

Identifikasi famili gen putatif penyandi protease inhibitor dengan pendekatan *in silico* komparatif pada genom *Hevea brasiliensis* Muell. Arg

Identification of putative gene family encoding protease inhibitors by in silico comparative analysis in Hevea brasiliensis Muell. Arg genome

Irfan MARTIANSYAH^{1)*}, Riza Arief PUTRANTO^{1)*} & Nurul KHUMAIDA²⁾

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

²⁾Pemuliaan dan Bioteknologi Tanaman, Fakultas Agronomi dan Hortikultura, IPB, Dramaga, Bogor 16680

Diterima tgl 29 Maret 2017 / disetujui tgl 1 Agustus 2017

Abstract

Protease inhibitors (PIs) are small proteins that form complexes with proteases and inhibits their proteolytic activity. Its potential application as an antimicrobial agent has been studied. Most of PIs' molecule size is around 8-22 kDa depending on their protein families. To date, on the basis of sequence homologies of inhibitor domains, PIs have been classified into 48 families in all organisms. In plant, more than 13 families of PIs have been identified but they were not widely identified in the rubber tree (Hevea brasiliensis Muell. Arg). In the present study, 40 putative HbPI genes, designated as HbPI01 to HbPI36, were identified from whole-genome sequence of rubber tree clone Reyan 7-33-97 using 7453 scaffolds available online in NCBI with the accession code: LVXX01000000. Multiple sequence alignment using MUSCLE algorithm discovered seven conserved motifs (Motifs I-VII) among HbPIs. Phylogenetic analysis of 50 and 36 PI amino acid residues of 32 scaffolds containing putative PI genes from Arabidopsis thaliana and H. brasiliensis showed three clusters (families): LTP-I, SERPIN and LTP-II. LTP-I has 23 putative HbPI genes (HbPI05 to HbPI27) and 12 AtPI genes. SERPIN, a family member of serine protease inhibitor group, has 11 putative HbPI genes (HbPI01 to HbPI04 and HbPI28 to HbPI34) and 22 AtPI genes. LTP-II has 2 putative HbPI genes (HbPI35 to HbPI36) and 16 AtPI genes. In conclusion, this work provides valuable information for further functional characterization of HbPI genes in H. brasiliensis.

[Key words: protease inhibitor, genome-wide, scaffold, in silico, Hevea brasiliensis].

Abstrak

Protease inhibitor (PI) merupakan protein yang membentuk kompleks dengan protease dan menghambat aktivitas proteolitik dari enzim tersebut. Potensi penggunaan protease inhibitor sebagai agensia antimikroba telah diketahui. Kebanyakan PI memiliki ukuran molekul sekitar 8-22 kDa bergantung pada familinya. Saat ini, PI dapat diklasifikasikan menjadi 48 famili di seluruh organisme berdasarkan kemiripan sekuen dari domain inhibitorynya. Pada tanaman, lebih dari 13 famili PI telah diketahui tetapi pada tanaman karet (*Hevea brasiliensis* Muell. Arg) belum diidentifikasi. Pada penelitian ini, sebanyak 40 gen putatif penyandi PI (*HbPI01* hingga *HbPI36*) telah berhasil diidentifikasi dari 7453 scaffold genom utuh tanaman karet klon Reyan 7-33-97 yang tersedia secara daring dengan kode akses LVXX01000000. Penjajaran sekuen menggunakan algoritma MUSCLE memperlihatkan tujuh konservasi motif (Motif I-VIII) pada famili gen putatif *HbPIs*. Analisis pohon filogenetik dari tanaman *Arabidopsis thaliana* dan *H. brasiliensis* sebanyak 50 dan 36 sekuen residu asam amino dari 32 scaffold yang mengandung gen putatif PI menunjukkan adanya tiga kluster besar, yaitu famili LTP-I, SERPIN dan LTP-II. LTP-I terdiri dari 23 gen putatif *HbPI* (*HbPI05* hingga *HbPI27*) dan 12 gen *AtPI*. SERPIN yang merupakan anggota kelas protease inhibitor serin terdiri dari 11 gen putatif *HbPI* (*HbPI01* hingga *HbPI04* dan *HbPI28* hingga *HbPI34*) dan 22 gen *AtPI*. LTP-II terdiri dari 2 gen putatif *HbPI* (*HbPI35* hingga *HbPI36*) dan 16 gen *AtPIs*. Penelitian ini menghasilkan informasi penting untuk melakukan karakterisasi fungsional lebih mendalam pada gen *HbPI* tanaman karet ke depannya.

[Kata kunci: protease inhibitor, genome-wide, scaffold, in silico, Hevea brasiliensis].

^{*)} Penulis korespondensi: irfan.martiansyah@iribb.org ; rizaputranto@iribb.org

Pendahuluan

Protease inhibitor (PI) merupakan protein yang berfungsi untuk menghambat aktivitas proteolitik enzim protease. Ukuran molekul PI berkisar antara 3-120 kDa, tetapi kebanyakan PI berukuran 8-22 kDa (Islamov *et al.*, 2012; Malolo *et al.*, 2015; Michaud, 2001). Berdasarkan domain inhibitor, lebih dari 13 famili protease inhibitor seperti SERPIN, Bowman Birk (BBIs), Squash, Kunitz, Cereal tripsin/ α -amilase, Mustard (Sinapis) tripsin (MSI), Potato tipe I PIs (PI1), Potato tipe II PIs (PI2), Cistein PIs (CYS), Stefin (Famili-1 cistatin), Cistatin (Famili-2 cistatin), Kininogen (Famili-3 cistatin), Fitocistatin (Famili-4 cistatin), serta Aspartil dan metalokarboksipeptidase telah berhasil diidentifikasi pada tanaman (Habib & Fazili, 2007; Malolo *et al.*, 2015; Michaud, 2001; Mosolov & Valueva, 2008; Rawlings *et al.*, 2004; Schlüter *et al.*, 2010). Berdasarkan situs reaktif aktivitas proteolitik, PI terbagi menjadi enam kelas utama yaitu serin, sistein, tripsin, aspartat, metalo dan treonin (Habib & Fazili, 2007; Ryan, 1990). PI dengan aktivitas penghambat serin, sistein dan tripsin merupakan yang paling dominan dan melimpah di dalam sel (Mosolov & Valueva, 2008).

Protease inhibitor disinyalir sebagai produk yang memiliki nilai ekonomi yang tinggi dalam sektor pertanian dan kesehatan dikarenakan memiliki kemampuan anti-mikroba dan anti-virus (Bijina *et al.*, 2011; Islamov *et al.*, 2012; Sack *et al.*, 2015). PI bekerja dengan cara menghambat, menghidrolisis, mengubah atau menonaktifkan protease yang diekskresikan oleh patogen sehingga menjadi avirulen (Ryan, 1990). Pada tanaman tebu transgenik, PI diketahui dapat menekan pertumbuhan larva serangga penggerek batang (*Scirpophaga excerptalis*). Sedangkan pada tembakau transgenik yang mengandung gen PI dapat menjadi resisten terhadap serangga hama dan patogen penyebab penyakit tanaman (Christy *et al.*, 2009; Senthilkumar *et al.*, 2010). Putranto *et al.* (2016) telah berhasil mengisolasi enzim PI pada lateks tanaman karet dan menunjukkan adanya aktivitas penghambatan pertumbuhan cendawan patogen tanaman kelapa sawit (*Ganoderma boninense*) dan penyebab penyakit jamur akar putih pada tanaman karet (*Rigidoporus lignosus*) skala laboratorium. Selain itu, PI yang dipurifikasi dari suspensi sel kalus karet, dilaporkan dapat menekan perkembangan miselium dan zoospora *Phytophthora palmivora* (Bunyatang *et al.*, 2016; Bunyatang *et al.*, 2013).

Tanaman karet (*Hevea brasiliensis* Muell. Arg) adalah komoditas perkebunan yang memiliki nilai ekonomi tinggi karena merupakan sumber utama karet alam dunia yang sudah komersial saat ini. Karet alam diproduksi oleh tanaman karet dalam

bentuk getah yang disebut lateks (D'Auzac & Jacob, 1989). Lateks merupakan sitoplasma dari jaringan sel kompleks bernama latisifer yang didalamnya terkandung tidak hanya partikel karet namun juga fraksi lain seperti protein pertahanan (Aoki *et al.*, 2014; De Fay & Jacob, 1989). Salah satu dari protein pertahanan tersebut adalah protease inhibitor. Dalam penelitian pendahuluan, Putranto *et al.*, (2016) telah mengidentifikasi, mengisolasi, dan mengkarakterisasi empat protein putatif PI (HbPI01-HbPI04) dari serum B lateks yang termasuk ke dalam famili protease inhibitor SERPIN. Meskipun demikian, famili-famili gen penyandi PI di tanaman karet belum banyak diketahui. Di sisi lain, identifikasi famili gen PI di tingkat laboratorium (*wet lab*) memerlukan waktu yang lama dan tenaga yang tidak sedikit. Oleh karena itu, pendekatan *in silico* melalui bioinformatika dapat digunakan untuk mempercepat analisis dalam prediksi famili gen PI (Penders *et al.*, 2008). Selain itu, informasi genom karet dari klon RRIM 600 dan Reyan 7-33-97 telah tersedia dan dapat diakses secara daring pada basis data European Nucleotide Archive (ENA) (Lau *et al.*, 2016; Rahman *et al.*, 2013; Tang *et al.*, 2016).

