

Landskap biologi pengkomputeran dalam penyelidikan padi di MARDI

(Landscape of computational biology in rice research at MARDI)

Rabiatul Adawiah Zainal Abidin, Norliza Abu Bakar, Sew Yun Shin dan Sanimah Simoh

Pengenalan

Nasi ialah makanan ruji penduduk di Asia Tenggara. Menjelang tahun 2050, dianggarkan populasi penduduk dunia akan mencecah sehingga 9 bilion (FAO 2009). Oleh yang demikian, sumber makanan yang mencukupi seperti nasi yang diperolehi daripada tanaman padi sangat penting bagi memenuhi permintaan penduduk dunia. Perubahan cuaca turut memberi tekanan yang hebat kepada petani kerana kemarau dan hujan yang berpanjangan akan mengurangkan hasil pertanian mereka. Permintaan dan cabaran ini mendorong pembaikan baka padi untuk menghasilkan ciri tanaman padi berkualiti seperti rintang penyakit, berhasil tinggi, tinggi kandungan nutrien, rintang banjir serta tahan kandungan garam yang tinggi bagi memenuhi keperluan keselamatan dan sekuriti makanan dunia.

Sejak dahulu lagi bidang pembiakbakaan telah menghasilkan pelbagai jenis varieti padi dengan pelbagai ciri seperti varieti padi wangi, rendah indeks glisemik, berhasil tinggi, ketahanan serangan penyakit dan serangga serta ketahanan banjir. Sejarah dengan kemajuan teknologi, pembiakbakaan padi telah bergabung dengan bidang bioteknologi untuk menghasilkan varieti baharu padi berpandukan maklumat genomik dan genetik tanaman atau lebih dikenali sebagai pembiakbakaan berbantuan penanda molekul (MAS).

Melalui teknologi terkini, ahli pembaikan baka tidak sekadar melihat kepada ciri fizikal tanaman berdasarkan mata kasar dalam menentukan pemilihan baka terbaik. Sebaliknya maklumat genomik dan genetik padi yang diperolehi melalui kemajuan bioteknologi membolehkan ahli pembaikan baka mendapat akses serta mengeksplorasi sumber data genomik dan genetik untuk digunakan dalam aktiviti pembiakbakaan padi. Penghasilan sumber data genomik dan genetik disokong oleh bidang biologi pengkomputeran yang semakin menjadi keperluan utama penyelidikan kerana diperlukan untuk menganalisis dan mengurus data genomik dan genetik berskala tinggi yang dihasilkan oleh kemajuan bioteknologi terkini seperti platform penjujukan, pemprofilan dan penjenutihan.

Biologi pengkomputeran adalah bidang sains yang bersifat multidisiplin yang menggabung dan mengaplikasikan teknik pengkomputeran untuk menganalisis data biologi seperti jujukan DNA, populasi genetik, gen serta protein dalam meramal, menganalisis serta memperoleh penemuan baharu

untuk penyelesaian masalah biologi yang rumit. Teknik pengkomputeran melibatkan kaedah analitik, permodelan dan simulasi dengan menggunakan perisian saintifik yang berteknologi tinggi.

Pada tahun 2009, *International Rice Genome Sequencing Project* (IRGSP) dengan kerjasama beberapa institut bersifat bukan menjana keuntungan dan akademik telah menjalankan projek penjujukan genom padi varieti Nipponbare daripada subspecies *japonica* dan varieti 93-11 daripada subspecies *indica*. Biologi pengkomputeran (BP) adalah salah satu pendekatan yang telah diaplikasikan untuk menghasilkan rujukan genom padi Nipponbare dan 93-11. Kaedah BP digunakan untuk menghimpunkan jujukan genom padi daripada saiz jujukan pendek sehingga membentuk keseluruhan genom yang bersaiz 430 Mb serta mencirikan gen dan protein putatif padi yang mengawal proses biologi dan fungsi molekul tertentu dalam pembentukan fenotip tanaman.

