

## ***CpEDb*: Repositori maklumat genetik dan genomik *Carica papaya* dan *Erwinia mallotivora* bagi aplikasi genomik kefungsiian dan pembangunan penanda molekul**

(*CpEDb*: Repository of genetics and genomics information in *Carica papaya* and *Erwinia mallotivora* for application in the functional genomics and molecular markers development)

Rabiatul Adawiah Zainal Abidin, Nor Syahira Mohamad Senin, Azrin Ahmad, Norliza Abu Bakar, Johari Sarip dan Sanimah Simoh

### **Pengenalan**

Platform penjujukan generasi hadapan (PGH) adalah teknologi penjujukan DNA dan RNA yang berskala tinggi, dapat dijujuk dalam jangka masa yang singkat, serta menawarkan kos yang berpatutan kepada pengguna. Platform PGH telah menghasilkan data genetik dan genomik yang dapat digunakan oleh para penyelidik bagi memahami asas genetik dan evolusi suatu tanaman. Data genetik dan genomik tanaman yang dianalisis dapat menghasilkan maklumat berguna seperti gen, protein, metabolit, tapak jalan dan penanda molekul, seterusnya diaplikasikan dalam kajian kefungsiian genomik dan penanda molekul. Namun, penghasilan maklumat genetik dan genomik yang bertruput tinggi memerlukan repositori atau pangkalan data sebagai rujukan komprehensif supaya maklumat genetik dan genomik tersebut diarkib secara teratur serta diakses dengan mudah oleh pengguna.

*Carica papaya* varieti Eksotika (betik Eksotika) telah berjaya menembusi pasaran eksport ke luar negara. Namun, penyakit mati rosot betik (*papaya dieback disease*) yang disebabkan oleh *Erwinia mallotivora* iaitu sejenis bakteria gram-negatif daripada famili Enterobacteriaceae telah menyebabkan kemusnahan yang teruk terhadap tanaman betik di negara ini.

Kekurangan data genetik dan genomik merupakan antara kekangan yang dihadapi oleh penyelidik bagi menjalankan penyelidikan betik dengan lebih komprehensif dan efektif. Penghasilan data jujukan genom betik dan bakteria dijangka dapat memudahkan analisis genetik dan genomik varieti betik tempatan bagi mengenal pasti, menganotasi dan mengklasifikasikan gen kepada beberapa famili gen dan faktor transkripsi yang dikaitkan dengan tekanan abiotik, rintang penyakit, kemasakan buah dan perkembangan bunga. Pengenalpastian dan anotasi famili gen yang terpilih dapat memberikan maklumat mengenai struktur gen, serta hubungan pokok filogenetik berkaitan dengan ortolog atau paralog spesies yang berbeza.

Penjujukan genom betik Eksotika telah dijalankan dengan menggunakan teknologi PGH Illumina HiSeq 2000. Pelbagai jujukan gen dan protein dalam genom betik Eksotika telah

dicirikan dengan menggunakan kaedah bioinformatik. Selain jujukan gen dan protein, metabolit betik Eksotika dan Sekaki turut diprofilkan dengan menggunakan platform metabolomik. Selain itu, penjujukan transkriptom betik Eksotika dan Viorica yang telah dijangkiti oleh patogen *E. mallotivora* turut dihasilkan bagi tujuan mengenal pasti gen-gen berkaitan pertahanan yang telah diekspreskan. Genom patogen *E. mallotivora* turut dijana bagi mengetahui dan memahami komposisi genetik dan genomik *E. mallotivora*, seterusnya mendapatkan mekanisme tindak balas patogen tersebut dalam penyakit mati rosot betik.