Studi komparatif genom merupakan salah satu pendekatan molekuler yang dapat digunakan untuk mengidentifikasi domain, motif dan famili gen dalam suatu genom organisme. Penjajaran sekuen dan analisis filogenetik yang menjadi bagian tak terpisahkan dari studi ini mampu membandingkan dua atau lebih organisme yang belum diketahui informasinya dengan suatu organisme model (Bush & Moore, 2012; Edgar, 2004). Identifikasi gen spesifik dan fungsinya membantu kita dalam memahami perkembangan evolusi molekuler gen dan protein serta lebih jauh lagi dapat mengenali perbedaan morfologi ataupun fenotipe tanaman (Sharma & Suresh, 2015). Analisis *in silico* komparatif *A. thaliana* dan *H. brasiliensis* telah berhasil mengidentifikasi beberapa famili gen penting seperti superfamili APETALA2/ETHYLENE RESPONSE FACTORS (AP2/ERF) (Duan *et al.*, 2010; Putranto & Montoro, 2016; Putranto *et al.*, 2015a; Putranto *et al.*, 2015b) famili *Calcium-Dependent Protein Kinase* (CDPK) (Xiao *et al.*, 2017), famili reseptor asam absisat (ABA) (Guo *et al.*, 2017), famili *Glucose-6-Phosphate Dehydrogenase* (G6PDH) (Long *et al.*, 2016), famili *Metacaspase* (Liu *et al.*, 2016), famili *Aquaporin* (Zou *et al.*, 2015), dan famili *ABC Transporter* (Zhiyi *et al.*, 2015). Di sisi lain, studi komparatif sejenis juga telah berhasil dilakukan pada spesies lain seperti *Oryza vs Shorgum* untuk identifikasi dan analisis filogenetik gen *transporter Ca²⁺* (Goel *et al.*, 2011) dan gen *siDof* pada tanaman millet (*Setaria italica*) (Zhang *et al.*, 2017). Baru-baru ini, penelitian menggunakan

studi komparatif juga dilakukan pada tanaman kentang (*Solanum tuberosum*) untuk mengidentifikasi gen putatif yang menyandi domain *SQUAMOSA promoter-binding protein* (faktor transkripsi spesifik yang berperan dalam pembungaan) dengan *A. thaliana* sebagai tanaman pembandingnya (Kavas *et al.*, 2017).

Arabidopsis thaliana merupakan tanaman model penting yang digunakan sebagai pembanding dalam identifikasi gen dan menentukan fungsi suatu gen. Tanaman *Arabidopsis* seringkali digunakan dalam analisis genomik karena memiliki berbagai keunggulan yaitu siklus hidup yang pendek (6 minggu), ukuran kecil, anakannya banyak dan memiliki genom inti yang relatif kecil (125 Mb). Seluruh genom *Arabidopsis* telah berhasil disekuensing (selama 1996-2000), termasuk anotasi prediksi gen fungsional sehingga banyak proses biologi seperti pertahanan, fotomorfogenesis, regulasi gen, perkembangan, metabolisme, transportasi, reparasi DNA berhasil dipelajari dengan relatif mudah dan terarah (Initiative, 2000). Genom *A. thaliana* mengandung lebih dari 41 gen yang berpotensi menyandi protease inhibitor, 8 diantaranya telah teridentifikasi mewakili 68 protease inhibitor yang sudah diketahui pada berbagai organisme (MEROPS peptidase database; <http://merops.sanger.ac.uk/>) (Rawlings *et al.*, 2004).

Penelitian ini bertujuan untuk mengidentifikasi famili gen putatif penyandi PI dengan pendekatan *in silico* komparatif pada genom *H. brasiliensis*. Genom pembanding yang dipilih adalah *A. thaliana*. Dalam studi komparatif ini, *scaffold* genom karet diunduh pada basis data publik. Identifikasi famili dan domain terkonservasi (*conserved domain*) yang terdapat pada CDS putatif teranotasi dilakukan secara daring melalui basis data NCBI. Analisis penjajaran sekuen jamak telah memprediksi tujuh motif potensial terkait aktivitas *HbPIs*. Tiga kluster famili gen penyandi PI telah berhasil diidentifikasi antara *A. thaliana* dan *H. brasiliensis* menggunakan analisis filogenetik.

Bahan dan Metode

Pengunduhan dan penyimpanan basis data genom utuh tanaman karet

Sekuen genom tanaman karet (*Hevea brasiliensis* Muell. Arg) klon Reyan 7-33-97 dengan kode akses LVXX01000000 dan SRP069104 secara berurutan diunduh dari basis data *European Nucleotide Archive* (www.ebi.ac.uk/ena) dari EMBL-EBI (Britania Raya). Panjang genom yang diunduh adalah 1,46 GB dengan jumlah *scaffold* sebanyak 7.453 yang meliputi panjang daerah genom 1,37 GB. Sejumlah 84.285 *contig* menghasilkan 43.792 gen

(Tang *et al.*, 2016). Sekuen genom tersebut disimpan dalam sistem basis data Galaxy pada *platform* bioinformatika SouthGreen, Prancis (<http://galaxy.southgreen.fr/galaxy/>). Dalam penelitian ini, istilah *contig* dan *scaffold* digunakan untuk membedakan secara berurutan sekuen nukleotida pendek (± 300 kb) dan panjang ($\pm 6,4$ Mb).

Identifikasi dan sortasi scaffold genom tanaman karet

Keseluruhan 7.453 *scaffold* genom tanaman karet yang telah dianotasi disortir berdasarkan tiga kata kunci yaitu “protease”, “inhibitor”, dan “proteinase”. Penelusuran basis data nukleotida menggunakan *query* berupa residu asam amino dilakukan pada *scaffold* terpilih via NCBI BLAST+ *tblastn* menggunakan *platform* Galaxy (Cock *et al.*, 2015). Lima puluh residu asam amino yang menyandi protease inhibitor dari *A. thaliana* diunduh dari basis data TAIR v.10 (<http://www.arabidopsis.org/>). Sekuen tersebut digunakan sebagai famili PI komparatif terhadap *H. brasiliensis*. Hasil analisis *tblastn* disusun dalam bentuk tabel Excel.

Anotasi manual sekuen scaffold dan ekstraksi sekuen putatif penyandi PI dari tanaman karet

Sekuen *scaffold* dari genom tanaman karet yang telah teridentifikasi kemudian dipisahkan dan disimpan. Anotasi manual sekuen *scaffold* meliputi daerah 5'-UTR, *coding sequence* (CDS), 3'-UTR dilakukan menggunakan program Genious 5.3.6 (Biomatters Ltd, USA) (Kearse *et al.*, 2012). Daerah CDS putatif penyandi protease inhibitor diekstrak. Sebagai konfirmasi hasil anotasi, analisis MEGA-BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) dilakukan pada sekuen CDS tersebut dengan parameter basis data *nucleotide collection* (nr/nt). Sekuen CDS putatif penyandi PI kemudian dianotasi kembali dengan nama gen dan spesies yang memiliki persentase kemiripan > 80 %, lalu sekuen tersebut diterjemahkan menjadi residu asam amino.

Identifikasi famili dan domain terkonservasi (conserved domain)

Identifikasi famili dan domain terkonservasi (*conserved domain*) yang terdapat pada CDS putatif teranotasi dilakukan secara daring melalui NCBI CDD (www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Sekuen residu asam amino dari CDS dimasukkan ke dalam kolom pencarian CDD dan hasil yang diperoleh disimpan dalam format PDF. Penentuan domain dilakukan dengan mencocokkan sekuen asam amino atau protein yang telah diperoleh terhadap sekuen nukleotida CDS putatif teranotasi. Domain terkonservasi

yang diperoleh kemudian dianotasi sesuai dengan nama famili dan nomor aksesinya. Sekuen tersebut kemudian diterjemahkan menjadi residu asam amino (Putranto *et al.*, 2015a). Data rekapitulasi disimpan dalam bentuk tabel. Motif terkonservasi dalam residu asam amino putatif PI pada tanaman karet diperoleh dengan analisis penjajaran sekuen jamak (*multiple sequence alignment*) dengan algoritma MUSCLE (Edgar, 2004).

Analisis filogenetik komparatif antara famili gen penyandi protease inhibitor dari A. thaliana dan H. brasiliensis

Workflow analisis filogenetik dilakukan melalui dua langkah yaitu (1) penjajaran (*alignment*) sekuen asam amino dan (2) konstruksi pohon filogenetik untuk menentukan ortologi dan paralogi dari gen. Penjajaran sekuen asam amino putatif PI dari tanaman model *A. thaliana* dan *H. brasiliensis* dieksekusi menggunakan algoritma penjajaran sekuen jamak MUSCLE dalam program MEGA 6.0. Parameter algoritma yang digunakan adalah 32 maksimum iterasi dengan pengukuran jarak kmer 4_6, metode pengklasteran UPGMB. Sekuen disejajarkan berdasarkan tingkat kesamaan 65%. Pohon filogenetik dirakit menggunakan algoritma statistik *Neighbor-Joining* pada program MEGA 6.0 (Senthikumar *et al.*, 2010). Bootstrap untuk pohon konsensus dibuat sebanyak 1000 ulangan. Jarak evolusi dihitung menggunakan metode berbasis matriks JTT (John-Taylor-Thornton). Perlakuan *missing data* (*gap*) menggunakan *pairwise deletion*.

Hasil dan Pembahasan

Pra-perlakuan sekuen genom utuh dan sortasi scaffold pada tanaman karet

Genom tanaman karet (*H. brasiliensis*) klon Reyan 7-33-97 telah berhasil dipetakan oleh Tang *et al.* (2016). Perakitan genom tanaman karet sebesar 1,37 Gb melingkupi 93,8% dari total ukuran genom karet (1,46 Gb), diketahui memiliki jumlah *scaffold* sebanyak 7453 *scaffold* (N50 *scaffold* = 1,28 Mb), jumlah *contig* sebanyak 84.285 *contig* (N50 *contig* = 30,6 kb), dan jumlah gen sebanyak 43.792 gen.

