

## **Pendekatan bioinformatik bagi merungkai maklumat gen dan penanda molekul tanaman**

(Bioinformatics approach to identify gene and molecular markers information in plant)

Rabiatul Adawiah Zainal Abidin, Lina Rozano, Rafidah Badrun dan Umi Kalsom Abu Bakar

### **Pengenalan**

Perkembangan teknologi penyelidikan melalui ciptaan mesin dan kit berteknologi tinggi telah mendorong penyelidik dalam bidang bioteknologi tumbuhan meningkatkan lagi kualiti penyelidikan berasaskan pertanian. Ini tidak dinafikan lagi berikutan projek genom tanaman dan haiwan yang giat dijalankan di seluruh dunia. Sehingga Jun 2012 dianggarkan 22 projek genom tanaman dan haiwan yang telah siap dihasilkan dan telah diterbitkan dalam pelbagai jurnal antarabangsa terulung seperti Science dan Nature. Projek genom tanaman ini sangat penting dalam membantu penyelidik dari serata dunia mengkaji sifat atau trait tanaman, menyiasat variasi dalam genom dan kehadiran gen serta perkaitan dengan sifat-sifat perwarisan yang dikawal oleh pelbagai gen suatu tanaman. Melalui projek genom, penyelidik mencari penyelesaian masalah kepada tanaman dengan menghasilkan varieti tanaman yang berkualiti, bebas serangan penyakit dan serangga, tahan cuaca serta penghasilan tanaman yang bermutu tinggi.

Genom tumbuhan yang pertama diujuk ialah *Arabidopsis thaliana* yang membantu merungkai permasalahan dan maklumat dalam tanaman. Sejak itu, *A. thaliana* telah menjadi model spesies tumbuhan oleh saintis serata dunia. Kajian pada tanaman padi tentang penjujukan dan pemetaan genom padi daripada varieti Nipponbarre di Jepun yang sudah lengkap membantu penyelidik serata dunia mengenal pasti gen padi yang mengekspreskan trait ekonomi. Antaranya ialah gen yang menyumbang kepada penghasilan padi yang lebih subur dan berhasil tinggi, padi aroma, tahan cuaca dan kemarau.

Sehingga kini, maklumat genom tanaman terus meningkat seperti genom tomato, timun, betik, kentang dan tanaman-tanaman lain (*Jadual 1*). Lambakan maklumat genom telah menjadi rujukan kepada tanaman yang masih tidak mempunyai maklumat genom terutamanya buah-buahan tropika seperti dokong, manggis serta rambutan. Kajian telah membuktikan bahawa maklumat genom tanaman yang sedia ada mampu menyokong kajian tanaman yang masih baru terutamanya dalam bidang penyelidikan penyakit dan biodiversiti tanaman.

Jadual 1. Senarai projek penjujukan genom tumbuhan sehingga Mei 2012

Bil.	Nama tumbuhan	Nama saintifik	Terbitan	Tahun terbitan
1.	Arabidopsis	<i>Arabidopsis thaliana</i>	Nature	2000
2.	Padi	<i>Oryza sativa Japonica</i>	Science	2002
3.	Poplar	<i>Populus trichocarpa</i>	Science	2006
4.	Anggur	<i>Vitis vinifera</i>	Nature	2007
5.	Betik	<i>Carica papaya</i>	Nature	2008
6.	Teratai	<i>Lotus japonicus</i>	DNA Research	2009
7.	Sorghum	<i>Sorghum bicolor</i>	Nature	2009
8.	Timun	<i>Cucumis sativus</i>	Nature Genetics	2009
9.	Jagung	<i>Zea mays</i>	Science	2009
10.	Castor bean	<i>Ricinus communis</i>	Nature	2010
11.	Epal	<i>Malus x domestica</i>	Nature Genetics	2010
12.	Strawberi	<i>Fragaria vesca</i>	Nature Genetics	2010
13.	Koko	<i>Theobroma cacao</i>	Nature Genetics	2010
14.	Kacang soya	<i>Glycine max</i>	Nature	2010
15.	Arabidopsis lyrata	<i>Arabidopsis lyrata</i>	Nature Genetics	2011
16.	Kubis china	<i>Brassica rapa</i>	Nature Genetics	2011
17.	Pokok kurma	<i>Phoenix dactylifera</i>	Nature	2011
18.	Kacang dal	<i>Cajanus cajan</i>	Nature	2011
19.	Ubi kentang	<i>Solanum tuberosum</i>	Nature	2011
20.	Tomato	<i>Solanum lycopersicum</i>	Nature	2012