Teknologi penjujukan juga berupaya menghasilkan genom pelbagai varieti betik sebagai genom rujukan bagi penjarangan penanda molekul penanda polimorfisme nukleotida tunggal (SNP) dan mikrosatelit (SSR), peta rangkaian genetik betik, pembinaan tapak jalan biosintesis metabolit, hubungan kait ciri genotip dan fenotip serta lain-lain lagi. Misalnya, peta rangkaian betik yang mengandungi 219 SNP dan 21 lokus ciri kuantitatif (QTL) bagi sifat kualiti telah dibangunkan dan akan diaplikasikan dalam strategi pemilihan varieti betik berpandukan penanda molekul SNP.

Walaupun pelbagai maklumat genetik serta genomik betik Eksotika dan *E. mallotivora* telah dihasilkan, namun masih belum ada pangkalan data genetik dan genomik betik Eksotika dibangunkan sebagai rujukan komprehensif kepada para penyelidik. Oleh itu, sebuah pangkalan data iaitu *CpEDb* (<http://www.mycpedb.org>) telah dibangunkan sebagai platform perlombongan maklumat genetik dan genomik bagi betik Eksotika dan *E. mallotivora* (Gambar 1). Buat masa ini, *CpEDb* menyimpan maklumat genetik dan genomik betik Eksotika dan *E. mallotivora* yang telah dihasilkan oleh penyelidik di MARDI dan turut diintegrasikan dengan maklumat omik daripada pangkalan data awam seperti NCBI (<http://ncbi.nlm.nih.gov/>), PlantGDB (<https://www.plantgdb.org/>), Phytozome (<https://phytozome-next.jgi.doe.gov/>) dan PLAZA (<https://bioinformatics.psb.ugent.be/plaza>).

### **Kaedah pembangunan *CpEDb***

Kerangka *CpEDb* dibina dengan menggunakan sistem pengurusan pangkalan data MySQL versi 5.7.25 dan sistem pengurusan reka bentuk laman sesawang 'WordPress' versi 5.0.3. 'WordPress' berfungsi sebagai antara muka bagi memaparkan data secara dalam talian. Bahasa pengaturcaraan 'Hypertext Preprocessor' (PHP5) dan 'Hypertext Markup Language' (HTML5) berperanan sebagai atur cara yang menerima dan menghantar permohonan pencarian data daripada pengguna kepada MySQL dengan menggunakan enjin pelayan Apache versi 2.4.29.

# Carica papaya Eksotika Database (*CpEDb*)

Resources for papaya genomics, genetics, breeding and disease research

HOME GENOME PATHWAY MOLECULAR MARKERS DISEASE TOOLS DOWNLOADS ABOUT US Q

## Exploring *C. papaya* Eksotika genome

*Carica papaya* Eksotika Database (CpEDb) is a curated biological database, which aims to centralise genomics and genetics information for papaya research.

*Carica papaya* L. cv. Eksotika is widely grown in Malaysia. Eksotika has contributed to the import and export industry in Malaysia. However, the genomics and genetics information of this variety is still scarce, which hinder the researcher to investigate and understand the Eksotika genome characteristics. The genome sequences data of Eksotika papaya will facilitate the genomic and genetic analysis of local papaya varieties to gain insight the underlying molecular mechanism in several agronomically papaya traits such as papaya dieback disease, nutritional and ripening.

**About Us**  
*Carica papaya* Eksotika Database (CpEDb) is supported by the following source: MARDI EPP14: Pembangunan Industri Benih.

**Citation**  
If you use *Carica papaya* Eksotika Database (CpEDb) please cite:  
<http://tafs.mardi.gov.my/jtafs/44-2/SNP.pdf>

**Contact Us**  
*Carica papaya* Eksotika Database (CpEDb) team can be reached via this email: [papayae@gmail.com](mailto:papayae@gmail.com)

Copyright 2020 *Carica papaya* Eksotika Database (CpEDb). All Rights Reserved.

Gambar 1. Laman utama sistem pangkalan data *CpEDb*

Buat masa ini, *CpEDb* beroperasi menggunakan pelayan komputer berasaskan sistem pengoperasian Linux-Ubuntu versi 18.06 LTS, 8 GB memori dan 150 GB ruang storan data. Kesemua perisian tersebut merupakan sumber terbuka, mudah dibangunkan, mempunyai nilai-nilai sekuriti dan sangat aktif digunakan oleh komuniti penyelidik dan pentadbir sistem.