Diagram alir kegiatan analisis *in silico* disajikan pada Gambar 1. *Scaffold* genom tanaman karet sebanyak 7.453 *scaffold* disortir berdasarkan tiga kata kunci. Metode sortasi dengan kata kunci tersebut dipilih karena relatif mudah dan sederhana. Dari hasil sortasi tersebut, sebanyak 108, 46, dan 31 *scaffold* diperoleh menggunakan kata kunci “protease”, “inhibitor” dan “proteinase”, secara berurutan. Empat puluh enam *scaffold* dari kata kunci “inhibitor”

diekstrak dari basis data genom tanaman karet klon Reyan 7-33-97. Analisis *tblastn* terhadap 50 residu asam amino penyandi protease inhibitor dari *A. thaliana* dilakukan. Sebagai tanaman model dalam biologi molekuler, basis data genom *A. thaliana* dipilih karena telah 100% teranotasi serta fungsi famili gennya telah diketahui (Putranto *et al.*, 2015a). Hasil *tblastn* menunjukkan sebanyak 33 *scaffold* memiliki kemiripan yang tinggi (80-99%) untuk sekuen nukleotida maupun asam amino terhadap *A. thaliana*.

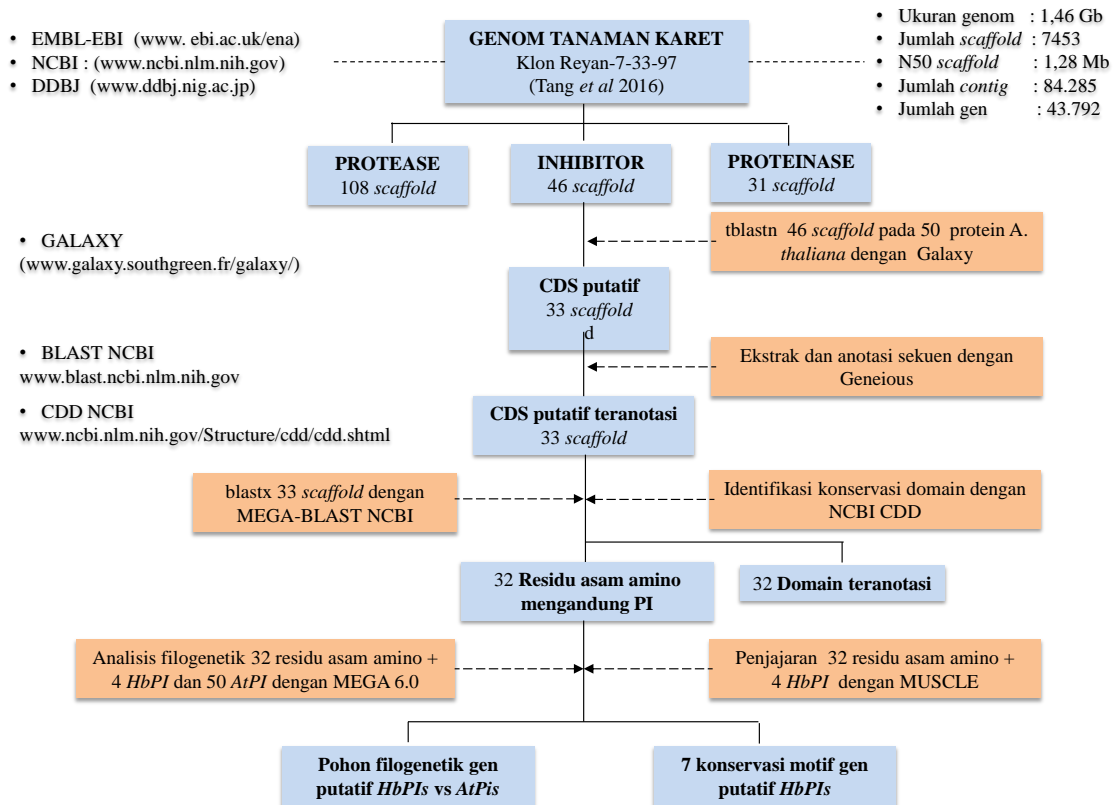
Anotasi manual sekuen scaffold dan ekstraksi sekuen putatif penyandi PI dari tanaman karet

Sekuen *scaffold* yang panjang ($\pm 6,4$ Mb) berpotensi memuat lebih dari 20 gen di dalamnya. Secara *in silico*, pendugaan gen yang dituju dalam suatu *scaffold* membutuhkan ketelitian dan pengalaman untuk menganalisis secara tepat bagian CDS yang menyandikan gen putatif PI. Setelah menemukan CDS yang tepat, dilakukan pencarian kodon *start* (AUG) dan kodon *stop* (UAA,UAG,UGA) lalu sekuen lengkap tersebut diekstrak dan dianotasi sementara menggunakan program Geneious. Hasil anotasi sementara terhadap 33 sekuen nukleotida menunjukkan adanya perbedaan sekuen *full-length* dan parsial dari bagian gen putatif PI yaitu 27 sekuen *full-length* dan 5 sekuen parsial. Perbedaan tersebut dapat dihasilkan karena sekuen *scaffold* berpotensi mengandung intron dan bagian sekuen *non-coding* lainnya (Morello & Breviaro, 2008, Hong *et al.*, 2006). Lima sekuen parsial yang mengandung intron secara *in silico* yaitu *scaffold* LVXX01000048 (*HbPI28*), LVXX01000103 (*HbPI18*), LVXX01000276 (*HbPI34*), LVXX01000440 (*HbPI16*) dan LVXX01000574 (*HbPI13*).

Hasil MEGA-BLAST pada 33 sekuen nukleotida tersebut menunjukkan 32 diantaranya memiliki tingkat kemiripan sekuen yang tinggi, bahkan sebanyak 22 sekuen memiliki skor kesamaan basa nukleotida lebih dari 200 nukleotida (garis merah) (Gambar 2). Sekuen LVXX01000246 merupakan satu-satunya sekuen yang memiliki skor kesamaan basa nukleotida kurang dari 80 nukleotida (garis hijau). Hal ini menunjukkan bahwa sekuen tersebut memiliki tingkat kemiripan yang rendah dibandingkan dengan sekuen gen pada organisme lain yang terdapat dalam basis data NCBI sehingga tidak dilanjutkan untuk proses translasi menjadi asam amino.

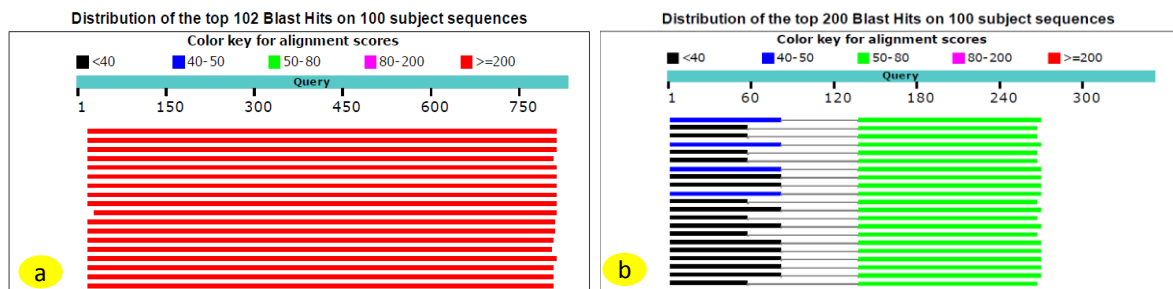
Identifikasi domain dan famili

Domain adalah suatu daerah dari protein yang secara struktural umumnya terdiri dari 25-350 asam amino. Protein sederhana rata-rata hanya memiliki satu domain. Pada protein yang lebih



Gambar 1. Diagram alir analisis *in silico* komparatif untuk identifikasi famili gen putatif penyandi protease inhibitor pada genom *Hevea brasiliensis*. Piranti lunak seperti EMBL-EBI, NCBI, DDBJ, Galaxy, Geneious, MEGA 6.0 dan algoritma MUSCLE untuk penjajaran sekuen digunakan dalam analisis.

Figure 1. Scheme of *in silico* comparative analysis for the identification of putative gene family encoding protease inhibitors in *Hevea brasiliensis* genome. EMBL-EBI, NCBI, DDBJ, Galaxy, Geneious, MEGA 6.0 and MUSCLE algorithm were used in this analysis.



Gambar 2. Profil pencarian MEGA-BLAST dari sekuen nukleotida (a) LVXX01000048 (*HbPI28*) tingkat kemiripan sekuen tinggi dan (b) LVXX010000246 tingkat kemiripan sekuen rendah. Analisis dilakukan di basis data NCBI. Warna merah menunjukkan tingkat kemiripan tinggi berbanding terbalik dengan warna hijau dengan tingkat kemiripan rendah.

Figure 2. Profile of MEGA-BLAST search from nucleotide sequence (a) with high similarity such as LVXX01000048 (*HbPI28*) and (b) with low similarity such as LVXX010000246. The analysis was carried out in NCBI. Red strip showed high similarity sequences as opposed to green strip for low similarity sequences.

kompleks, beberapa domain dapat terkandung di dalamnya. Domain bertanggung jawab terhadap aktivitas protein dan biasanya memiliki fungsi yang spesifik (Levitt, 2009). Pemanfaatan basis data dapat digunakan untuk melacak keberadaan suatu motif dan domain dari protein.

