artikel, data komputer berbentuk teks atau gambar. Dari mana datangnya data bioinformatika yang dianalisis oleh komputer? Jawabnya adalah dari eksperimen “basah” – disebut “basah” untuk membandingkan dengan eksperimen “kering” oleh komputer yang tidak menghendaki adanya air karena merusakkan komponen listriknya – seperti pembacaan sekuen DNA, kristalografi sinar X dan DNA microarray. Karena teknik-tekniknya berbeda maka karakter datanya pun berbeda. Sehingga sebelum kita menggunakan perangkat lunak dan internet untuk menerapkan bioinformatika, sangat penting untuk pertama memahami bagaimana fenomena biologi yang diteliti itu dan teknik bagaimana mendapatkan data bioinformatika itu dilakukan.

Dalam ranah biologi molekuler dikenal prinsip Dogma Sentral. Menurut prinsip ini, sumber informasi adalah urutan/sekuen DNA berupa rangkaian 4 jenis basa yaitu A (adenin), T (timin), G (guanin) dan C (sitosin). Dalam inti sel, informasi ini selanjutnya dirubah oleh enzim RNA polimerasi dalam sebuah proses yang disebut transkripsi menjadi urutan RNA. Pada RNA basa yang berubah hanya T menjadi U (urasil). Selebihnya sama persis, termasuk urutannya. RNA dari inti sel akan keluar menuju sitoplasma yang kemudian informasinya dirubah oleh Ribosom menjadi protein. Proses ini disebut translasi dimana setiap 3 pb RNA yang disebut kodon diterjemahkan menjadi 1 asam amino. Terakhir, rantai asam amino atau sering disebut rantai polipeptida ini melipat membuat satu struktur tertentu yang akhirnya menentukan fungsi dari protein tersebut apakah sebagai enzim, hormon, antibodi dan sebagainya. Sebagian besar DNA manusia berada dalam inti sel. Berbeda dengan DNA bakteri yang berbentuk lingkaran sehingga tidak ada ujungnya, DNA manusia berbentuk linier. Karena panjangnya itu, DNA linier ini dikemas oleh protein-protein tertentu seperti histone menjadi nukleosom, selanjutnya beberapa nukleosom membentuk kromatin sebagai unit aktif terkecil berukuran panjang 30 nm. Kromatin lebih jauh bergandengan membentuk struktur kompleks yang kita kenal sebagai kromosom. Jadi kromosom adalah pembawa fisik informasi sekuen DNA. Seandainya urutan basa DNA

⁸ Allen, F. et al. Blue Gene: A vision for protein science using a petaflop supercomputer. IBM Systems Journal, vol. 40, num. 2, 2001.



adalah bit dalam komputer, maka kromosom adalah hardisk-nya. Kita memiliki 46 buah kromosom. 44 di antaranya adalah pasangan yang masing-masing bersumber dari bapak dan ibu kita. Sementara 2 kromosom lainnya yang disebut X dan Y menentukan jenis kelamin kita apakah laki-laki (XY) atau perempuan (XX) dan tidak pernah ada pasangan YY. Di luar DNA yang terdapat dalam inti sel, manusia memiliki DNA di organela mitokondria. DNA mitokondria ini walau jumlahnya sangat sedikit tapi penting sekali dalam fungsi sel, sekaligus menyimpan “rahasia” asal-usul manusia yang akan dijelaskan nanti.