Sehingga 4 Ogos 2017, sejumlah 35,828 gen telah dikenal pasti dalam genom padi yang bersaiz 430 Mb. Setiap gen mempunyai lokasi fizikal di dalam kawasan genom serta bertaburan pada 12 kromosom padi. Analisis maklumat genomik dan genetik yang terhasil daripada teknologi penjujukan melalui aplikasi biologi pengkomputeran berupaya menghasilkan genom pelbagai varieti padi sebagai rujukan genom bagi penjajaran penanda molekul penanda polimorfisme tunggal (SNP) dan mikrosatelit (SSR), pencirian SNP dalam kawasan gen sama ada bersifat mutasi salah erti dan mutasi sama erti, peta rangkaian genetik padi, pembinaan tapak jalan biosintesis metabolit, hubung kait ciri genotip dan fenotip serta lain-lain lagi.

Sehingga kini, terdapat pertambahan dua genom padi varieti IR64 dan Kasalath (<http://snp-seek.irri.org/>) yang sesuai dijadikan sebagai genom rujukan. IR64 ialah varieti padi daripada subspecies *indica* sementara Kasalath daripada subspecies *aus*. Kemudahan akses kepada empat jujukan genom padi tersebut didapati telah banyak memanfaatkan para penyelidik padi dan memberi impak positif menerusi pemantapan aktiviti penyelidikan padi. Ini disebabkan oleh penghasilan maklumat genomik dan genetik yang penting telah digunakan untuk membangunkan serta menambah baik varieti padi mengikut ciri yang diinginkan.

Antara kajian dan penyelidikan padi yang menggunakan kaedah biologi pengkomputeran di serata dunia adalah penjujukan genom padi varieti tahan penyakit karah bagi memperoleh senarai gen rintang penyakit karah seperti NB-ARC, LRR dan protein kinase (PK), penemuan penanda molekul polimorfisme nukleotida tunggal (SNP) dan mikrosatelit (SSR) yang diramal terlibat dengan ciri rintang kemarau dan ketahanan garam tinggi bagi varieti Nagina 22 dan Pokkali, pengenalpastian gen-gen yang terlibat dengan ciri tekanan dan lain-lain.

Selain itu, kaedah BP turut diaplikasikan dalam pencirian profil metabolit pelbagai subspesies padi yang memetakan metabolit kepada tapak jalan biokimia supaya ciri metabolik dan nutrien terlibat pada subspesies padi tersebut dapat diramalkan.

MARDI juga tidak ketinggalan dalam penyelidikan tanaman padi berasaskan aplikasi biologi pengkomputeran. Pengenalpastian gen-gen yang diekspreskan dalam tisu tanaman atau sela masa tertentu, penemuan penanda SNP yang terlibat dengan penyakit padi seperti hawar seludang serta pemprofilan kandungan nutrien padi tempatan telah berjaya diperolehi.

Secara ringkasnya, aplikasi dan biologi pengkomputeran dalam penyelidikan padi (*Gambar rajah 1*) terbahagi kepada kategori berikut: 1) penemuan dan pencirian terhadap gen, protein, metabolit dan penanda molekul, 2) pembangunan protokol analisis, perisian dan pangkalan data serta 3) pembangunan infrastruktur

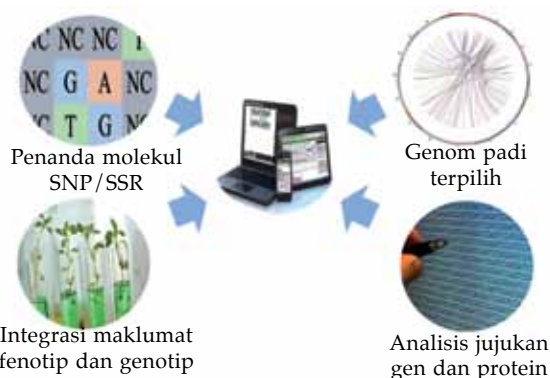
Penemuan dan pencirian gen, protein, metabolit dan penanda molekul melalui kaedah biologi pengkomputeran

Antara objektif analisis data padi dengan menggunakan kaedah biologi pengkomputeran adalah untuk memperoleh penanda molekul berdensiti tinggi, gen dan protein yang diekspreskan serta metabolit yang dikenal pasti sebagai bahan bioaktif tanaman.