## Antara muka dan kandungan sistem pangkalan data *CpEDb*

Maklumat omik yang boleh diperolehi dalam pangkalan data *CpEDb* ialah jujukan gen, protein, ontologi gen, metabolit dan penanda molekul. Maklumat omik tersebut telah dipilih menerusi hasil analisis genom, transkriptom dan metabolomik bagi *C. papaya* serta *E. mallotivora*. Jadual 1 menunjukkan maklumat genetik dan genomik yang terkandung dalam *CpEDb* serta penerangan bagi maklumat tersebut. Pembersihan data dijalankan terlebih dahulu bagi memastikan data yang disimpan mengikut format fail sistem pengurusan pangkalan data iaitu MySQL. Maklumat omik *CpEDb* adalah berbentuk teks, angka dan jujukan.

Jadual 1. Senarai set data yang terdapat dalam pangkalan data *CpEDb* berserta penerangan

Set data	Bilangan data	Penerangan
Jujukan gen	16,165	Jujukan gen betik Eksotika yang telah dicirikan
Jujukan protein	10,125	Jujukan protein betik Eksotika yang telah dicirikan
Ontologi gen	2053	Senarai ontologi gen bagi jujukan gen betik Eksotika
Data metabolit	94	Data metabolit betik Eksotika
Penanda molekul SSR, SNP, InDel	951 SSR 37,510 SNP 16,446 InDel	Penanda molekul SSR, SNP dan InDel daripada data genom betik Eksotika
Gen rintang	208	Jujukan gen rintang betik Eksotika
Gen efektor	13	Jujukan gen efektor <i>E. mallotivora</i>

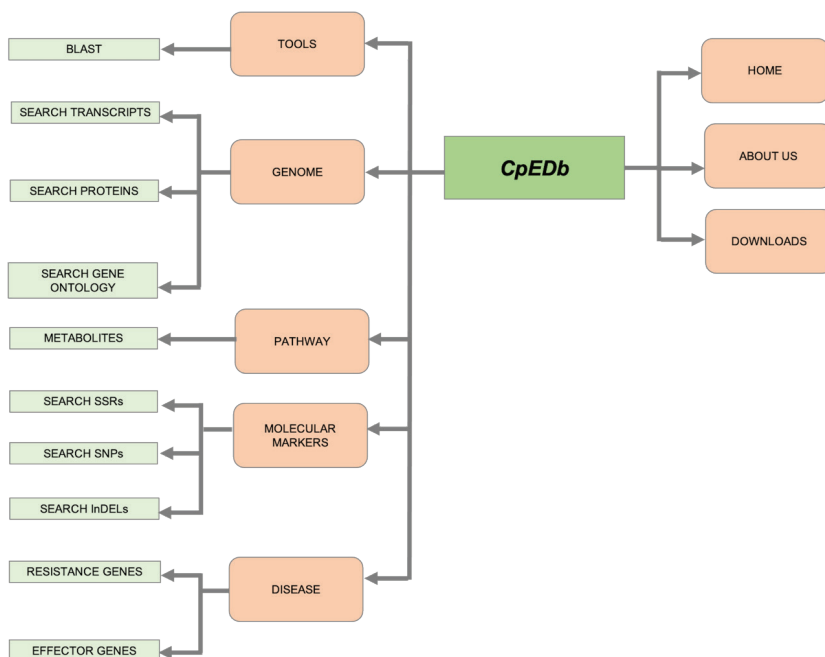
*Rajah 1* menunjukkan peta menu dan submenu yang terdapat dalam *CpEDb*. Terdapat tujuh menu utama yang boleh dipilih oleh pengguna dalam *CpEDb*. Setiap menu utama dan submenu mewakili pelbagai jenis maklumat genetik dan genomik betik Eksotika dan *E. mallotivora* (*Jadual 2*).