Dalam dogma sentral DNA digunakan untuk membuat protein. Sekuen DNA yang mengkode protein disebut gen. Pada mamalia rata-rata hanya 10 % dari keseluruhan DNA yang berupa gen. 90 % sisanya berguna untuk menjaga keutuhan kromosom seperti menjadi centromere (bagian tengah) atau telomere (bagian ujung), dan sebagian besar lainnya fungsinya belum diketahui sehingga muncul istilah junk DNA⁹. Junk bukan berarti sampah. Junk food tidak memiliki nilai gizi tapi disukai. Barang yang membuat sakit dan dibuang disebut sampah. Jadi junk DNA bukan parasit yang tidak berguna, hanya kita belum memahami keseluruhan fungsinya saja di luar prinsip dogma sentral. Oleh karena itu, konsep genom pada era bioinformatika ini pun telah berubah. Dahulu genom adalah satu set gen yang dimiliki individu untuk hidup. Sekarang genom adalah seluruh DNA, termasuk yang mengkode gen atau tidak. Genom berbagai makhluk hidup telah selesai dibaca. Selain manusia, juga tikus yang menjadi model penelitian biologi selama ini, juga padi yang menjadi makanan pokok bangsa Indonesia. Bila genom adalah obyek maka ilmu yang mempelajarinya disebut genomik. Selanjutnya mengikuti prinsip dogma sentral, gen yang dirubah menjadi RNA melahirkan obyek RNAom atau transkriptom dengan ilmunya masing-masing RNAomik dan Transkriptomik. Kemudian RNA dirubah menjadi protein, sehingga obyeknya adalah proteom dan ilmunya adalah proteomik. Akhiran-akhirian “om” ini berarti “keseluruhan” yang menunjukkan kemampuan mempelajari obyek itu secara keseluruhan pada satu sel atau individu. Inilah yang disebut revolusi “om”. Seperti penulis sajikan dalam tulisan di majalah Warta Ekonomi (4-4-2008), revolusi “om” tidak hanya berhenti pada meja lab tapi sudah merambah dunia bisnis, terutama kedokteran.

Fenomena biologi lain yang mendasar dan penting untuk memahami data bioinformatika adalah tentang fungsi gen. Walaupun manusia dewasa memiliki kira-kira 3 trilyun sel dengan 3 ribu jenis sel yang berbeda, tapi semuanya memiliki genom yang sama. Artinya DNA pada sel darah yang berwarna merah, sama persis dengan DNA pada sel kulit yang berwarna cokelat. Kenapa sel darah bisa berwarna merah? 33 % dari kandungan sel darah merah/eritrosit adalah protein hemoglobin. Protein ini mengandung zat besi yang melekat pada senyawa organik heme. Seperti kita melihat besi berkarat yang berwarna merah, zat besi ini berwarna merah. Oleh karena kandungannya yang tinggi, hal ini membuat sel darah berwarna merah. Fungsi zat besi itu sendiri adalah mengikat molekul udara

⁹ “I said it was “junk” DNA, not “trash”. Everyone knows that you throw away trash. But junk we keep in the attic until there are may be some need for it.”, di S. Brenner. The human genome: The nature of the enterprise (in: Human Genetic Information: Science, Law and Ethics – No.149; Science, Law, and Ethics – Symposium Proceedings (CIBA Foundation Symposia)), John Wiley and Sons, Ltd, 1990.