Sumber: Pangkalan data Plant Genom Duplication ([chibba.agtec.uga.edu/duplication](http://chibba.agtec.uga.edu/duplication)) dan Phytozome ([www.phytozome.net/](http://www.phytozome.net/))

Teknologi penjujukan terkini yang telah menyokong projek genom tanaman ini dikenali sebagai Jujukan Generasi atau *Next Generation Sequencing* (NGS). Teknologi Jujukan Generasi telah menjadi fenomena dalam dunia penyelidikan tanaman kerana berkebolehan menghasilkan data jujukan yang terdiri daripada berbilion nukleotida dalam masa yang singkat, cepat dan berkualiti. Walau bagaimanapun, lambakan data biologi ini telah menjadi sebahagian masalah kepada penyelidik dalam bidang biologi molekul kerana kesukaran menganalisis sejumlah data berskala besar. Malah ia sesuatu yang mustahil untuk dicapai tanpa menggunakan pendekatan bioinformatik.

#### **Pendekatan bioinformatik sebagai peralatan menganalisis data tumbuhan**

Bioinformatik ialah bidang yang sangat luas yang merangkumi pelbagai disiplin ilmu. Ia termasuk bidang yang menggunakan kaedah sains komputer, teknologi maklumat dan statistik bagi menyelesaikan permasalahan dalam bidang biologi seperti biologi molekul, biokimia, kesihatan dan biodiversiti.

Jadual 2. Senarai pangkalan data umum tanaman yang boleh dijadikan rujukan penyelidikan

Bil.	Nama pangkalan data	URL	Kegunaan
1.	Plant Genome Duplication Database (PGDD)	<a href="http://chibba.agtec.uga.edu/duplication/">chibba.agtec.uga.edu/duplication/</a>	Katalog maklumat genomik tumbuhan berbunga terutamanya tumbuhan berbunga yang mempunyai genom lengkap
2.	Phytozome	<a href="http://www.phytozome.net">www.phytozome.net</a>	Menyimpan maklumat genomik tumbuhan yang telah dianotasi. Tertumpu pada tumbuhan yang mempunyai jujukan genom
3.	Array Express	<a href="http://www.ebi.ac.uk/arrayexpress/">www.ebi.ac.uk/arrayexpress/</a>	Menyimpan maklumat eksperimen pengekspresan gen yang telah dianotasi daripada pelbagai organisma seperti tumbuhan, haiwan dan manusia
4.	IntAct	<a href="http://www.ebi.ac.uk/intact/">www.ebi.ac.uk/intact/</a>	Katalog interaksi antara molekul yang diperolehi daripada pelbagai eksperimen
5.	Gramene	<a href="http://www.gramene.org">www.gramene.org</a>	Maklumat spesifik berkaitan pertanian seperti penanda molekul tanaman seperti SSR, SNP dan RFLP, gen dan fenotip
6.	PathoPlant	<a href="http://www.pathoplant.de">www.pathoplant.de</a>	Pangkalan data yang menyimpan maklumat interaksi tumbuhan dan patogen
7.	MetaCrop	<a href="http://metacrop.ipk-gatersleben.de/">metacrop.ipk-gatersleben.de/</a>	Pangkalan data tapak jalan metabolik di dalam tumbuhan
8.	Ensembl Plants	<a href="http://plants.ensembl.org/">plants.ensembl.org/</a>	Pangkalan data genom tumbuhan
9.	Genetic & Genomic Information Systems	<a href="http://urgi.versailles.inra.fr/gnpis/">http://urgi.versailles.inra.fr/gnpis/</a>	Menyimpan maklumat genetik dan genomik tumbuhan seperti fenotip, genotip, jujukan, peta genetik dan <i>array</i>
10.	Plant Resistance Genes	<a href="http://niblrrs.ucdavis.edu/">http://niblrrs.ucdavis.edu/</a>	Menyimpan maklumat gen rintang yang terlibat dalam sistem pertahanan tumbuhan

Bioinformatik sangat membantu dalam menganalisis data mentah menjadi maklumat yang bernilai. Ia diperlukan bagi pencarian data dan maklumat biologi seperti DNA, RNA, protein, genom, gen dan motif melalui pangkalan data dan jurnal. Penggunaan pangkalan data bagi pencarian data dan maklumat biologi seperti DNA, protein, genom dan gen merupakan komponen penting dalam pendekatan bioinformatik. Kini telah banyak pangkalan data tanaman yang sesuai untuk dijadikan rujukan oleh penyelidik (*Jadual 2*).