Pengenalpastian gen padi menggunakan kaedah biologi pengkomputeran bermula dengan inisiatif kajian pengekspresan gen padi rintang dan tahan kemarau pada tahun 2011. Ia bertujuan untuk mengenal pasti jenis dan bilangan gen yang diekspreskan dalam varieti padi Aeron 1. Pengenalpastian gen tersebut penting sebagai asas pengetahuan tentang gen yang berpotensi terlibat dengan mekanisme sistem tahan kemarau.

Selain itu, pemencilan RNA daripada sampel beras berwarna untuk kajian pengekspresan gen yang terlibat dalam mekanisme penghasilan pigmen hitam dan merah pada perikap padi turut dijalankan bagi tujuan mengenal pasti gen-gen yang mengawal penghasilan pigmen merah dan hitam.

Di Jepun, sekumpulan penyelidik daripada NARO *Agricultural Research Centre* Jepun telah mengenal pasti gen *Pi9* yang berpotensi terlibat dalam mekanisme pertahanan penyakit karah padi melalui kaedah penjujukan dan analisis data penjujukan gen. Kemudian gen *Pi9* dipencilkan daripada padi varieti Jepun dan dijujuk.



Gambar rajah 1. Antara peranan biologi pengkomputeran dalam penyelidikan padi

Melalui penemuan jujukan DNA gen *Pi9*, boleh diklonkan dalam varieti padi tempatan. Seterusnya, perbandingan antara varieti padi rintang karah tempatan dengan varieti padi rentan karah tempatan dilakukan melalui penjajaran jujukan gen *Pi9* di kalangan varieti tersebut dengan menggunakan kaedah biologi pengkomputeran.

Pengekspresan gen-gen yang terlibat dengan mekanisme biologi bagi ciri padi seperti yang dinyatakan di atas melibatkan teknik penjujukan yang menghasilkan data jujukan DNA berskala tinggi dengan anggaran saiz fail mencecah puluhan giga bait (Gb) yang bergantung pada genom saiz tanaman. Data jujukan yang terkandung dalam fail tersebut diproses menggunakan teknik biologi pengkomputeran bagi mengenal pasti jujukan gen berpotensi. Ia menggunakan pelbagai perisian biologi pengkomputeran tertentu dan pangkalan data biologi yang spesifik bagi menyokong kualiti analisis data. Melalui penggunaan perisian dan pangkalan data berkaitan, maklumat gen akan dapat dicirikan. Seterusnya penyelidik dapat memilih gen yang diinginkan untuk eksperimen dan pengesahihan dalam aktiviti penambahbaikan ciri padi rintang dan tahan kemarau.

Selain penjujukan dan pengekspresan gen padi, penjujukan genom varieti padi tempatan di MARDI yang bermula sekitar tahun 2016 turut menggunakan teknik BP. Antara jujukan genom padi yang telah dijujuk ialah padi berwarna hitam (Bali, Pulut Hitam 9) dan padi merah (MRM 16, MRQ 100) untuk kajian padi spesialiti, padi berhasil tinggi (MR 297), padi wangi (MR Q76) serta padi varieti untuk kajian penyakit hawar daun bakteria (BLB) (IRBB66, IRBB7) dan padi varieti bagi kajian penyakit karah (BLAST) (IRBL20 dan Tetep). Fokus utama penjujukan ini adalah untuk mengenal pasti penanda molekul SNP dan gen yang berpotensi mewakili ciri agronomi padi seperti nutrien antioksidan, folat dan rintang penyakit. Memandangkan saiz data jujukan genom adalah berskala tinggi, maka penyelidik bioinformatik di MARDI telah membangunkan protokol dan manual yang bersesuaian untuk dirujuk bagi menetapkan penggunaan perisian yang terlibat serta meminimumkan sebarang kesalahan atau bias semasa menganalisis data.