(O₂) sehingga sel darah merah dapat menghantarkan oksigen ke seluruh bagian tubuh. Apakah yang membuat gen penyandi protein hemoglobin hanya aktif di sel darah merah, tapi tidak di sel kulit, sel rambut dan lainnya? Gen tidak selalu aktif. Gen bisa dimatikan seperti kita mematikan lampu dengan saklar. Saklar dalam gen adalah promoter dan operator. Enzim RNA polimerase yang mentranskripsi DNA menjadi RNA berikatan dengan promoter dengan kekuatan sebesar 190 nano molar (nM)¹⁰. Di depan promoter terdapat operator yang menjadi tempat protein lain berikatan yaitu repressor. Repressor mengikat operator dengan kekuatan 20 piko molar (pM). Ketika repressor berikatan dengan promoter, ikatan RNA polimerase pada promoter meningkat 100 x menjadi 2,5 nM. Hal ini menyebabkan 2 hal. Pertama, RNA polimerase lebih memilih posisi stand-by seperti atlet lari yang bersiap di garis start daripada berkeliaran di sekelilingnya yang membuang energi dan waktu. Kedua, karena ikatan repressor (20 pM) lebih kuat 10.000 x daripada ikatan RNA polimerase (2,5 nM), maka gen menjadi tidak aktif karena RNA polimerase seperti kereta api yang tertahan oleh barang yang jauh lebih kuat di depannya. Seperti inilah posisi gen tidak aktif. Untuk mengaktifkan gen, sel memelurkan satu molekul kecil yang bernama inducer. Inducer ini berikatan dengan protein repressor yang menyebabkan ikatan repressor dengan operator turun 1.000 x menjadi hanya 0,2 nM. Dengan hanya perbedaan kekuatan ikatan 10 x (RNA polimerase 2,5 nM vs repressor 0,2 nM), maka RNA polimerase yang berukuran jauh lebih besar, dapat dengan mudah menyingkirkan rintangan di depannya yaitu repressor itu. Inilah saatnya gen menjadi aktif. Tapi dalam kehidupan nyata kita mengenal lampu berdaya 10 W yang redup dan 60 W yang terang. Demikian pula dalam gen ada yang tingkat ekspresinya rendah atau tinggi, tidak sekedar mati/hidup. Secara umum, faktor-faktor yang mengatur itu disebut protein-protein faktor transkripsi (TF). Misalnya di belakang promoter ada sekuen DNA berurutan TATA. Pada sekuen ini, berikatan protein faktor transkripsi II D (TFIID) yang kemudian “mengundang” protein lain seperti TFIIA, TFIIB dan seterusnya yang akhirnya membuat enzim RNA polimerase “nyaman” mendatangi “tahtanya” di sekuen promoter. Sekuen TATA yang kemudian disebut TATA box menjadi salah satu penunjuk lokasi gen yang hanya berjumlah 10 % itu pada “lautan” genom yang 90 %-nya tidak mengkode gen. Setelah transkripsi terjadi, pada sel-sel eukariot (sel berinti seperti pada manusia), sebelum memasuki tahap translasi masih terjadi proses lanjutan yaitu splicing yaitu penghilangan bagian gen yang tidak mengkode protein atau intron. Gen sel eukariot mengandung bagian yang disebut exon yaitu bagian yang akhirnya digunakan untuk mengkode protein dan intron. Selain splicing, juga ada proses penstabilan RNA dengan memberikan cap/topi yaitu sejenis gugus kimia di ujung-ujungnya yaitu gugus metil guanosin di ujung 5' dan gugus poli adenin di ujung 3'. Intron hanya dimiliki oleh sel eukariot. Bakteri yang berjenis sel tak berinti/prokariot tidak memilikinya. Tetapi DNA manusia di mitokondria uniknya juga tidak memiliki intron jadi mirip DNA bakteri. Hal ini yang menjadi landasan teori evolusi sel eukariot yang diramalkan berasal dari sel prokariot yang dimakan oleh sel lain tapi bertahan sehingga lambat laun sel ini menjadi sel eukariot. Kembali kepada sel eukariot, proses splicing/pemotongan intron bisa menghasilkan protein

¹⁰ R.C. Hardison. Working with Molecular Genetics. The Pennsylvania State University, 2005. Buku teks ini gratis dan digunakan penulis untuk mengajar kuliah genetika molekuler untuk S-2 di Program Studi Biologi UI, Depok.



yang berbeda dari gen yang sama yaitu dengan mengkombinasikan bagian yang berbeda. Protein yang dihasilkan biasanya disebut ARF/alternate reading frame. Misalnya perbedaan exon nomer 1 pada protein p16 (disebut exon 1 α) dan p14^{ARF} (disebut exon 1 β) membuat terjadinya dua protein yang berbeda ini walau exon lainnya setelah itu sama persis¹¹. Ukuran intron yang bisa jauh lebih besar daripada exon juga membuat variasi seperti ini menjadi sangat mungkin terjadi.