Protein yang berbeda dapat memiliki motif dan domain yang sama sehingga digolongkan dalam satu famili yang sama (Cherukuri, 2007). Famili dari suatu protein pada organisme dapat mengalami spesiasi (ortolog) maupun duplikasi(paralog) selama proses evolusi. Struk-

tur, sekuen, dan fungsi protein dapat diketahui dengan mengklasifikasikan famili gen yang mengekspresikannya sehingga dengan mudah dipahami hubungan kekerabatan sekaligus mekanisme molekuler yang terkait dengan fungsi tersebut (Das et al., 2015). Domain dari suatu protease inhibitor merupakan segmen dari sekuen asam amino yang berisi sebuah situs reaktif setelah melepaskan bagian tertentu yang tidak terlibat langsung dalam aktivitas inhibitor. Hasil identifikasi domain dan famili gen menunjukkan bahwa satu *scaffold* dapat memiliki lebih dari satu famili dan domain.

Empat puluh gen putatif penyandi PI yang terkandung di dalam 32 *scaffold* telah berhasil diidentifikasi (Tabel 1). Sebagian besar *scaffold* hanya memiliki satu domain, tetapi empat di antaranya memiliki lebih dari satu domain, yaitu *scaffold* LVXX01000019 (domain PKc dan S_TKc), LVXX01000079 (domain *Hydrophob_seed* dan ROM1), LVXX01000238 (domain PKc dan S_TKc), LVXX01000249 (domain S_TKc dan STKc_CAMK). Sebanyak 40 domain telah berhasil diidentifikasi secara *in silico*. Domain protein kinase (PKc) ditemukan sebanyak 24 kali dengan kisaran panjang 60-321 asam amino.

Domain PKc berukuran sekitar 30 kDa dan memiliki panjang domain residu sebanyak 250-350 asam amino. Lebih dari 1000 PKc berhasil diidentifikasi pada saat memetakan gen-gen dari genom lengkap *A. thaliana* (Initiative, 2000). Selama 20 tahun terakhir, perkembangan fungsi PKc yang dipelajari berperan dalam proses induksi sinyal, seperti adanya cekaman biotik, kualitas dan kuantitas cahaya, hormon, dan kondisi lingkungan yang berbeda-beda. Selain itu, regulasi jam biologis tanaman, regulasi siklus sel, perkembangan, aktivitas transportasi dan metabolisme selular (Lehti-Shiu & Shiu, 2012)

Dua domain SERPIN, lima domain *serine-threonine kinase* (S_TKc), satu domain *non-specific lipid transfer protein* (nsLTP), satu domain *hydropob seed*, satu domain ROM1 dan satu domain *Thaumatococcus* juga berhasil diidentifikasi (Tabel 1). SERPIN merupakan domain terbesar dan terbanyak dari protease inhibitor. SERPIN berukuran 39-43 kDa dan memiliki bentuk *loop* fleksibel serta kovalen kompleks protease ireversibel. SERPIN memiliki panjang 340-440 asam amino relatif lebih banyak dari pada KunitzP dan potato inhibitor 1 (30-200 asam amino) yang memiliki bentuk *loop* kaku dan kompleks non-kovalen dengan protease (Christeller & Laing, 2005). SERPIN merupakan protease inhibitor spesifik yang berinteraksi dan berikatan dengan lambat dengan substrat atau protease target serta dapat membentuk *loop* dalam konformasi kanonikal (Fluhr et al., 2012). Secara

umum, keanekaragaman dari domain yang dikandung oleh keempat puluh gen putatif penyandi *HbPI* menunjukkan fungsi potensial yang berbeda pada tanaman karet.

Penjajaran sekuen asam amino putatif *HbPIs*

Untuk mencari bentuk motif dari gen putatif PI dalam tanaman karet, dilakukan penjajaran 36 sekuen residu asam amino yang telah teridentifikasi sebelumnya menggunakan algoritma MUSCLE (Gambar 3). Penjajaran 36 sekuen residu asam amino dari 32 *scaffold* yang mengandung gen putatif *HbPI* menghasilkan tujuh motif yang memiliki tingkat kemiripan homologi yang tinggi. Motif I dan II memiliki homologi yang sama untuk 27 gen putatif *HbPI* yang terkonservasi di dalam kedua motif tersebut. Hampir semua asam amino pada motif III dan IV terdapat pada seluruh gen putatif *HbPI* dan menunjukkan kesamaan homologi yang tinggi serta asam amino terkonservasi pada kedua motif ini.

Tabel 2 memperlihatkan sekuen konsensus dari motif asam amino yang ditemukan dalam famili gen penyandi *HbPIs*. Motif I merupakan motif dengan konsensus sekuen terpanjang. Motif I dan II memiliki jumlah residu asam amino masing-masing 88 dan 62 asam amino. Konsensus kedua motif tersebut diawali dengan asam amino yang sama sekaligus didominasi oleh asam amino prolin. Asam amino prolin termasuk dalam asam amino berukuran kecil dan bersifat alifatik serta hidrofobik. Sifat tersebut menunjukkan bahwa prolin ditemukan di permukaan dari struktur protein (Betts & Russell, 2003). Motif III dan IV memiliki jumlah residu sebanyak 36 dan 37 asam amino. Secara umum, kedua motif tersebut dipenuhi oleh asam amino hidrofobik seperti A, G, T, F, Y, W, I, V, L, M. Hal tersebut memprediksi bentuk dari protease inhibitor serta aktivitasnya pada permukaan protein. Motif V dan VII memiliki jumlah residu asam amino berturut-turut 42 dan 87 asam amino, sedangkan motif VI memiliki jumlah residu asam amino terpendek yaitu 14 asam amino.

Lipid transfer protein (LTP) merupakan kelompok protein yang berukuran sekitar 7-10 kDa yang dapat ditemukan di jaringan tanaman (Edstam et al., 2014). Secara struktural, LTP dibagi menjadi dua yaitu LTP-I berukuran 9-10 kDa dengan jumlah asam amino sebanyak 90-95 asam amino dan LTP-II berukuran 7 kDa (65-75 asam amino) (Finkina et al., 2016 ; Kader, 1996). LTP menarik untuk dipelajari karena memiliki tiga fungsi utama yaitu kemampuan untuk mengikat dan mentransfer lipid, protein pertahanan endogen untuk imunitas tanaman, dan alergen tanaman (Finkina et al., 2016). LTP dapat ditemukan secara luas di

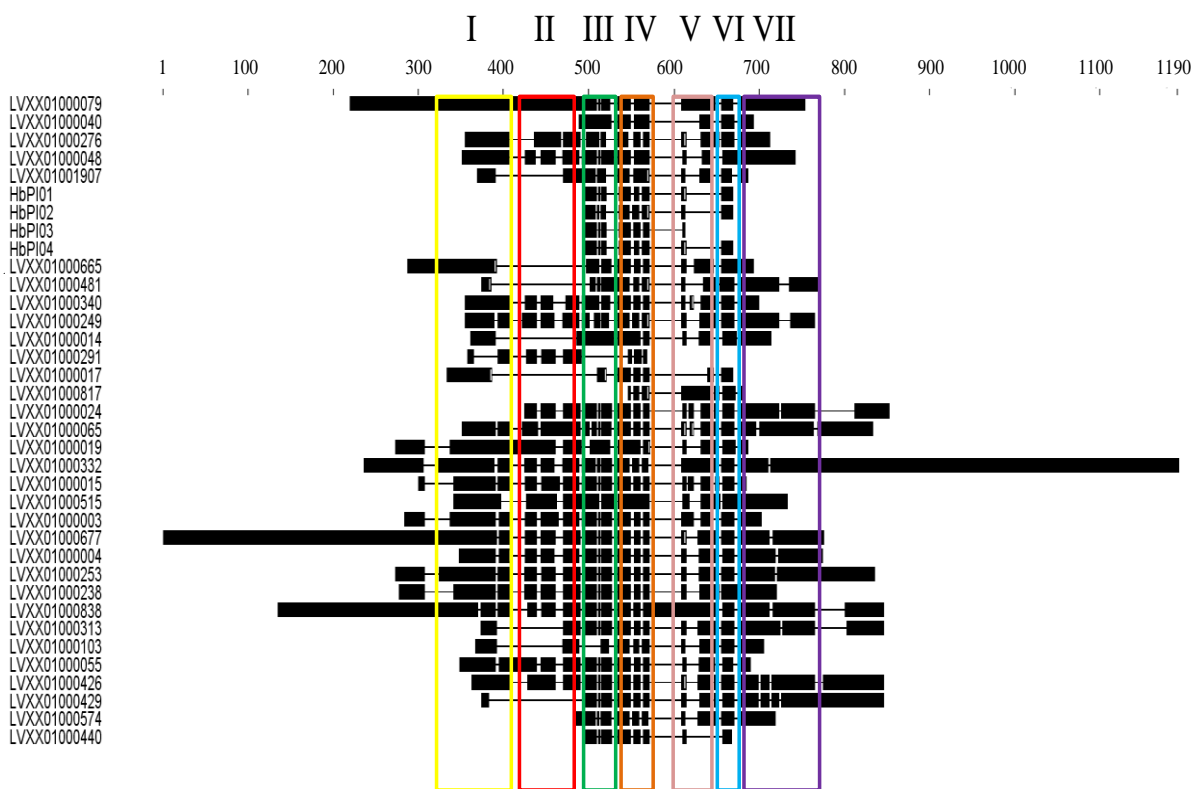
Tabel 1. Hasil identifikasi 40 gen putatif penyandi PI dengan analisis *in silico* komparatif pada 32 *scaffold* genom tanaman karet (*H. brasiliensis*). Identifikasi domain protein dilakukan pada NCBI *Conserved Domains Database* (CDD).