Kebanyakan data pula berbentuk data heterogenus yang merupakan data daripada pelbagai sumber dan format

seperti data eksperimen dan pangkalan data umum. Oleh itu, bioinformatik diperlukan bagi mengintegrasikan data dan maklumat biologi tersebut dengan menggunakan skrip pengaturcaraan atau perisian saintifik. Terdapat penyelidik bioinformatik yang mahir mereka bentuk dan membangunkan perisian saintifik bagi tujuan analisis data jujukan DNA, protein dan kajian statistik.

### **Aplikasi bioinformatik dalam pengekspresan gen tanaman**

Pengekspresan gen merupakan proses gen diekspreskan dalam waktu dan tempat tertentu bagi memahami proses fisiologi yang berlaku. Walau bagaimanapun, bioinformatik diperlukan untuk mengenal pasti pelbagai gen yang diekspreskan oleh tanaman kerana jumlah data yang dihasilkan adalah berskala besar. Pengenalpastian gen yang terlibat dalam pelbagai trait ekonomi seperti pemasakan, pemanisan berhasil tinggi dan tahan penyakit penting sebagai kefahaman terhadap gen yang terlibat dengan trait yang berpotensi untuk dimajukan. Ia dapat meningkatkan tahap kecekapan pemilihan tanaman dan seterusnya penghasilan tanaman bermutu. Misalnya di Pusat Penyelidikan Bioteknologi, penjujukan buah dokong bertujuan mengenal pasti gen yang terlibat dengan aktiviti pembungaan. Penjujukan padi pula dijalankan bagi mengenal pasti gen dan penanda molekul yang terlibat dengan tahan kemarau.

Kajian pengekspresan gen tanaman melalui projek penyelidikan *Discovery of new genes from plants using genomics and bioinformatics* telah dijalankan dengan menggunakan teknologi jujukan generasi kedua bagi mengenal pasti gen yang terlibat dalam proses pembungaan pokok dokong (*Lansium domesticum*). Sebanyak 215 juta jujukan pendek yang bersaiz 92 bes berpasangan (bp) telah dihasilkan daripada 8 sampel dokong yang telah dirawat dalam keadaan stress yang berbeza. Analisis yang dijalankan telah menemui lebih 85,000 gen umum dengan pelbagai fungsi manakala sebanyak 348 gen unik yang berpotensi terlibat dalam tapak jalan pembungaan buah dokong.

Data jujukan gen yang dihasilkan tidak terhad kepada satu maklumat sahaja, tetapi turut memberi maklumat dari segi pengenalpastian gen yang terlibat dengan tapak jalan pembungaan, penemuan fungsi suatu gen, penemuan gen baru yang tidak diketahui fungsinya dan pembinaan tapak jalan pembungaan. Kaedah bioinformatik yang digunakan bagi buah dokong yang tiada rujukan genom ialah analisis jujukan *de novo*. Ini sangat mencabar tetapi yang menariknya adalah penemuan lebih 39,000 bilangan gen yang tidak diketahui fungsinya. Penemuan gen daripada kategori ini memerlukan analisis genomik yang lebih mendalam untuk mengetahui fungsi sebenar gen tersebut.

### **Aplikasi bioinformatik dalam penanda molekul**

Penanda molekul digunakan dalam bidang pembiakbakaan bagi membantu menentukan pemilihan genotip induk bagi trait terpilih. Melalui bantuan penanda molekul, permasalahan pada trait dapat dikenal pasti pada peringkat awal pertumbuhan tanaman. Penanda molekul yang terkini dan kerap digunakan ialah penanda jujukan tunggal atau *Single Nucleotide Polymorphism (SNP)*. Ini kerana taburan SNP yang tinggi dalam tanaman dan kos pengesanan SNP yang menjimatkan.

SNP berperanan sebagai penanda yang membawa trait dalam tumbuhan. Kaedah penjujukan melalui pengekspresan gen atau transkriptomik mampu menemui SNP yang boleh dikaitkan dengan gen yang terlibat dengan ciri yang dikehendaki. Usaha ini berupaya menghasilkan padi wangi yang bertekstur baik, padi rintang kemarau dan padi berhasil tinggi.

Kajian pengekspresan padi Aeron1 dan MR 219 untuk kajian padi rintang kemarau yang terdiri daripada 4 sampel telah menghasilkan lebih kurang 94,000 jujukan pendek bersaiz 92 bp. Pengenalpastian gen dan SNP daripada pengekspresan padi tahan kemarau lebih mudah kerana padi telah mempunyai jujukan genom lengkap iaitu Nipponbarre dan Indica. Sebanyak 897 bilangan SNP berpotensi telah ditemui dalam padi Aeron1 manakala 882 bilangan SNP dalam padi MR 219. Daripada bilangan tersebut hanya 150 bilangan SNP berpotensi mempunyai gen yang terlibat dengan rintang kemarau.