Setelah gen berhasil ditranskripsi menjadi RNA yang matang melalui proses splicing, berikutnya informasi ini dibaca oleh molekul Ribosom menjadi protein. Ribosom adalah molekul yang terbentuk dari protein dan RNA (ribosomal RNA/rRNA) dan awalnya terdiri dari dua bagian terpisah yaitu bagian 50S dan 30S. Hanya ketika Ribosom mengenali satu sekuen tertentu pada RNA yang disebut sekuen Shine-Dalgarno, maka dua bagian itu bersatu membentuk Ribosom utuh yang berbentuk seperti cincin dengan RNA di tengahnya. Ribosom akan menerjemahkan informasi RNA menjadi informasi asam amino yang drangkai. Rumus yang digunakan ribosom telah dipecahkan oleh para ilmuwan dan dirangkum dalam tabel yang disebut Tabel Kode Genetik. Dalam tabel ini, 3 basa RNA yang berurutan (disebut kodon) diterjemahkan menjadi satu asam amino. Karena ada 20 jenis asam amino sementara ada 64 kemungkinan kombinasi 3 basa RNA dari 4 jenis basa (A.,U,G,C) maka terdapat asam amino yang bisa dikode oleh lebih dari satu kodon. Keadaan ini yang disebut sifat degenerasi kodon. Walaupun seluruh individu di muka bumi menggunakan tabel yang sama, tapi terdapat perbedaan pada frekuensi penggunaan kodon yang dipakai untuk membuat asam amino. Misalnya antara tanaman dengan bakteri, dengan manusia dan seterusnya. Hal inilah yang disebut codon usage.

Rantai polipeptida yang merupakan untaian “manik-manik” asam amino yang disintesis oleh ribosom ini disebut dengan struktur primer. Proses berikutnya terjadilah pelipatan/folding menjadi satu struktur 3 dimensi yang unik/sejenis saja untuk satu protein. Proses ini dalam hirarkinya dimulai dari struktur primer, kemudian struktur sekunder dengan contohnya struktur alfa heliks, beta-sheet, turn dan random. Struktur sekunder ini bersifat lokal terhadap struktur utuh protein. Berbagai struktur sekunder berkumpul menjadi struktur tersier yang merupakan struktur utuh satu rantai polipeptida. Ada yang berbentuk seperti pegas, bola, keranjang dan lain-lain. Bila protein itu terdiri lebih dari satu rantai polipeptida maka struktur akhir yang dibentuk disebut struktur kuarternar. Inilah akhir dari prinsip dogma sentral yang mendasari seluruh fenomena biologi molekuler yang penting dimengerti sebelum mengolah data dengan perangkat-perangkat lunak untuk penerapan bioinformatika.

UNSUR BIOINFORMATIKA

Saat ini informasi dasar utama yaitu sekuen genom manusia telah dibaca. Ada setidaknya 4 jenis sekuen genom manusia yang didepositkan di pusat data yaitu data campuran genom puluhan manusia tanpa nama yang dibaca oleh

¹¹ Subiyanto, P., Witarto, A.B., Rukmo, M., and Putra, S.T. Methylation of p14ARF gene in patients from RS Moewardi, Solo with oral cancer. Presentasi pada 5th IDA-FDI Joint Meeting, Denpasar, 14 November 2009.