Terdapat tiga submenu di bawah menu 'Genome' yang membolehkan pengguna mencari dan menyaring data jujukan gen dan protein serta ontologi gen. Bagi menu 'Pathway' memaparkan data metabolit betik Eksotika yang telah diprofilkan dengan menggunakan platform metabolomik. Tiga submenu di bawah menu 'Molecular Markers' adalah pencarian SSR, SNP dan InDel bagi betik Eksotika.

Kesemua hasil carian dipaparkan dalam bentuk jadual. Pengguna boleh memuat turun hasil pencarian mengikut format CSV, Microsoft Excel dan teks (.txt). Perisian pencarian persamaan jujukan iaitu BLAST yang turut dipautkan bawah menu 'Tools'. Melalui perisian BLAST, pengguna boleh melakukan analisis pencarian persamaan jujukan terhadap pangkalan data terpilih. Perisian BLAST yang digunakan ialah BLAST yang diubah suai oleh Sequence Server dari segi antara muka dan fungsi muat turun yang mesra pengguna (*Gambar 2*).

### **Kelebihan dan keunikan sistem pangkalan data *CpEDb***

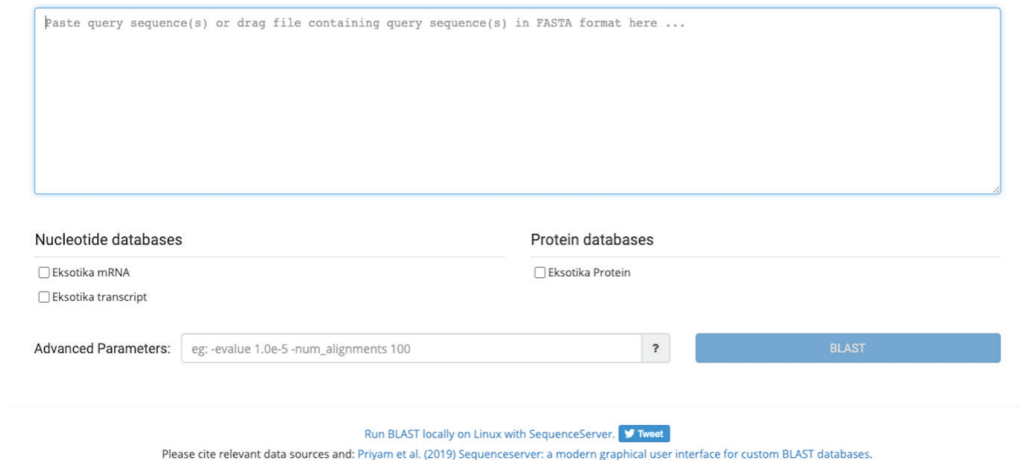
Sehingga kini, sistem pangkalan data *CpEDb* merupakan satu-satunya pangkalan data yang menyediakan maklumat genetik dan genomik yang berkaitan dengan betik Eksotika dan *E. mallotivora*. *CpEDb* disediakan dengan reka bentuk antara muka yang mesra pengguna bagi memudahkan pencarian data yang diinginkan. Pengguna boleh membuat pilihan pencarian secara manual atau menggunakan fungsi muat turun secara pukal mengikut kesesuaian dan pilihan masing-masing. Penggunaan perisian ketiga iaitu BLAST memberi kemudahan kepada pengguna membandingkan jujukan nukleotida atau protein serta mendapatkan maklumat awal berkaitan fungsi dan penglibatan



Rajah 1. Peta menu dan submenu bagi pangkalan data CpEDb

Jadual 2. Senarai menu utama yang disediakan dalam sistem pangkalan data CpEDb berserta penerangan bagi setiap menu tersebut