Table 1. Identification of 40 putative encoding PI genes using *in silico* comparative analysis of 32 genomic scaffold of *H. brasiliensis*. The identification of protein domain was carried out using NCBI *Conserved Domains Database* (CDD).

Scaffold	Nama Putatif /Putative name	Basis data / Database NCBI CDD			Referensi/References
		Nama domain/ Domain name	Akresi/ Accession	Panjang/ Length (aa)	
LVXX01000481	<i>HbPI-01</i>	1 PKc	cl21453	172	(Putranto <i>et al.</i> , 2016)
LVXX01000481	<i>HbPI-02</i>	1 PKc	cl21453	172	(Putranto <i>et al.</i> , 2016)
LVXX01000481	<i>HbPI-03</i>	1 PKc	cl21453	172	(Putranto <i>et al.</i> , 2016)
LVXX01000481	<i>HbPI-04</i>	1 PKc	cl21453	172	(Putranto <i>et al.</i> , 2016)
LVXX01000004	<i>HbPI-05</i>	1 STKc_IRAK	cd14066	270	(Venturini <i>et al.</i> , 2013)
LVXX01000253	<i>HbPI-06</i>	1 PKc	cl21453	272	(Venturini <i>et al.</i> , 2013)
LVXX01000426	<i>HbPI-07</i>	1 PKc	cl21453	208	(Pan <i>et al.</i> , 2014)
LVXX01000429	<i>HbPI-08</i>	1 PKc	cl21453	178	(Li <i>et al.</i> , 2010)
LVXX01000677	<i>HbPI-09</i>	1 PKc	cl21453	269	(Li <i>et al.</i> , 2010)
LVXX01000313	<i>HbPI-10</i>	1 PKc	cl21453	204	(Li <i>et al.</i> , 2010)
LVXX01000838	<i>HbPI-11</i>	1 PKc	cl21453	321	(Rivarola <i>et al.</i> , 2011)
LVXX01000332	<i>HbPI-12</i>	1 PKc	cl21453	286	(Li <i>et al.</i> , 2010)
LVXX01000574	<i>HbPI-13</i>	1 PKc	cl21453	127	(Tuskan <i>et al.</i> , 2006)
LVXX01000055	<i>HbPI-14</i>	1 PKc	cl21453	182	(Li <i>et al.</i> , 2010)
LVXX01000238	<i>HbPI-15a</i>	1 PKc	cl21453	168	(Li <i>et al.</i> , 2010)
	<i>HbPI-15b</i>	2 S_TKc	smart00220	176	
LVXX01000440	<i>HbPI-16</i>	1 PKc	cl21453	66	(Zeng <i>et al.</i> , 2014)
LVXX01000065	<i>HbPI-17</i>	1 PKc	cl21453	278	(Li <i>et al.</i> , 2010)
LVXX01000103	<i>HbPI-18</i>	1 PKc	cl21453	92	(Li <i>et al.</i> , 2010)
LVXX01000024	<i>HbPI-19</i>	1 PKc	cl21453	236	(Li <i>et al.</i> , 2010)
LVXX01000003	<i>HbPI-20</i>	1 PKc	cl21453	181	(Li <i>et al.</i> , 2010)
LVXX01000015	<i>HbPI-21</i>	1 PKc	cl21453	180	(Li <i>et al.</i> , 2010)
LVXX01000249	<i>HbPI-22a</i>	1 S_TKc	smart00220	253	(Li <i>et al.</i> , 2010)
	<i>HbPI-22b</i>	2 STKc_CAMK	cd05117	255	
LVXX01000340	<i>HbPI-23</i>	1 PKc	cl21453	221	(Zeng <i>et al.</i> , 2014)
LVXX01000019	<i>HbPI-24a</i>	1 PKc	cl21453	219	(Li <i>et al.</i> , 2010)
	<i>HbPI-24b</i>	2 S_TKc	smart00220	219	
LVXX01000515	<i>HbPI-25</i>	1 PKc	cl21453	92	(Li <i>et al.</i> , 2010)
LVXX01000665	<i>HbPI-26</i>	1 PKc	cl21453	126	(Li <i>et al.</i> , 2010)
LVXX01000017	<i>HbPI-27</i>	1 PKc	cl21453	60	(Zhu <i>et al.</i> , 2015)
LVXX01000048	<i>HbPI-28</i>	1 SERPIN	cl00137	79	(Pan <i>et al.</i> , 2014)
LVXX01000291	<i>HbPI-29</i>	1 PKc	cl21453	69	(Li <i>et al.</i> , 2010)
LVXX01001907	<i>HbPI-30</i>	1 SERPIN	cl00137	132	(Li <i>et al.</i> , 2010)
LVXX01000014	<i>HbPI-31</i>	1 PKc	cl21453	90	(Li <i>et al.</i> , 2010)
LVXX01000481	<i>HbPI-32</i>	1 PKc	cl21453	172	(Zeng <i>et al.</i> , 2014)
LVXX01000817	<i>HbPI-33</i>	1 PKc	cl21453	72	(Zhang & Gao, 2016)
LVXX01000276	<i>HbPI-34</i>	1 Thaumatin	pfam00314	218	(Li <i>et al.</i> , 2010)
LVXX01000040	<i>HbPI-35</i>	1 nsLTP1	cd01960	79	(Rivarola <i>et al.</i> , 2011)
LVXX01000079	<i>HbPI-36a</i>	1 Hydrophob_seed	pfam14547	78	(Zeng <i>et al.</i> , 2014)
	<i>HbPI-36b</i>	2 ROM1	COG5422	79	

berbagai organisme dengan fungsi dan ukuran yang bervariasi (Wang *et al.*, 2012). Chen *et al.*, (2017) melaporkan bahwa LTP terlibat dalam respon cekaman abiotik berupa suhu rendah dan kekeringan pada tanaman tebu. LTP juga berperan dalam toleransi suhu tinggi pada tanaman gandum (Yu *et al.*, 2014). Selain itu, LTP terekspresi lebih banyak ketika terjadi infeksi bakteri patogen pada tanaman barley (Petti *et al.*, 2010). *A. thaliana*

yang terserang penyakit akar dan tercekam salinitas tinggi melibatkan protein LTP sebagai protein pertahanan yang ditunjukkan dengan peningkatan profil ekspresi gen tersebut (Jülke & Ludwig-Müller, 2015). Famili gen LTP banyak ditemukan di berbagai tanaman seperti *A. thaliana* (110 famili), gandum (150 famili) dan padi (52 famili) (Boutrot *et al.*, 2008; Lamesch *et al.*, 2012).



Gambar 3. Penjajaran 36 sekuen residu asam amino dari 32 scaffold yang mengandung gen putatif PI pada genom *Hevea brasiliensis* (klon Reyan 7-33-97). Sekuen dikelompokkan berdasarkan tingkat kemiripannya dengan jarak kmer 4_6. Pembobotan sekuen menggunakan ClustalW. Warna yang berbeda menunjukkan perbedaan motif terkonservasi (I-VII).

Figure 3. Alignment of 36 amino acid residues from 32 scaffolds containing putative PI in *Hevea brasiliensis* genome (clone Reyan 7-33-97). Sequences were grouped according to its similarity using kmer 4_6 distance. Sequence weighing was carried out using ClustalW. Different colors showed different conserved motifs (I-VII).

Tabel 2. Konsensus residu asam amino dari tujuh motif hasil penjajaran sekuen residu asam amino dari 32 scaffold. Sekuen konsensus diperoleh menggunakan program Geneious.

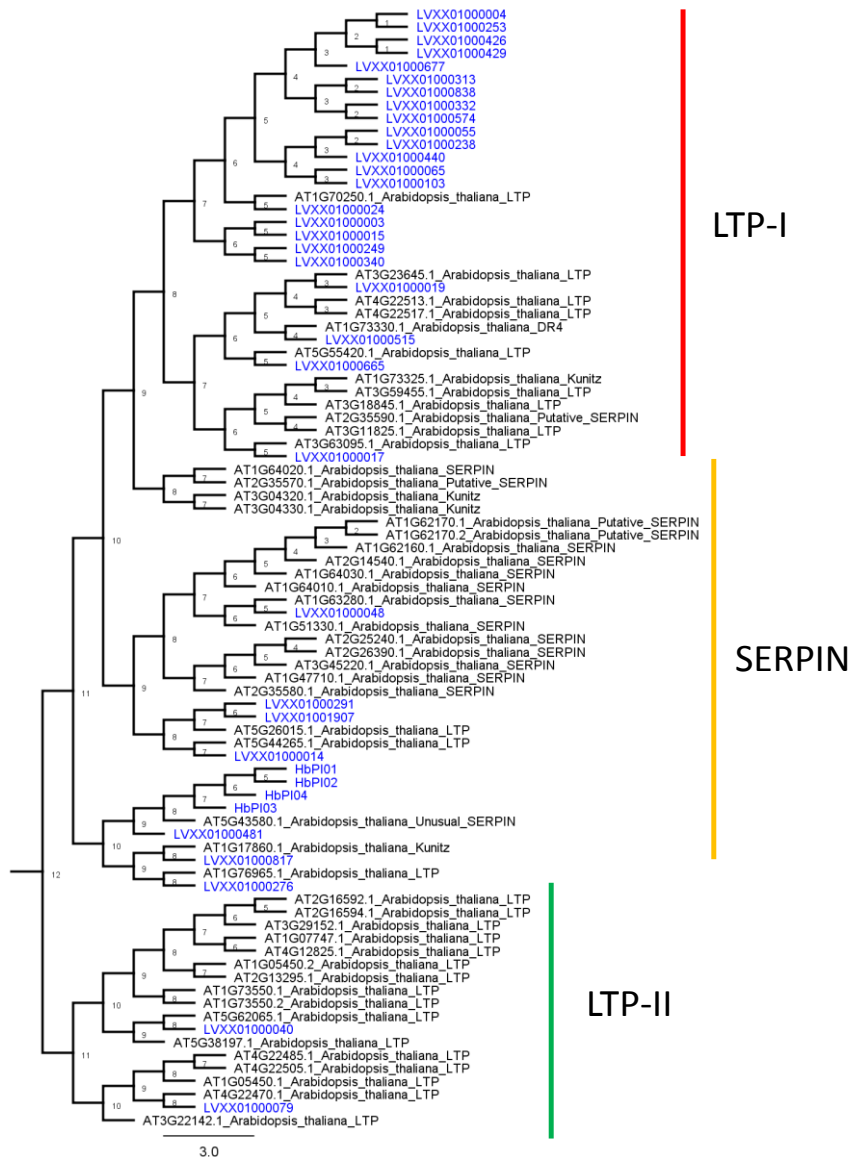
Table 2. Consensus amino acid residues from seven motifs resulted from alignment of 36 amino acid residues from 32 scaffolds. Consensus sequences were obtained using Geneious program.