Konsorsium Pembacaan Genom Manusia¹², data campuran genom 4 orang manusia tanpa manusia yang dibaca oleh perusahaan genom Amerika Serikat Celera Genomics¹³, kemudian data genom manusia bernama Craig Venter pemilik Celera Genomics¹⁴, dan terakhir data manusia tanpa manusia dari Korea yang dibaca oleh para ilmuwan dan perusahaan pembacaan genom dari Korea. Seluruh informasi sekuen DNA dari genom-genom itu bisa dibaca dari 3 pilihan pusat data DNA di dunia yaitu GenBank di Amerika Serikat (www.ncbi.nlm.nih.gov/), DDBJ/DNA Databank of Japan di Jepang dan EBI/European Bioinformatics Institute di Inggris. Trend bioinformatika saat ini adalah mengolah data biologi molekuler dengan berbagai jenis penerapan pada ilmu biologi/bioteknologi. Misalnya untuk cabang mikrobiologi, penerapan bioinformatika untuk indentifikasi genus/spesies dan mencari kekerabatan (filogenetik). Cabang ilmu biokimia mengandalkan bioinformatika untuk prediksi dan pemodelan struktur protein. Cabang genetik tertarik menggunakan bioinformatika untuk mencari kandidat-kandidat gen baru. Ilmu Biologi Sel berminat pada aplikasi bioinformatika untuk melihat interaksi gen atau protein satu dengan yang lain dalam sebuah metabolisme yang kompleks. Kegiatan ini melahirkan ilmu baru yang disebut Biologi Sistem. Cabang ilmu Biologi Pertumbuhan (Developmental Biology) termasuk Biologi Kanker misalnya menggunakan bioinformatika untuk melihat gen-gen tertentu yang aktif pada salah satu tahap pertumbuhan sel/transformasi sel normal menjadi sel kanker menggunakan data-data dari DNA microarray. Dan banyak contoh penerapan lainnya. Walau ada banyak sekali perbedaan aplikasinya, semua kegiatan analisis bioinformatika mengandung 3 unsur yang sama. Yaitu, 1. “Pusat data/database”. Kegiatan di sini bisa berbentuk menggunakan atau membuat pusat data. Contoh pusat data antara lain GenBank untuk data DNA, SwissProt untuk data protein, Protein Data Bank untuk data struktur protein/DNA, dan pusat data ekspresi RNA. 2. Unsur “Analisis”. Ada dua bentuk utama analisis yaitu mencari kesamaan/homologi. Sebab kesamaan sekuen/struktur menunjukkan kesamaan fungsi biologi. Untuk kesamaan informasi yang berbentuk linier (sekuen DNA, sekuen protein) digunakan teknik alignment/pensejajaran. Sementara untuk informasi yang berbentuk struktur ruang 3D digunakan teknik superimpose. Analisis yang kedua adalah pencarian pola. Misalnya adalah pola ekspresi gen pada sel kanker vs pada sel normal yang datanya diperoleh dari eksperimen DNA microarray. Dan unsur terakhir, 3. adalah unsur “Prediksi”. Dari “Analisis”, dilakukan prediksi yang disebut in-silico sebagai analogi dari in-vivo (fenomena dalam lingkungan hidup yang asli), in-vitro (dalam lingkungan buatan/tabung reaksi) dan in-silico untuk fenomena yang dianalisis menggunakan chip komputer yang bahan utamanya adalah silikon. Prediksi ini termasuk kemudian lebih jauh menjadi simulasi. Setiap kegiatan bioinformatika, apapun penerapannya, pasti mengandung salah satu atau semua tiga unsur tersebut.

¹² Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921.

¹³ Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. *Science* 291: 1304–1351.

¹⁴ Levy, S. et al. The diploid genome sequence of an individual human. *PLoS Biol.* 5(10): e254, 2007.



Sebelum menjelaskan contoh masing-masing unsur, sekali lagi pembaca perlu diingatkan bahwa bioinformatika bukan sekedar penggunaan perangkat lunak. Sebab perangkat lunak bila diberikan input pasti keluar output tanpa peduli apakah hasilnya benar atau tidak. Dari mulai input, perlu dilakukan penyaringan data, penetapan parameter yang cocok untuk fenomena biologi yang akan dilihat. Dan ketika output diperoleh, perlu dicocokkan dengan fenomena biologi yang sudah dipahami apakah sejalan atau tidak. Untuk itu hasil-hasil eksperimen in-vivo atau in-vitro harus didahulukan daripada in-silico atau sebaliknya hasil eksperimen in-silico ini dibuktikan dengan eksperimen in-vitro atau in-vivo.