Teknik biologi pengkomputeran juga digunakan untuk menyaring dan memilih data SNP berskala tinggi berpandukan kriteria dan parameter tertentu. Semasa proses penyaringan dan pemilihan SNP, terdapat kriteria serta format yang telah ditetapkan bagi mendapatkan SNP yang berpotensi sebagai penanda molekul. Dalam proses ini, penyelidik biologi pengkomputeran akan menulis beberapa program komputer bagi menyaring SNP yang berpotensi dengan berpandukan kriteria tersebut. Ia adalah mustahil untuk melaksanakan proses penyaringan data SNP, gen atau protein satu persatu tanpa menggunakan perisian dan bahasa pengaturcaraan.

Pembangunan protokol dan perisian serta pangkalan data

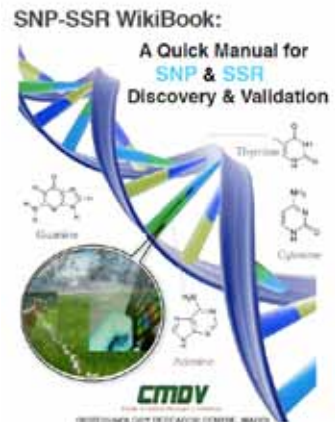
Protokol analisis biologi pengkomputeran bagi penemuan penanda molekul SNP, SSR dan gen juga telah dibangunkan dan diperkenalkan semasa pameran MSTE 2013 bagi kategori pengetahuan. Protokol analisis biologi pengkomputeran telah dibukukan sebagai manual analisis penemuan SNP, SSR dan gen. Buku ini menyediakan langkah-langkah menganalisis data penanda molekul daripada jujukan, bermula daripada pemasangan perisian, pemilihan perisian, perlombongan SNP dan SSR, sehinggalah pemilihan kriteria SNP dan SSR.

Penyediaan manual dan protokol analisis menggunakan teknik biologi pengkomputeran amat penting supaya analisis data mencapai objektif dan berjalan dengan lancar. Di samping itu, dapat menjimatkan masa dalam mencari protokol atau perisian sebelum memulakan analisis. Tambahan lagi, penyediaan protokol atau manual akan memudahkan analisis pada masa akan datang kerana dokumentasi telah dijalankan secara sistematik. Manual analisis ini (*Gambar 1*), didapati sangat membantu menjadikan aktiviti penjenotipan dan penyelidikan padi di *Centre for Marker Discovery and Validity (CMDV)* berjalan dengan lebih lancar dan sistematik.

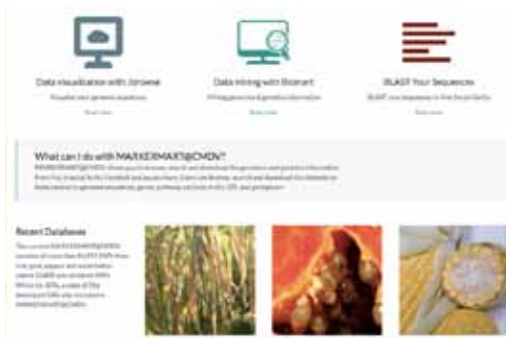
Sehingga kini, terdapat lebih kurang 20 jenis perisian biologi pengkomputeran yang digunakan dalam penyelidikan padi dan dipasang di pelayan serta komputer. Penggunaan perisian dalam biologi pengkomputeran juga merupakan sebahagian daripada alat analitik dan sokongan keputusan (*decision support*) yang membantu penyelidik mendapatkan jawapan berkaitan dengan peramalan atau maklumat pencirian suatu gen yang diinginkan.

Pangkalan data adalah satu platform pengurusan data yang penting dalam bidang penyelidikan kerana menjadi bahan rujukan pada masa akan datang kepada penyelidik lain. Pangkalan data penyelidikan padi telah dibangunkan oleh CMDV yang menyediakan maklumat genomik dan genetik tanaman padi sebagai sumber rujukan dalam penyelidikan molekular padi. Sumber data bagi pangkalan data ini tidak terhad kepada sumber data kajian dalaman, tetapi terdapat sumber data luar.