Motif/ Motifs	Jumlah residu/ Residues number (aa)	Sekuen motif konsensus/ Consensus sequence motifs
Motif I	88	PKPPYVPKPPIVKPPTIPKPPYVPKPPIVKPPTRPKPPHEPKPPITPKPPIVHPPYV PKPPIVKPPPFVPPKPPVVKPPHFPPKPIIFP
Motif II	62	PPYYKPPVTPAPPKPPVSPTPPTLPPKPPVTPTPKPPVTPTPPILPPKPPVTPT PPYFPN
Motif III	36	QKMKIGVLILLSWAATVSTDVECGTVTGLLSVCSTC
Motif IV	37	RKKKRDSTLGSYGNPLLKLSYQSLLKATNGFSLEILI
Motif V	42	EHELEYQTKIPSMRNRQFRLAESTRIGGTIGYLPPEFSQKRS
Motif VI	14	EYTEKSDVYSFGVV
Motif VII	87	LMELVSGRRNSEQSKDGMKEYFPPIRVARLVNKDGLLEDDVLSLLDPFLEGNAN VEELTRICKVACWCIQDDKTLRPSMSDVVYILEGS

Analisis filogenetik dari sekuen asam amino HbPIs dan AtPIs

Dari pohon filogenetik, diperoleh tiga kluster famili besar yaitu *lipid transfer protein* (LTP-I dan LTP-II) serta SERPIN (Gambar 4). LTP-I terdiri dari 23 gen putatif HbPI (*HbPI05-HbPI27*) dan 12 gen AtPI. Tiga gen AtPI diketahui bukan kelompok LTP-I, tetapi ikut tersisipkan dalam kelompok tersebut yaitu AT1G73330.1 (DR4=*drought resistant*), AT1G73325.1 (Kunitz) dan AT2G35590. 1 (*putative SERPIN*) diduga

karena adanya kemiripan sekuen asam amino dengan famili LTP. LVXX01000515 diduga merupakan salah satu gen putatif *drought resistant* karena memiliki hubungan kekerabatan yang lebih dekat dengan AT1G73330.1 yang berada dalam satu cabang yang sama. LTP-II terdiri dari 2 gen putatif HbPI (*HbPI35-HbPI36*) dan 16 gen AtPI.



Gambar 4. Analisis filogenetik 50 dan 36 residu asam amino dari 32 scaffold yang mengandung gen putatif PI dari *Arabidopsis thaliana* dan *Hevea brasiliensis* dilakukan dengan metode *Neighbor-Joining* pada program MEGA 6.0. *Bootstrap* untuk pohon konsensus sebanyak 1000 ulangan. Gen putatif PI dibagi menjadi 3 kluster besar, yaitu LTP-I, SERPIN dan LTP-II.

Figure 4. Phylogenetic analysis of 50 and 36 amino acid residues from 32 scaffolds containing putative PI genes in *Arabidopsis thaliana* and *Hevea brasiliensis* using *Neighbor-Joining* method in MEGA 6.0 program. *Bootstrap* for consensus tree was fixed at 1000 replicates. Putative PI genes were classified in three large clusters (LTP-I, SERPIN and LTP-II.)

Analisis pohon filogenetik dilakukan untuk mengetahui hubungan kekerabatan dan fungsi dari gen putatif yang diperoleh. Pohon filogenetik dibuat dari residu sekuen asam amino dari gen putatif *HbPIs* yang dihasilkan dibandingkan dengan gen PI yang berasal dari *A. thaliana* (*AtPIs*). *A. thaliana* merupakan spesies tanaman yang gen-gennya telah diketahui dan dikarakterisasi secara lengkap sehingga digunakan sebagai pembanding untuk analisis filogenetik. Analisis genom pada *A. thaliana* berhasil menemukan 29 gen SERPIN (Silverman *et al.*, 2001). SERPIN 1 dari *A. thaliana* berperan dalam menghambat protease metacaspase dan mengatur respon kekebalan tanaman sehingga dapat mengontrol jumlah hama dan patogen. SERPIN tidak langsung berinteraksi dengan patogen tetapi melalui jalur kompleks yang terlibat dalam *up-regulating* respon imun inang (Alvarez-Alfageme *et al.*, 2011). Kebanyakan famili dalam kelas serin PI berperan dalam menghambat protease serine, tetapi SERPIN dapat menghambat caspase dan protease sistein papain. Pada tanaman *barley* (*Hordeum vulgare*) dan gandum (*Triticum aestivum*), SERPIN berpotensi untuk menghambat tripsin, kimotripsin, cathepsin G asam glutamik, lisin atau arginin pada situs reaktif yang tumpang tindih (Roberts *et al.*, 2011).

Klaster SERPIN terdiri dari 11 gen putatif *HbPI* (*HbPI01-HbPI04;HbPI28-HbPI34*) dan 22 gen *AtPI*. Dari 22 *AtPI* terdapat empat gen yang diduga bukan famili SERPIN yaitu AT3G04320.1, AT3G04330.1, AT1G17680.1 (Kunitz) dan AT1G76965.1 (LTP). LVXX01000817 diduga lebih dekat ke famili kunitz karena memiliki cabang yang sama dengan AT1G17680.1. *HbPI01-HbPI04* merupakan gen putatif PI yang diperoleh pada penelitian sebelumnya (Putranto *et al.*, 2016). *HbPI01-HbPI04* berada dekat dengan LVXX01000481 dan AT5G43580.1 yang diketahui merupakan famili SERPIN (*unusual*). SERPIN merupakan protein kecil atau domain protein dengan banyak ujung inhibitor yang tersusun dari 29-190 residu asam amino yang berperan dalam menghambat protease serin. SERPIN memiliki paling sedikit 16 famili yang dikelompokkan berdasarkan kesamaan sekuen, topologi dan mekanisme ikatan (Bode & Huber, 1993). PI ini telah berhasil diidentifikasi pada beberapa tanaman, baik aspek molekular maupun fungsionalnya. Ekspresinya diinduksi oleh respon pelukaan atau serangan hama dan patogen. Pada tanaman, inhibitor ini efektif dalam menekan proses metabolisme dalam pencernaan hama serangga pertanian serta ramah lingkungan dan berkelanjutan (Jamal *et al.*, 2013).

Penelitian ini memberikan gambaran identifikasi famili gen potensial penyandi protease inhibitor pada genom *H. brasiliensis* dengan menggunakan analisis komparatif terhadap famili protease inhibitor pada *A. thaliana*. Sayangnya, lokalisasi gen-gen tersebut dalam peta fisik genom tanaman karet belum dapat dilakukan karena pemetaan *scaffold* dalam kromosom belum tersedia. Meskipun demikian, analisis komparatif ini dapat dilanjutkan dengan teknik laboratorium basah untuk validasi ekspresi dari tiap-tiap gen penyandi protease inhibitor pada tanaman karet.

Kesimpulan

Studi komparatif secara *in silico* merupakan pendekatan bioinformatika molekuler yang dapat digunakan untuk mengidentifikasi domain, motif dan famili suatu gen dalam suatu genom melalui informasi yang terdapat di dalam suatu basis data gen. Sebanyak 40 gen putatif penyandi protease inhibitor (*HbPI01-HbPI36*) telah berhasil diidentifikasi dari 7.453 *scaffold* genom utuh tanaman karet klon Reyan 7-33-97. Analisis penajajaran sekuen membagi gen putatif PI menjadi tujuh konservasi motif (Motif I-VIII). Analisis filogenetik yang membandingkan residu asam amino dari 32 *scaffold* pada *H. brasiliensis* dan *A. thaliana* menunjukkan adanya tiga klaster yang terkait pada suatu famili yang diduga gen penyandi PI, yaitu famili LTP-I sebanyak 23 *scaffold* yang mengandung gen putatif *HbPI* (*HbPI05-HbPI27*), famili SERPIN sebanyak 11 *scaffold* yang mengandung gen putatif *HbPI* (*HbPI01-HbPI04;HbPI28-HbPI34*) serta famili LTP-II sebanyak 2 *scaffold* yang mengandung gen putatif *HbPI* (*HbPI35-HbPI36*). Hasil ini memperlihatkan gambaran studi komprehensif dari famili gen penyandi protease inhibitor yang dapat dijadikan sebagai acuan untuk melakukan karakterisasi gen lebih jauh pada tanaman karet.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Prof. Chaorong Tang dan kolega yang telah menyediakan data publik genom *Hevea brasiliensis* Muell. Arg. Ucapan terima kasih juga disampaikan kepada Kementerian Pertanian atas dana penelitian yang diberikan melalui program KKP3N pada tahun 2015.