Sekarang marilah kita lihat bagaimana unsur “Analisis” dilakukan. Sekuen DNA atau sekuen protein dicari kesamaannya terhadap sekuen yang ada lainnya di pusat data GenBank atau SwissProt menggunakan teknik alignment. Ada dua jenis alignment yaitu local dan global. Local alignment menerapkan satu batasan/window dengan panjang sekuen tertentu yang diterapkan pada sekuen yang dicari. Jadi misalnya panjang sekuennya 1.000 pb, perangkat lunak akan memotong-motongnya menjadi katakan 20-100 pb, baru mencari kesamaan sekuen di pusat data. Kentungannya waktu yang diperlukan untuk mencari jadi pendek, tapi kerugiannya hasil yang didapatkan adalah potongan pendek bukan sekuen utuh. Perangkat lunak on-line yang paling banyak digunakan menggunakan teknik local alignment ini misalnya adalah BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi). Walaupun dilakukan pencarian terhadap jutaan entri data, BLAST bisa menghasilkan sekuen yang mirip dalam waktu detik sampai menit saja. Jadi local alignment cocok digunakan untuk pencarian awal sekuen yang mirip. Sementara global alignment membandingkan keseluruhan sekuen dari ujung ke ujung. Bila tidak ada bagian yang sama/mirip maka perangkat lunak menentukan itu sebagai gap/lubang sebagai insersi atau dilesi. Keuntungannya bisa mendapatkan alignment dengan sekuen utuh tapi kerugiannya waktu yang sangat lama. Perangkat lunak yang sering digunakan misalnya CLUSTAL, versi on-line-nya hanya mampu menampung alignment terhadap 500 entri. Bayangkan bila kita mau membandingkan sekuen gen penyandi hemaglutinin dari virus flu yang jumlahnya sekarang mencapai 10.000 ribu entri di GenBank, untuk menemukan sekuen yang khas virus flu H5N1 saja. Jadi global alignment cocok untuk pencarian lanjutan setelah kita menyaring jumlah entri berdasar kemungkinan paling tinggi. Keberhasilan pertama untuk pencarian gen seperti ini adalah pembuktian bahwa ada gen virus yang juga terdapat pada manusia pada tahun 1983¹⁵. Sekarang fenomena seperti ini sudah menjadi hal yang biasa diterima karena diketahui adanya virus-virus yang dapat menyisipkan gennya ke dalam genom manusia seperti HPV, HIV dan lain-lain.

Data pada hirarki berikutnya adalah RNA. Informasi yang penting umumnya bukan sekuen RNA karena sekuen RNA sama persis dengan sekuen DNA. Tapi informasi yang penting adalah tingkat ekspresi RNA yang datanya diperoleh dari eksperimen menggunakan DNA microarray. Bentuk datanya adalah angka yang menunjukkan tingkat ekspresi gen tertentu pada satu kondisi. Misalnya ekspresi 30 ribu gen pada sel kulit kanker berarti ada 30 ribu baris data versus

¹⁵ Viral oncogene v-sis was found to be a modified form of the normal cellular gene that encodes platelet-derived growth factor!!! Discovered by R.F. Doolittle et al. Science 221, 275 (1983) & M.D. Waterfield et al. Nature 304, 35 (1983).



ekspresi 30 ribu gen yang sama sama pada sel kulit normal. Analogi sel kulit kanker vs sel kulit normal ini mungkin lebih mudah dibayangkan bila kita melihat fenomena ulat yang berubah menjadi kupu-kupu. Mengikuti prinsip biologi molekuler yang sudah diterangkan sebelumnya, kita tahu setiap sel ulat dan setiap sel kupu-kupu sebenarnya memiliki sekuen genom yang sama persis. Perbedaan fisiologi yang mencolok antara ulat yang merambat dengan kupu-kupu yang terbang terdapat pada gen mana yang aktif. Ilmu ini disebut ilmu genomika fungsional (functional genomics). DNA microarray adalah sebuah chip yang di atasnya dilekatkan potongan pendek DNA tunggal/single strand yang akan berpasangan dengan DNA yang cocok dari sampel sel yang diteliti bila DNA itu mengkode gen dan aktif menjadi RNA. Proses berpasangan ini disebut hibridisasi. Tingkat ekspresi gen diketahui dari kekuatan sinyal fluoresens dari potongan DNA dari sel yang sudah dilabel dengan fluorosens. Semakin banyak DNA yang aktif, berarti cahaya fluoresens semakin terang. Scanner khusus yang akan membaca kekuatan sinyal. Ukuran DNA microarray hanya sebesar korek api yang mampu menampung sampai puluhan ribu informasi gen. Tak heran harganya masih sangat mahal mencaai hampir 5 juta Rupiah per-keping untuk sekali pakai. Setelah diperoleh data berupa angka yang mewakili tingkat ekspresi gen, maka dilakukan “analisis” pencarian pola¹⁶. Sebab umumnya sebuah fenomena kompleks dalam sebuah sel tidak disebabkan oleh satu dua gen saja, tapi banyak gen yang saling berinteraksi. Pencarian pola yang disebut clustering ini seperti mengelompokkan antara orang yang berwajah bulat, berdagu lancip, berambut pirang, berkaca mata sebagai ciri-ciri yang dicari di antara kumpulan orang yang hanya berdagu lancip saja, atau berambut hitam saja dan seterusnya.