Menu	Penerangan
'Home'	Laman utama CpEDb. Laman ini menyediakan maklumat ringkas mengenai CpEDb, jenis data dan statistik data yang terdapat dalam pangkalan data CpEDb.
'About Us'	Laman 'About Us' memberikan maklumat berkaitan latar belakang CpEDb dan soalan yang kerap ditanya (FAQ).
'Download'	Laman 'Download' menyediakan data yang boleh dimuat turun secara pukal. Kebiasaannya penyelidik bioinformatik menggunakan fungsi muat turun secara pukal kerana lebih cepat dan data boleh dimanipulasi sebagai kegunaan analisis mereka.
'Tools'	Laman 'Tools' menyediakan perisian BLAST sebagai alat analisis pencarian persamaan jujukan. Pengguna boleh menggunakan data masing-masing bagi melakukan analisis pencarian persamaan jujukan.
'Genome'	Laman ini menyediakan statistik data genetik dan genomik betik Eksotika. Terdapat tiga submenu di bawah 'Genome', iaitu pencarian jujukan transkrip, jujukan protein dan ontologi gen. Pengguna boleh menggunakan kata carian dan fungsi penyaringan bagi mendapatkan data yang diinginkan.
'Pathway'	Laman yang menyediakan data metabolit betik Eksotika.
'Molecular Markers'	Laman yang menyediakan data SSR, SNP dan InDel bagi betik Eksotika.
'Disease'	Laman yang menyediakan data gen rintang bagi betik dan gen efektor bagi <i>E. mallotivora</i> .



Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...

**Nucleotide databases**

Eksotika mRNA  
 Eksotika transcript

**Protein databases**

Eksotika Protein

Advanced Parameters:  ? [BLAST](#)

[Run BLAST locally on Linux with SequenceServer.](#) [Tweet](#)

Please cite relevant data sources and: Priyam et al. (2019) Sequenceserver: a modern graphical user interface for custom BLAST databases.

Gambar 2. Antara muka BLAST yang dapat digunakan untuk analisis pencarian persamaan jujukan

jujukan gen atau protein. Selain itu, *CpEDb* mengikut konsep laman sesawang responsif kerana boleh dilayari dengan pelbagai pelayar seperti 'Chrome', 'Firefox' dan 'Safari' serta boleh juga dilayari dengan pelayar dalam telefon bimbit.

### Penutup

Sistem pangkalan data *CpEDb* menyediakan maklumat saintifik bagi membantu meningkatkan produktiviti penyelidikan genomik kefungsi dan pembaikbakaan tanaman. Pangkalan data *CpEDb* dibangunkan bagi memudahkan pengguna menguruskan maklumat genetik dan genomik yang diperlukan dalam kajian genomik kefungsi dan pembangunan penanda molekul. Melalui pembangunan *CpEDb*, maklumat genetik dan genomik yang telah dihasilkan dapat disepadukan dalam satu platform yang memudahkan pencarian dan pengumpulan maklumat tersebut yang secara tidak langsung meningkatkan kecekapan pengurusan penyelidikan genomik kefungsi dan penanda molekul. Pembangunan dan pengurusan *CpEDb* akan sentiasa dikemas kini dan dikembangkan dari masa ke masa bagi menjamin kualiti dan nilai-nilai saintifik maklumat yang didepositkan.

### Penghargaan

Kajian ini telah dijalankan di Pusat Penyelidikan Bioteknologi dan Nanoteknologi, MARDI dengan peruntukan daripada KRBN001: Pembangunan Industri Benih dan PRB5002: Projek Pembangunan RMK-11