Daftar Pustaka

Alvarez-Alfageme F, J Maharramov, L Carrillo, S Vandenabeele, D Vercammen, F Van Breusegem & G Smagghe (2011). Potential use of a serpin from *Arabidopsis* for pest control. *PLoS ONE*, 6(5), 1-9

- Aoki Y, S Takahashi, D Takayama, Y Ogata, N Sakurai, H Suzuki, K Asawatreratanakul, D Wititsuwannakul, R Wititsuwannakul & D Shibata (2014). Identification of laticifer-specific genes and their promoter regions from a natural rubber producing plant *Hevea brasiliensis*. *Plant Science* 225,1-8.
- Betts MJ, & RB Russell (2007). Amino-acid properties and consequences of substitutions. *Bioinformatics for geneticists* 2, 311-339.
- Bijina B, S Chellappan, SM Basheer, K Elyas, AH Bahkali & M Chandrasekaran (2011). Protease inhibitor from *Moringa oleifera* leaves: isolation, purification, and characterization. *Process Biochemistry* 46(12), 2291-2300.
- Bode W & R Huber (1993). Natural protein proteinase inhibitors and their interaction with proteinases. In, *EJB Reviews*. Springer. 43-61
- Boutrot F, N Chantret & M-F Gautier (2008). Genome-wide analysis of the rice and *Arabidopsis* non-specific lipid transfer protein (nsLtp) gene families and identification of wheat nsLtp genes by EST data mining. *BMC Genomics* 9(86), 1-19.
- Bunyatang O, N Chirapongsatunkul, P Bangrak, R Henry & N Churngchow (2016). Molecular cloning and characterization of a novel bifunctional α -amylase/subtilisin inhibitor from *Hevea brasiliensis*. *Plant Physiology and Biochemistry* 101, 76-87.
- Bunyatang O, N Chirapongsatunkul & N Churngchow (2013). Purification of a Protease Inhibitor from *Hevea brasiliensis* cell suspension and its effect on the growth of *Phytophthora palmivora*. *Journal of plant biochemistry and biotechnology* 22(2), 185-192.
- Bush WS & JH Moore (2012). Genome-wide association studies. *PLoS Comput Biol* 8(12), 1-11
- Chen Y, J Ma, X Zhang, Y Yang, D Zhou, Q Yu, Y Que, L Xu & J Guo (2017). A Novel Non-specific Lipid Transfer Protein Gene from Sugarcane (NsLTps), Obviously Responded to Abiotic Stresses and Signaling Molecules of SA and MeJA. *Sugar Tech* 19(1), 17-25.
- Cherukuri PF (2007) *Protein Domain Superfamilies: An Evolutionary Perspective Towards a Non-redundant Classification of Protein Domains*. Boston University, Thesis 2007,
- Christeller J & W Laing (2005). Plant serine proteinase inhibitors. *Protein and peptide letters* 12(5), 439-447.
- Christy LA, S Arvinth, M Saravanakumar, M Kanchana, N Mukunthan, J Srikanth, G Thomas & N Subramonian (2009). Engineering sugarcane cultivars with bovine pancreatic trypsin inhibitor (aprotinin) gene for protection against top borer (*Scirpophaga excerptalis* Walker). *Plant cell reports* 28(2), 175-184.
- Cock PJA, JM Chilton, B Grüning, JE Johnson & N Soranzo (2015). NCBI BLAST+ integrated into Galaxy. *BioRxiv* 1-9
- D'Auzac J & J Jacob (1989). The composition of latex from *Hevea brasiliensis* as a laticiferous cytoplasm. In ed. J. J. d'Auzac J, Chrestin H., *Physiology of Rubber Tree Latex*. Boca Raton, Florida: CRC Press Inc. 58-96
- Das S, NL Dawson & CA Orengo (2015). Diversity in protein domain superfamilies. *Current opinion in genetics & development* 35,40-49.
- De Fay E & J Jacob (1989). Anatomical organization of the laticiferous system in the bark. In ed J J d'Auzac J, Chrestin H *Physiology of Rubber Tree Latex*. Boca Raton, Florida: CRC Press Inc. 4-14
- Duan C, M Rio, J Leclercq, F Bonnot, G Oliver & P Montoro (2010). Gene expression pattern in response to wounding, methyl jasmonate and ethylene in the bark of *Hevea brasiliensis*. *Tree Physiology* 30(10), 1349-1359.
- Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32(5), 1792-1797.
- Edstam MM, M Laurila, A Höglund, A Raman, KM Dahlström, TA Salminen, J Edqvist & K Blomqvist (2014). Characterization of the GPI-anchored lipid transfer proteins in the moss *Physcomitrella patens*. *Plant Physiology and Biochemistry* 7, 55-69.
- Finkina E, D Melnikova, I Bogdanov & T Ovchinnikova (2016). Lipid Transfer Proteins As Components of the Plant Innate Immune System: Structure, Functions, and Applications. *Acta naturae* 8(2), 47.
- Fluhr R, N Lampl & TH Roberts (2012). Serpin protease inhibitors in plant biology. *Physiologia Plantarum* 145(1), 95-102.
- Goel A, G Taj, D Pandey, S Gupta & A Kumar (2011). Genome-wide comparative in silico

- analysis of calcium transporters of rice and sorghum. *Genomics, proteomics & bioinformatics* 9(4), 138-150.
- Guo D, Y Zhou, H-L Li, J-H Zhu, Y Wang, X-T Chen & S-Q Peng (2017). Identification and characterization of the abscisic acid (ABA) receptor gene family and its expression in response to hormones in the rubber tree. *Scientific Reports* 1-10.
- Habib H & KM Fazili (2007). Plant protease inhibitors: a defense strategy in plants. *Biotechnology and Molecular Biology Reviews* 2(3), 68-85.
- Hong X, DG Scofield & M Lynch (2006). Intron size, abundance, and distribution within untranslated regions of genes. *Mol Biol Evol* 23(12), 2392-404.
- Initiative TAG (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814), 796-815.
- Islamov R, A Ilin & T Kustova (2012). *Inhibitors of Serine Proteinase-Application in Agriculture and Medicine*. INTECH Open Access Publisher. Available from: <http://www.intechopen.com/books/medicinal-chemistry-and-drugdesign/inhibitors-of-serineproteinase-application-in-agriculture-and-medicine> 103-118
- Jamal F, PK Pandey, D Singh & MY Khan (2013). Serine protease inhibitors in plants: nature's arsenal crafted for insect predators. *Phytochemistry Reviews* 12(1), 1-34.
- Jülke S & J Ludwig-Müller (2015). Response of *Arabidopsis thaliana* roots with altered lipid transfer protein (ltp) gene expression to the clubroot disease and salt stress. *Plants* 5(1), 2.
- Kader JC (1996). Lipid-transfer proteins in plants. *Annual review of plant biology* 47(1), 627-654.
- Kavas M, AK Kızıldoğan & B Abanoz (2017). Comparative genome-wide phylogenetic and expression analysis of *SBP* genes from potato (*Solanum tuberosum*). *Computational Biology and Chemistry* 67 (2017) 131-140
- Kearse M, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, T Thierer, B Ashton, P Meintjes & A Drummond (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12), 1647-1649.
- Lamesch P, TZ Berardini, D Li, D Swarbreck, C Wilks, R Sasidharan, R Muller, K Dreher, DL Alexander & M Garcia-Hernandez (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40(1), 1202-1210.
- Lau NS, Y Makita, M Kawashima, TD Taylor, S Kondo, AS Othman, AC Shu-Chien & M Matsui (2016). The rubber tree genome shows expansion of gene family associated with rubber biosynthesis. *Scientific Report* 6(28594), 1-14.
- Lehti-Shiu MD & S-H Shiu (2012). Diversity, classification and function of the plant protein kinase superfamily. *Phil. Trans. R. Soc. B* 367(1602), 2619-2639.
- Levitt M (2009). Nature of the protein universe. *Proceedings of the National Academy of Sciences* 106(27), 11079-11084.
- Li K, W Zhu, K Zeng, Z Zhang, J Ye, W Ou, S Rehman, B Heuer & S Chen (2010). Proteome characterization of cassava (*Manihot esculenta* Crantz) somatic embryos, plantlets and tuberous roots. *Proteome science* 8(10), 1-12.
- Liu H, Z Deng, J Chen, S Wang, L Hao & D Li (2016). Genome-wide identification and expression analysis of the metacaspase gene family in *Hevea brasiliensis*. *Plant Physiology and Biochemistry* 105, 90-101.
- Long X, B He, Y Fang & C Tang (2016). Identification and characterization of the glucose-6-phosphate dehydrogenase gene family in the para rubber tree, *Hevea brasiliensis*. *Frontiers in plant science* 7(215), 1-14
- Malolo F-AE, AB Nougá, A Kakam, K Franke, L Ngah, O Flausino, EM Mpondo, F Ntie-Kang, JC Ndom & V da Silva Bolzani (2015). Protease-inhibiting, molecular modeling and antimicrobial activities of extracts and constituents from *Helichrysum foetidum* and *Helichrysum mechowianum* (compositae). *Chemistry Central Journal* 9(1), 1-11.
- Michaud D 2001. *Recombinant Protease Inhibitors in Plants*. USA: Landes Bioscience.
- Morello L & D Breviario (2008). Plant spliceosomal introns: not only cut and paste. *Current genomics* 9(4), 227-238.
- Mosolov VV & TA Valueva (2008). Proteinase inhibitors in plant biotechnology: A review. *Applied Biochemistry and Microbiology* 44(3), 233-240.