Bilangan SNP yang dihasilkan adalah banyak sama ada dalam ratusan dan ribuan bilangan SNP. Oleh itu, SNP tidak mudah dianalisis satu persatu. SNP yang berpotensi dikenal pasti dengan kaedah bioinformatik yang menggunakan perisian saintifik dan juga pembinaan skrip pengaturcaraan. Selepas pengenalpastian SNP yang berpotensi, pengesanan SNP boleh dilakukan dengan bantuan alat Sequenom atau iScan yang didapati di makmal CMDV. SNP ini boleh diaplikasi bagi pembinaan peta genetik dan analisis *Quantitative Trait Loci (QTL)* dan seterusnya digunakan untuk aktiviti pembiakbakaan. Pemetaan peta genetik dan analisis QTL daripada data panel SNP juga memerlukan kaedah bioinformatik tertentu iaitu melalui penggunaan perisian saintifik seperti MapMaker, JoinMap dan QTL Cartographer.

Pengesanan SNP dengan menggunakan mesin Sequenom dan Illumina Scan (iScan) dapat menghasilkan data genotip yang berskala besar. Namun begitu, adalah mustahil untuk menyusun satu persatu data genotip mengikut format yang dikehendaki oleh penyelidik biologi molekul. Oleh itu, kaedah bioinformatik diperlukan bagi pembinaan skrip pengaturcaraan supaya penyusunan data dilakukan secara automatik bukan secara manual. Kini terdapat 1,536 panel

SNP padi dan 800 panel SNP tembikai telah dibangunkan di makmal CMDV dengan menggunakan pendekatan bioinformatik yang boleh diaplikasikan dalam aktiviti pembiakbakaan.

Aplikasi bioinformatik dalam kajian penyakit tanaman bioinformatik juga telah diaplikasikan bagi menangani penyakit tanaman menerusi penjana tapak jalan biologi yang merungkai interaksi antara patogen penyakit terhadap tanaman yang diserang. Kajian penyakit mati rosot betik yang telah menular sejak tahun 2003 melibatkan asas pengetahuan mengenai sistem imun tanaman betik serta sistem serangan patogen *Erwinia* sp. Asas dalam kajian penyakit tanaman memerlukan data jujukan betik, *Erwinia* sp. serta maklumat interaksi spesifik antara betik dengan patogen *Erwinia* sp. yang terdiri daripada genom, gen, motif, domain dan protein. Maklumat genom, gen dengan protein diperlukan bagi penjana tapak jalan yang melibatkan interaksi antara protein tanaman dengan patogen.

Maklumat daripada genom tanaman dengan analisis terhadap motif dan domain protein membantu penyelidik bagi mengenal pasti gen yang rintang terhadap serangan penyakit. Malah interaksi gen yang menghasilkan protein yang rintang terhadap penyakit tanaman telah banyak diketengahkan oleh penyelidik melalui kajian keperpustakaan dan pangkalan data.

Berdasarkan kefahaman dan maklumat gen, protein, motif dan domain yang spesifik bagi kajian penyakit tanaman boleh dimanipulasi dengan menggunakan aplikasi bioinformatik yang menjana tapak jalan biologi penyakit tanaman. Antara perisian yang digunakan bagi pembinaan tapak jalan biologi ialah Pathway Studio, Ingenuity Systems Pathway Analysis, GeneGo, Ondex, Cytoscape serta Gene Interaction Miner. Hasilnya adalah peramalan tapak jalan mekanisme interaksi antara gen rintang penyakit dalam tanaman terhadap gen virulen bakteria. Mekanisme ini bukan sahaja terhad kepada data yang telah diinputkan malah maklumat interaksi secara tidak langsung antara tanaman dengan patogen seperti penghasilan hormon akibat faktor persekitaran juga akan diintegrasikan secara automatik bagi menyokong kajian penyakit tanaman.

### **Isu dan cabaran**

Lambakan maklumat melalui pelbagai aktiviti penyelidikan yang dijalankan telah mendatangkan impak yang positif kepada penyelidik. Namun demikian, pendekatan bioinformatik sebagai alat bantuan penyelidikan memerlukan pembangunan infrastruktur IT yang berkeupayaan tinggi bagi menyokong kelajuan memproses data serta rangkaian internet yang stabil. Tanpa infrastruktur IT yang kukuh, aktiviti penyelidikan bioinformatik tidak dapat berjalan dengan lancar

dan baik. Penguasaan bidang biologi dan sains komputer juga sangat penting bagi penyelidik bioinformatik.