Pada hirarki struktur protein, pusat data yang kita kunjungi akan menyuguhkan data-data berbentuk teks yang bila disimak berbentuk angka koordinat pada sumbu X, Y, Z dari masing-masing atom penyusun makromolekul protein atau DNA. Kita memerlukan perangkat lunak khusus seperti Raswin yang tersedia gratis di Internet untuk bisa memvisualisasikan struktur 3D dari protein tersebut. Menggunakan komputer dengan kecepatan yang lebih tinggi, informasi 3D itu bisa diolah menjadi simulasi dinamika protein. Seperti bagaimana molekul air melewati protein aquaporin secara spesifik.

MANFAAT BIOINFORMATIKA DALAM PENGEMBANGAN KEILMUAN BIOLOGI

Marilah berikutnya kita lihat dua contoh di dunia dan Indonesia untuk melihat manfaat bioinformatika dalam pengembangan keilmuan biologi.

Dalam biologi ada misteri besar yang coba dipecahkan ilmuwan dengan teori evolusi Darwin. Sebagian umat beragama termasuk bangsa Indonesia meyakini teori penciptaan yaitu manusia diciptakan oleh Tuhan. Sementara teori Darwin menyandarkan pada fenomena seleksi alam untuk menjelaskan bahwa nenek moyang manusia adalah kera. Mana yang benar?

Ledakan informasi biologi molekuler telah memungkinkan membaca genom manusia Eropa, Amerika, Asia, Afrika juga manusia purba homo

¹⁶ Nugroho, A.S., Witarto, A.B., and Handoko, D. Support Vector Machine: Teori dan Aplikasinya Dalam Bioinformatika. IlmuKomputer.com, pp. 1-11 (2003).



Neanderthal yang berasal dari Jerman. Sebuah proyek internasional yang didanai oleh lembaga Amerika Serikat, National Geographic, perusahaan komputer IBM dan Yayasan keluarga Waitt pemilik perusahaan komputer Gateway beberapa tahun terakhir ini diluncurkan dengan nama Genographic Project yang diketuai oleh ahli genetika, Spencer Wells¹⁷. Proyek ini pertama membaca sekuen gen tertentu dari suku-suku yang masih asli dari seluruh penjuru dunia lalu membandingkannya dengan bioinformatika. Tujuannya untuk melacak siapa Ibu pertama dan Bapak pertama manusia modern Homo sapiens di muka bumi ini. Untuk melacak jejak Ibu pertama digunakan sekuen gen dari mitokondria. Sebabnya karena jumlah mitokondria pada sel telur 99,99 % dari pada sel sperma yang membuahnya jadi setiap manusia, DNA mitokondrianya bisa dibilang pasti berasal dari Ibu, bukan Bapak. Itulah yang dijelaskan di bagian pertama tulisan ini bahwa DNA mitokondria menyimpan rahasia asal-usul manusia. Sementara untuk melacak Bapak pertama digunakan sekuen dari kromosom Y yang hanya terdapat pada laki-laki. Bagaimana bioinformatika menjawab soal besar ilmu biologi ini? Pertama, kita perlu mengakses pusat data GenBank melalui internet yang gratis. Kita perlu unduh gen mitokondria manusia dari suku Tengger dan suku Manado misalnya. Setelah dapat, kita gunakan perangkat lunak CLUSTAL yang sudah diperkenalkan sebelumnya. Hasilnya dari 546 pb yang telah ditentukan, hanya 2 yang berbeda. Ini menunjukkan suku Tengger dan suku Manado relatif berkerabat dekat. Maklum tinggal di daerah yang berdekatan. Sekarang apa yang terjadi kalau kita membandingkan DNA mitokondria dari suku Tengger dengan suku Bamileke dari Afrika pada wilayah DNA yang sama. Hasilnya kali ini terdapat 11 perbedaan basa. Jumlah yang sangat besar dibanding 2 sebelumnya. Hal ini menunjukkan kekerabatan yang jauh antara kedua suku ini yang bisa juga diamati dari bentuk badan, warna kulit, bentuk rambut dan seterusnya. Menggunakan 78.590 data sekuen DNA mitokondria dari suku-suku yang ada di seluruh dunia, proyek Genographic bisa membuat peta silsilah yang para ilmuwan menyebutnya sebagai pohon filogenetik tentang asal-usul manusia pertama berasal dari mana. Teori yang menyatakan bahwa manusia berasal dari banyak tempat melalui proses seleksi alam ternyata terbantahkan dan memastikan kejadian manusia dari satu tempat yang berada di Afrika, saat ini wilayah Somalia. Dengan demikian bioinformatika mendukung teori satu Ibu dan satu Bapak dari seluruh umat manusia saat ini.