- Pan B-Z, M-S Chen, J Ni & Z-F Xu (2014). Transcriptome of the inflorescence meristems of the biofuel plant *Jatropha curcas* treated with cytokinin. *BMC Genomics* 15(974), 1-20.
- Penders B, K Horstman & R Vos (2008). Walking the line between lab and computation: The “moist” zone. *BioScience* 58(8), 747-755.
- Petti C, M Khan & F Doohan (2010). Lipid transfer proteins and protease inhibitors as key factors in the priming of barley responses to *Fusarium* head blight disease by a biocontrol strain of *Pseudomonas fluorescens*. *Functional & integrative genomics* 10(4), 619-627.
- Putranto R-A, C Duan, Kuswanhadi, T Chaidamsari, M Rio, P Piyatrakul, E Herlinawati, J Pirrello, F Dessailly, J Leclercq, F Bonnot, C Tang, S Hu & P Montoro (2015a). Ethylene Response Factors Are Controlled by Multiple Harvesting Stresses in *Hevea brasiliensis*. *PLoS ONE* 10(4), 1-26.
- Putranto R-A, E Herlinawati, M Rio, J Leclercq, P Piyatrakul, E Gohet, C Sanier, F Oktavia, J Pirrello, Kuswanhadi & P Montoro (2015b). Involvement of Ethylene in the Latex Metabolism and Tapping Panel Dryness of *Hevea brasiliensis*. *International Journal of Molecular Sciences* 16(8), 17885-17908.
- Putranto RA & P Montoro (2016). The *Hevea brasiliensis* AP2/ERF superfamily: From ethylene signalling to latex harvesting and physiological disease response. *Menara Perkebunan* 84(1), 49-62.
- Putranto RA, Siswanto, AS Mulyatni, A Budiani & R Tistama (2016). Purification, characterization, and bioassay of putative protease inhibitors from *Hevea brasiliensis* latex. *E-Journal Menara Perkebunan* 84(2), 76-87.
- Rahman AYA, A Usharraj, B Misra, G Thottathil, K Jayasekaran, Y Feng, S Hou, SY Ong, FL Ng, LS Lee, HS Tan, MKLM Sakaff, BS Teh, B Khoo, SS Badai, NA Aziz, A Yuryev, B Knudsen, A Dionne-Laporte, N Mchunu, Q Yu, B Langston, TA Freitas, A Young, R Chen, L Wang, N Najimudin, J Saito & M Alam (2013). Draft genome sequence of the rubber tree *Hevea brasiliensis*. *BMC Genomics* 14(1), 1-15.
- Rawlings ND, DP Tolle & AJ Barrett (2004). Evolutionary families of peptidase inhibitors. *Biochemical Journal* 378(3), 705-716.
- Rivarola M, JT Foster, AP Chan, AL Williams, DW Rice, X Liu, A Melake-Berhan, HH Creasy, D Puiu & M Rosovitz (2011). Castor bean organelle genome sequencing and worldwide genetic diversity analysis. *PLoS ONE* 6(7), 1-9.
- Roberts TH, J-W Ahn, N Lampl & R Fluhr (2011). Plants and the study of serpin biology. *Methods Enzymol* 499, 347-366.
- Ryan CA (1990). Protease inhibitors in plants: genes for improving defenses against insects and pathogens. *Annual review of phytopathology* 28(1), 425-449.
- Sack M, A Hofbauer, R Fischer & E Stoger (2015). The increasing value of plant-made proteins. *Current opinion in biotechnology* 32, 163-170.
- Schlüter U, M Benchabane, A Munger, A Kiggundu, J Vorster, M-C Goulet, C Cloutier & D Michaud (2010). Recombinant protease inhibitors for herbivore pest control: a multitrophic perspective. *Journal of experimental botany* 61(15), 4169-4183
- Senthilkumar R, CP Cheng & KW Yeh (2010). Genetically pyramiding protease-inhibitor genes for dual broad-spectrum resistance against insect and phytopathogens in transgenic tobacco. *Plant biotechnology journal* 8(1), 65-75.
- Sharma R & CG Suresh (2015). Genome-wide identification and structure-function studies of proteases and protease inhibitors in *Cicer arietinum* (chickpea). *Computers in Biology and Medicine* 56, 67-81.
- Silverman GA, PI Bird, RW Carrell, PB Coughlin, PG Gettins, JI Irving, DA Lomas, CJ Luke, RW Moyer & PA Pemberton (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins: Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *Journal of Biological Chemistry* 276(36), 33293-33296
- Tang C, M Yang, Y Fang, Y Luo, S Gao, X Xiao, Z An, B Zhou, B Zhang, X Tan, H-Y Yeang, Y Qin, J Yang, Q Lin, H Mei, P Montoro, X Long, J Qi, Y Hua, Z He, M Sun, W Li, X Zeng, H Cheng, Y Liu, J Yang, W Tian, N Zhuang, R Zeng, D Li, P He, Z Li, Z Zou, S Li, C Li, J Wang, D Wei, C-Q Lai, W Luo, J Yu, S Hu & H Huang (2016). The rubber tree genome reveals new insights into rubber production and species adaptation. *Nature Plants* 2, 1-10.

- Tuskan GA, S Difazio, S Jansson, J Bohlmann, I Grigoriev, U Hellsten, N Putnam, S Ralph, S Rombauts, A Salamov, J Schein, L Sterck, A Aerts, RR Bhalariao, RP Bhalariao, D Blaudez, W Boerjan, A Brun, A Brunner, V Busov, M Campbell, J Carlson, M Chalot, J Chapman, GL Chen, D Cooper, PM Coutinho, J Couturier, S Covert, Q Cronk, R Cunningham, J Davis, S Degroeve, A Dejardin, C Depamphilis, J Detter, B Dirks, I Dubchak, S Duplessis, J Ehlting, B Ellis, K Gendler, D Goodstein, M Gribskov, J Grimwood, A Groover, L Gunter, B Hamberger, B Heinze, Y Helariutta, B Henrissat, D Holligan, R Holt, W Huang, N Islam-Faridi, S Jones, M Jones-Rhoades, R Jorgensen, C Joshi, J Kangasjarvi, J Karlsson, C Kelleher, R Kirkpatrick, M Kirst, A Kohler, U Kalluri, F Larimer, J Leebens-Mack, JC Leple, P Locascio, Y Lou, S Lucas, F Martin, B Montanini, C Napoli, DR Nelson, C Nelson, K Nieminen, O Nilsson, V Pereda, G Peter, R Philippe, G Pilate, A Poliakov, J Razumovskaya, P Richardson, C Rinaldi, K Ritland, P Rouze, D Ryaboy, J Schmutz, J Schrader, B Segerman, H Shin, A Siddiqui, F Sterky, A Terry, CJ Tsai, E Uberbacher, P Unneberg (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313(5793), 1596-604.
- Venturini L, A Ferrarini, S Zenoni, GB Tornielli, M Fasoli, SD Santo, A Minio, G Buson, P Tononi, ED Zago, G Zamperin, D Bellin, M Pezzotti & M Delledonne (2013). De novo transcriptome characterization of *Vitis vinifera* cv. Corvina unveils varietal diversity. *BMC Genomics* 14(41), 1-13.
- Wang N-J, C-C Lee, C-S Cheng, W-C Lo, Y-F Yang, M-N Chen & P-C Lyu (2012). Construction and analysis of a plant non-specific lipid transfer protein database (nsLTPDB). *BMC Genomics* 13(1), 1-9.
- Xiao XH, M Yang, JL Sui, JY Qi, YJ Fang, SN Hu & CR Tang (2017). The calcium-dependent protein kinase (CDPK) and CDPK-related kinase gene families in *Hevea brasiliensis* comparison with five other plant species in structure, evolution, and expression. *FEBS Open Bio* 7(1), 4-24.
- Yu G, W Hou, X Du, L Wang, H Wu, L Zhao, L Kong & H Wang (2014). Identification of wheat non-specific lipid transfer proteins involved in chilling tolerance. *Plant cell reports* 33(10), 1757-1766.
- Zeng L, Q Zhang, R Sun, H Kong, N Zhang & H Ma (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nature Communications* 5(4956), 1-12.
- Zhang L, B Liu, G Zheng, A Zhang & R Li (2017). Genome-wide characterization of the SiDof gene family in foxtail millet (*Setaria italica*). *Biosystems* 151, 27-33.
- Zhang QJ & LZ Gao (2016). The complete chloroplast genome sequence of desert poplar (*Populus euphratica*). *Mitochondrial DNA Part A* 27(1), 721-723.
- Zhiyi N, K Guijuan, L Yu, D Longjun & Z Rizhong (2015). Whole-transcriptome survey of the putative ATP-binding cassette (ABC) transporter family genes in the latex-producing laticifers of *Hevea brasiliensis*. *PloS ONE* 10(1), 1-23.
- Zhu H, Y Wang, H Yin, M Gao, Q Zhang & Y Chen (2015). Genome-wide identification and characterization of the LRR-RLK gene family in two *Vernicia* species. *International journal of genomics* 1-17
- Zou Z, J Gong, F An, G Xie, J Wang, Y Mo & L Yang (2015). Genome-wide identification of rubber tree (*Hevea brasiliensis* Muell. Arg.) aquaporin genes and their response to ethephon stimulation in the laticifer, a rubber-producing tissue. *BMC Genomics* 16(1001), 1-18.