### **Penutup**

Bioinformatik bukanlah sesuatu yang asing bagi negara maju seperti Kanada, Amerika Syarikat, Jepun dan Netherland yang sudah lama mengamalkan pertanian moden. Pelbagai tanaman utama dengan pelbagai trait yang berjaya meningkatkan ekonomi negara telah dihasilkan seperti padi, gandum, tomato, ubi kentang, epal, tembikai, kanola, jagung dan barli. Oleh itu, tidak mustahil bagi negara membangun seperti Malaysia untuk mencontohi kejayaan negara luar yang telah berjaya menyepadukan bidang bioinformatik dalam pertanian moden. Di samping itu, kerjasama di kalangan penyelidik pelbagai bidang yang berkaitan biologi molekul, pembiakbakaan, teknologi maklumat dan bioinformatik perlu dipertingkatkan supaya matlamat penyelidikan berjaya dicapai.

### **Penghargaan**

Kajian ini telah dijalankan di Pusat Penyelidikan Bioteknologi, MARDI dengan peruntukan dari pihak MOSTI dan MOA.

### **Bibliografi**

- Jena, K.K. dan Mackill, D.J. (2008). Molecular markers and their use in Marker Assisted Selection in rice. *Crop Science* 48: 1266 – 1276
- Hong, C.L., Lorence, M.T., Imelfort, M., Duran, C. dan Edwards, D. (2011). Bioinformatics tools and databases for analysis of next generation sequence data. *Briefings in Functional Genomics* 11(1): 12 – 24
- Martin, J.A. dan Zhong, W. (2011). Next generation transcriptome assembly. *Nature Reviews Genetics* 12: 671 – 682
- Mathiyalagan, R., Subramaniyam, S. dan Jegadeesan, R. (2010). Insilico analysis of drought tolerant genes in rice. *International Journal Biology Medical Research* 1(3): 36 – 40
- Michelmore, R. (2000). Genomic approaches to plant disease resistance. *Current Opinion In Plant Biology* 3: 125 – 131
- Mochida, K. dan Shinozaki, K. (2011). Advances in Omics and Bioinformatics Tools for System Analyses of Plant Functions. *Plant & Cell Physiology* 52(12): 2017 – 2038
- Nielsen, R., Paul, J.S., Albrechtsen, A. dan Yun, S.S. (2011). Genotype and SNP calling from next generation sequencing data. *Nature Reviews Genetics* 12: 443 – 451
- Salzberg, S.L. dan Pop, M. (2007). Bioinformatics challenges of new sequencing technology. *Trend in Genetics* 24(3): 142 – 149
- Thakur, V. dan Varshney, R. (2010). Challenges and strategies for Next Generation Sequencing (NGS) Data Analysis. *Journal Computer Science System Biology* 3(3): 40 – 42

### **Ringkasan**

Pendekatan bioinformatik semakin meluas diaplikasi dalam bidang penyelidikan bioteknologi pertanian. Bioinformatik diperlukan bagi menyelesaikan masalah biologi dalam penyelidikan bioteknologi pertanian. Data asas yang diperlukan bagi penyelidikan ini ialah data jujukan sama ada dalam bentuk genom, gen, motif dan domain protein. Data jujukan tersebut akan dianalisis menggunakan perisian saintifik dan peralatan komputer bagi menghasilkan maklumat biologi yang berguna. Beberapa penyelidikan bioteknologi pertanian di MARDI telah mengenal pasti kepentingan bioinformatik sebagai alat sokongan penyelidikan seperti penanda molekul, pengekspresan gen dan kajian penyakit tanaman.

### **Summary**

Bioinformatics is being widely applied in research and become a vital tool capable of solving biological problems especially in the field of agricultural biotechnology. The basic data input from biological research are mainly sequence data in the form of genomes, genes, motifs and protein domains. These sequences are analysed by dedicated bioinformatics software and computer tools to produce meaningful biological information. Several agricultural biotechnology research conducted in MARDI had demonstrated the importance of bioinformatics as a supporting tool for research in the field of molecular markers, gene expression and crop disease studies.

### **Pengarang**

Rabiatul Adawiah Zainal Abidin  
Pusat Penyelidikan Bioteknologi, Ibu Pejabat MARDI, Serdang,  
Peti Surat 12301, 50774 Kuala Lumpur  
E-mel: rabiatul@mardi.gov.my

Lina Rozano, Rafidah Badrun dan Umi Kalsom Abu Bakar  
Pusat Penyelidikan Bioteknologi, Ibu Pejabat MARDI, Serdang,  
Peti Surat 12301, 50774 Kuala Lumpur