Bila contoh di atas berskala besar, menggunakan komputer super cepat dan dana besar, bagaimana mungkin kita di Indonesia menggunakan teknologi bioinformatika untuk memajukan ilmu biologi dengan segala keterbatasan yang ada?

Ketika kembali ke Indonesia, April 2002, penulis menerima seorang mahasiswa S-1 dari Departemen Kimia, FMIPA IPB yang ingin melakukan penelitian pra-tugas akhir yang singkat hanya dalam waktu 1 bulan. Kebetulan minat penulis waktu itu dari studi penulis di Jepang sebelumnya adalah pengembangan enzim untuk sensor glukosa. Sehingga penulis memberi tugas

¹⁷ Behar, D.M., S. Rosset, J. Blue-Smith, O. Balanovsky, S. Tzur, D. Comas, R.J. Mitchell, L. Quintana-Murci, C. Tyler-Smith, R.S. Wells, The Genographic Consortium. The genographic project public participation mitochondrial DNA database. PLoS Genetics 3(6): e104



kepada Noviyon Darmawan, mahasiswa itu untuk menggunakan perangkat-perangkat lunak yang ada untuk mencari kandidat protein yang sama dari pusat data. Penelitian singkat itu berhasil menemukan pola-pola tertentu pada enzim glukosa dehidrogenase dan memprediksi enzim serupa dari sekuen DNA-nya pada beberapa bakteri yang lazim ada di Indonesia. Hasil penelitian yang diajukan kepada kompetisi Pemilihan Peneliti Remaja Indonesia II tahun 2003 yang diselenggarakan oleh LIPI dan Asuransi Bumiputera ini berhasil menjadi juara I untuk kategori IPA¹⁸. Setelah lulus S-1, sang mahasiswa tadi mendapat beasiswa Erasmus Mundus dari Uni Eropa untuk S-2 dan sekarang sedang menempuh S-3 di Universitas Muenster di Jerman.

KESIMPULAN

Ilmu biologi memiliki lahan yang sangat besar untuk dikembangkan di Indonesia, negara dengan sumber daya hayati terbesar di dunia. Kemajuan teknologi informasi dan komunikasi (TIK) melahirkan bioinformatika yang memacu perkembangan ilmu biologi lebih cepat daripada sebelumnya. Soal-soal besar dalam ilmu biologi, sampai hal-hal kecil yang bisa dikerjakan oleh mahasiswa S-1 untuk tugas akhir, bisa digarap oleh dosen dan mahasiswa yang berminat dengan modal komputer laptop seharga 3-4 juta-an dan jaringan internet dari warnet dengan biaya 2.000-3.000 Rupiah/jam. Informasi sekuen DNA manusia mana yang membuat orang Jawa lebih rentan terkena diabetes dan orang Padang terkena penyakit jantung, bisa dicari dari pusat data GenBank yang gratis daripada mencari informasi-informasi selebriti yang tidak pantas dicontoh. Semua kembali kepada keinginan kita sendiri untuk maju karena semua kemungkinan tersedia di depan mata.

¹⁸ Darmawan, N. Studi bioinformatika, pencairan berbasis struktur terhadap enzim PQQ dehidrogenase untuk biosensor glukosa pada genom mikroba. Prosiding Pemilihan Peneliti Remaja Indonesia II tahun 2003. Pemenang I bidang Ilmu Pengetahuan Alam.