## Bibliografi

- Ling, A.C.K., Zaidan, M.W.A.M., Wee, C.Y., Rozano, L., Supian, S. dan Halim, M.M.A. (2016). A transcriptome next-generation sequencing of *Carica papaya*. *Proceedings of the 3rd Plant Genomics Congress*, Kuala Lumpur, m.s. 11 – 12. Serdang: MARDI
- Nantawan, U., Kanchana-udomkan, C., Drew, R. dan Ford, R. (2018). Development of polymorphic simple sequence repeat (SSR) markers from genome re-sequencing of *Carica papaya* L. 'Sunrise Solo' and 'RB20 for marker-assisted breeding. *Acta Hort.* 1205: 687 – 695
- Redzuan, R.A., Abu Bakar, N., Rozano, L., Badrun, R., Mat Amin, N. dan Mohd Raih, M.F. (2014). Draft genome sequence of *Erwinia mallotivora* BT-MARDI, causative agent of papaya dieback disease. *Genome Announcement* 2: 1 – 14
- Sanimah, S., Maheswary, V., Sarip, J., NurQistina, O. dan Vasanthi, S. (2013). Identification of phytochemicals and the associated genes in Eksotika papaya at ripening index 5 using functional genomics. *J. Trop. Agric. and Fd. Sc.* 41(2): 283 – 308
- Zainal-Abidin, R., Abu-Bakar, N., Yusof, M.M.F. dan Abdullah, N. (2016). Sequence information on single nucleotide polymorphism (SNP) through genome sequencing analysis of *Carica papaya* variety Eksotika and Sekaki. *J. Trop. Agric. Food Sci.* 44: 219 – 228

## Ringkasan

Sistem pangkalan data *CpEDb* (<http://www.mycpedb.org>) telah dibangunkan sebagai pelantar perlombongan maklumat genetik dan genomik betik Eksotika serta *E. mallotivora* yang telah dianalisis dengan menggunakan pendekatan bioinformatik. Selain itu, *CpEDb* turut dihubungkan dengan perisian ketiga, iaitu BLAST yang berfungsi sebagai alatan pencarian persamaan jujukan DNA dan protein bagi betik dan *E. mallotivora*. Sehingga kini, *CpEDb* mengandungi 16,165 jujukan gen dan 10,125 jujukan protein, 94 metabolit, 951 SSR putatif serta 35,934 SNP putatif. Pembangunan *CpEDb* adalah sebagai penyebaran maklumat supaya dapat digunakan oleh saintis dan ahli pembaik baka dalam kajian genomik kefungsiian, pembangunan penanda molekul dan pembaikbakaan betik.

## Summary

The *CpEDb* web-based database (<http://www.mycpedb.org>) was developed as a data mining platform for genetic and genomic information of Eksotika papaya and *E. mallotivora*, which were analysed using a bioinformatics approach. In addition, *CpEDb* is linked to third-party software BLAST, which serves as a sequence similarity search tool for DNA and protein sequences in papaya and *E. mallotivora*. To date, *CpEDb* consists of 16,165 genes and 10,125 protein sequences, 94 metabolites, 951 putative SSRs as well as 35,934 putative SNPs. The aim of *CpEDb* development was to disseminate information for scientists and breeders to use in functional genomics, molecular marker development and molecular breeding in papaya.



**Pengarang**

Rabiatul Adawiah Zainal Abidin (Dr.)

Pusat Penyelidikan Bioteknologi dan Nanoteknologi, Ibu Pejabat MARDI  
Persiaran MARDI-UPM, 43400 Serdang, Selangor

E-mel: rabiatul@mardi.gov.my

Nor Syahira Mohamad Senin dan Norliza Abu Bakar (Dr.)

Pusat Penyelidikan Bioteknologi dan Nanoteknologi, Ibu Pejabat MARDI  
Persiaran MARDI-UPM, 43400 Serdang, Selangor

Johari Sarip (Dr.)

Pusat Penyelidikan Hortikultur, Ibu Pejabat MARDI  
Persiaran MARDI-UPM, 43400 Serdang, Selangor

Sanimah Simoh (Dr.)

Pusat Pengurusan Strategik dan Inovasi, Ibu Pejabat MARDI  
Persiaran MARDI-UPM, 43400 Serdang, Selangor

Azrin Ahmad

Institut Genom Malaysia (MGI)  
Jalan Bangi, 43000 Kajang, Selangor