Pangkalan data ini dikenali sebagai Padi@MarkerMart (*Gambar rajah 2*). Hingga kini, pangkalan data Padi@MarkerMart mengandungi panel 1536 SNP bagi 47 varieti padi berwarna, QTL dengan empat ciri tanaman (kandungan amilosa, wangi, kepanjangan dan rintang penyakit RTBV). Antara matlamat lain pembangunan pangkalan data genomik dan genetik padi adalah untuk menjadikan penyelidikan padi di MARDI ke arah pertanian digital. Salah satu perancangan pada masa hadapan adalah



Gambar 1. Muka hadapan buku manual analisis penemuan SNP dan SSR yang menggunakan teknik biologi pengkomputeran



Gambar rajah 2. Pangkalan data MarkerMart@CMDV yang menyimpan data genomik dan genetik padi untuk rujukan pada kajian-kajian lain

untuk mengintegrasikan maklumat genomik dan genetik padi dengan maklumat fenotip padi serta bank gen padi. Kewujudan pangkalan data bertujuan dapat meningkatkan produktiviti dan keefisienan dalam aktiviti penyelidikan padi. Seterusnya membantu meningkatkan pengeluaran padi di seluruh dunia memandangkan menjelang 2050, populasi manusia mencecah sehingga 9.6 bilion.

Infrastruktur

Infrastruktur merupakan elemen penting dalam menayakan pelaksanaan penyelidikan biologi pengkomputeran. Ia terdiri daripada pelayan komputer yang berkeupayaan tinggi, ruang memori dan storan yang mencukupi supaya dapat menampung kelajuan dan operasi aktiviti pemprosesan data berjalan dengan lancar. Hal ini kerana, lebih tinggi trupert data, lebih banyak ruang storan dan memori yang digunakan. Oleh itu, saintis biologi pengkomputeran bukan setakat merancang strategi analisis data, tetapi perlu mengambil kira keupayaan infrastruktur IT dalam mengendalikan analisis-analisis yang dikenali sebagai analisis berkeupayaan tinggi.

Sistem rangkaian juga sangat penting bagi membolehkan penghantaran data, muat naik dan muat turun data dapat dikendalikan dengan lebih efisien. Ini adalah elemen utama yang menyokong aktiviti penyelidikan biologi pengkomputeran. Maka, infrastruktur hendaklah disediakan memenuhi keperluan serta spesifikasi yang bersesuaian dengan corak pemprosesan dan analisis data. Sekiranya menggunakan teknologi atau peralatan berbentuk pengkomputeran di ladang bagi mengumpul data lapangan, maka infrastruktur merupakan elemen yang harus diberi perhatian.

Kesimpulan

Antara impak yang diperolehi melalui penggunaan biologi pengkomputeran dalam penyelidikan padi adalah pengenalanpastian gen, protein dan metabolit berpotensi yang terlibat dengan ciri padi rintang dan tahan kemarau, pigmen perikap, tahan penyakit, pembangunan cap jari DNA, penanda molekul nutrien padi dan tahan penyakit serta pembangunan pangkalan data genomik dan genetik padi. Secara umum, biologi pengkomputeran bukanlah satu bidang yang asing bagi negara maju kerana aplikasi biologi pengkomputeran telah membantu mereka bentuk ubatan penyakit kanser, memperbaiki ciri tanaman dan ternakan, pemuliharaan sumber alam sekitar dan lain-lain. Oleh itu, tidak mustahil bagi Malaysia untuk mencontohi kejayaan negara luar yang telah berjaya menyepadukan bidang biologi pengkomputeran dalam perubatan, pertanian moden dan alam sekitar. Di samping itu, kerjasama dalam kalangan penyelidik daripada pelbagai bidang yang berkaitan seperti biologi molekul, pembiakbakaan, teknologi maklumat dan bioinformatik perlu dipertingkatkan supaya matlamat penyelidikan berjaya dicapai.

Penghargaan

Kajian ini telah dijalankan di Program Pusat Penemuan dan Pengesanan Penanda Molekul (CMDV) dan Program Agri-omik dan Bioinformatik dengan peruntukan dari pihak EPP14, Kementerian Pertanian dan Industri Asas Tani (MOA) serta projek pembangunan RMK-11 MARDI.

Bibliografi

- Badrun, R., Nasir, K.H., Hashim, H. Abu-Bakar, N. dan Abd-Rahman, Z. (2016). Transcriptome sequencing for discovery of potential drought-tolerant related genes in rice variety, Aeron1.
- Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., Li, Y., Liu, X., Zhang, H., Dong, H., Zhang, W., Zhang, L., Yu, S., Wang, G., Lian, X. dan Luo, J. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature Genetics* 46(7): 714 – 721
- FAO 2009. High-level Expert Forum- How to feed the world 2050. Diperoleh dari http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf
- Fukuoka, S., Yamamoto, S.-I., Mizobuchi, R., et al. (2014). Multiple functional polymorphisms in a single disease resistance gene in rice enhance durable resistance to blast. *Scientific Reports* 4: 1 – 7
- Jain, M., Moharana, K.C., Shankar, R., Kumari, R. dan Garg, R. (2014). Genome wide discovery of DNA polymorphisms in rice cultivars with contrasting drought and salinity stress response and their functional relevance. *Plant Biotechnology Journal* 12(2): 253 – 264
- Mahesh, H.B., Shirke, M.D., Singh, S., et al. (2016). Indica rice genome assembly, annotation and mining of blast disease resistance genes. *BMC Genomics* 17: 1 – 12
- Parida, S.K., Mukerji, M., Singh, A.K., Singh, N.K. dan Mohapatra, T. (2012). SNPs in stress-responsive rice genes: Validation, genotyping, functional relevance and population structure. *BMC Genomics* 13: 1 – 17
- Shahril, A.R., Muhammad Fairuz, M.Y., Norhelwa Ezzah, N.A., Siti Norhayati, I., Ramachandran, K., Rahiniza, K. dan Asfaliza, R. (2018). Single nucleotide polymorphism (SNP) mining of Xa21 gene: A gene controlling bacterial leaf blight (BLB) disease resistance in rice
- Shin, S.Y., Ahmad, M.A., Abd Rashid, M.R., Abu-Bakar, N., Machap, C. dan Zainal Abidin, R.A. (2016). Antioxidant activities, macro-microelement composition of selected Malaysian local rice varieties. *Transactions of Persatuan Genetik Malaysia*, m.s. 2015 – 2012

Ringkasan

Bidang biologi pengkomputeran telah lama digunakan dalam pembangunan tanaman dan komoditi seperti padi, buah-buahan, sawit dan getah. Dalam artikel ini peranan dan sumbangan bidang biologi pengkomputeran dalam penyelidikan padi di MARDI telah dibincangkan. Pelbagai jenis projek penyelidikan padi telah mengaplikasikan biologi pengkomputeran bagi membantu meningkatkan produktiviti penyelidikan. Melalui aplikasi biologi pengkomputeran dapat membuka peluang dan potensi supaya penyelidikan tanaman padi lebih produktif dan efisien.

Summary

Computational biology has been used to develop new and improved plant varieties and commodities such as rice, fruits, oil palm and rubber. This article has discussed role and contribution of computational biology in rice research at MARDI. Computational biology has been applied in various rice research projects, which can help researchers to increase their productivity in research activity. Through the application of computational biology, it provides great opportunity and potential in shaping a more productive and efficient rice research in MARDI.

Pengarang

Rabiatul Adawiah Zainal Abidin
Pusat Penyelidikan Bioteknologi dan Nanoteknologi,
Ibu Pejabat MARDI, Persiaran MARDI-UPM,
43400 Serdang, Selangor
E-mel: rabiatul@mardi.gov.my

Norliza Abu Bakar, Sew Yun Shin dan Sanimah Simoh
Pusat Penyelidikan Bioteknologi dan Nanoteknologi,
Ibu Pejabat MARDI, Persiaran MARDI-UPM,
43400 Serdang, Selangor