

KONSTRUKSI PUSTAKA GENOM KAKAO (*Theobroma cacao* L.) UNTUK SEKUENSING GENOM TOTAL MENGGUNAKAN NEXT GENERATION SEQUENCING HiSeq2000

GENOMIC LIBRARY CONSTRUCTION OF COCOA (*Theobroma cacao* L.) FOR WHOLE GENOME SEQUENCING USING A NEXT GENERATION SEQUENCER HISEQ2000

I Made Tasma¹⁾, Dani Satyawan¹⁾, Habib Rijzaani¹⁾, dan Rubiyo²⁾

¹⁾Balai Besar Penelitian dan Pengembangan Bioteknologi dan Sumber Daya Genetik Pertanian

Jalan Tentara Pelajar 3A Bogor 16111, Indonesia

tasma12@yahoo.com

²⁾Balai Penelitian Tanaman Industri dan Penyegar

Jalan Raya Pakuwon km 2 Parungkuda, Sukabumi 43357

(Tanggal diterima: 28 Mei 2012, direvisi: 3 Juni 2012, disetujui terbit: 2 Juli 2012)

ABSTRAK

Pemuliaan kakao secara konvensional memerlukan waktu panjang (10-15 tahun). Pemanfaatan marka DNA akan memperpendek siklus pemuliaan kakao. Tujuan penelitian ini adalah mengkonstruksi pustaka genom tiga genotipe kakao yang dapat digunakan untuk sekuensing genom total kakao menggunakan NGS HiSeq2000 dan mendapatkan data resequencing genom total tiga genotipe kakao. Bahan tanaman terdiri dari tiga klon unggul kakao (ICCR02, ICCR04, dan SUL02) diperoleh dari Balittri, Pakuwon. DNA genomik diisolasi dari daun muda sebagai bahan konstruksi pustaka genom total. Sekuensing pustaka dilakukan pada mesin HiSeq2000 mengikuti protokol dari Illumina. Pustaka genom yang telah berhasil dikonstruksi berukuran 300 pasang basa (bp) masing-masing dengan konsentrasi 14,70 ng/ μ L (ICCR02), 15,20 ng/ μ L (ICCR04), dan 12,90 ng/ μ L (SUL02). Ukuran dan konsentrasi pustaka genom yang dihasilkan sangat ideal untuk sekuensing menggunakan HiSeq2000. Sekuensing ketiga genom menghasilkan data sekuen 52,9 x 10⁹ bp. Kluster DNA pustaka genom memiliki nilai $Q_{scores} > 30$ (75,0%) dengan tingkat kesalahan pembacaan basa rendah (1,47%). Nilai densitas kluster, persen kluster PF, intensitas basa, persen *phasing*, dan persen *prephasing* menunjukkan kualitas kluster pustaka genom ketiga genotipe kakao termasuk kategori pustaka ideal. Data sekuen yang dihasilkan juga sangat ideal untuk identifikasi marka SNP genom kakao. Koleksi marka SNP digunakan untuk identifikasi gen pengendali karakter penting kakao dan pemuliaan berbasis marka DNA untuk memperpendek siklus pemuliaan kakao.

Kata Kunci : Pustaka genom, genom total, sekuensing, HiSeq2000, kakao

ABSTRACT

Conventional cocoa breeding is slow and takes about 10-15 years to complete a breeding cycle. Applying genomic technology using DNA markers will significantly decrease cocoa breeding cycle. The objectives of this study were to construct cocoa whole genome genomic libraries to be used for resequencing the whole genome of cocoa and obtain whole genome resequence data of three cocoa genotypes. Three Indonesian cocoa genotypes (ICCR02, ICCR04, and SUL02) were used. DNA genomic was isolated from young leaf and used to construct genomic DNA libraries and generate DNA clusters. DNA clusters were sequenced using a HiSeq2000 platform. The whole genome libraries of the cocoa genotypes were successfully constructed. The library size was 300 bp with concentrations of 14.70 ng/ μ L (ICCR02), 15.20 ng/ μ L (ICCR04), and 12.90 ng/ μ L (SUL02), respectively. The genomic library size and concentrations are suitable for sequencing study using the NGS HiSeq2000. Total sequencing output obtained was 52.9 x 10⁹ bp. The genomic library clusters resulted during the sequencing process demonstrated the $Q_{scores} > 30$ of 75.0% with low error sequencing rate of 1.47%. Cluster densities, percentage of cluster PF, base intensity, and percentage of phasing and prephasing indicated the cluster quality of the genomic libraries is classified as an ideal one to be used for resequencing study using NGS HiSeq2000. The resequence data were ideal for SNP marker discovery. SNP markers are used to identify economically important genes of cocoa and marker-aided cocoa breeding to decrease the cocoa breeding cycle.

Keywords : Genomic library, whole genome, sequencing, HiSeq2000, cocoa

PENDAHULUAN

Usaha untuk meningkatkan produktivitas kakao dapat dilakukan melalui perbaikan genetik tanaman. Genom total kakao sudah disekuen dan sekuen rujukan kakao telah tersedia untuk publik (Argout *et al.*, 2011). Dengan tersedianya sekuen rujukan membuka jalan untuk eksploitasi plasma nutfah kakao nasional untuk mengidentifikasi gen atau QTL pembawa sifat unggul bernilai ekonomi tinggi seperti produktivitas dan ketahanan hama dan penyakit utama kakao.

Kakao merupakan tanaman diploid dengan jumlah kromosom diploid 20 pasang ($2n = 2x = 20$). Ukuran genom kakao relatif kecil sekitar $0,43 \times 10^9$ bp sangat mirip dengan ukuran genom padi. Dengan demikian berdasar ukuran genomnya tanaman kakao merupakan tanaman perkebunan model untuk studi genomika. Sekuen acuan beberapa genom kakao juga telah tersedia dan dapat diakses oleh publik. Dari sekuen genom acuan ini genom kakao mengandung sekitar 28.798 gen (Argout *et al.* 2011).

Resekuensing genom total berbagai plasma nutfah kakao nasional diperlukan untuk mengidentifikasi gen atau bagian kromosom yang mengendalikan karakter bernilai ekonomi tinggi. Variasi genetik dan gen unggul ini digunakan sebagai materi pemuliaan untuk mempercepat peningkatan kemajuan pemuliaan kakao. Informasi genomik ini juga bermanfaat untuk meningkatkan pemahaman genetik serta struktur genom kakao. Teknologi pemuliaan berdasarkan data genomik melalui *Marker-assisted Selection* (MAS) mempercepat siklus pemuliaan kakao. Pemanfaatan teknologi genomik meningkatkan akurasi dan efisiensi pemuliaan kakao karena seleksi tanaman pembawa karakter dapat dilakukan pada stadia awal pertumbuhan tanaman.

Next Generation Sequencing (NGS) mampu menyekuen genom total dalam waktu singkat dan biaya per unit data yang jauh lebih murah (Schuster, 2008; Pattersson *et al.*, 2009; Zhang *et al.*, 2011). Hal ini membuka peluang untuk mengetahui genotipe tanaman kakao secara komprehensif dan dalam waktu yang relatif singkat. Dengan demikian resekuensi genom total setiap individu tanaman kakao menjadi target untuk mengidentifikasi individu tanaman yang membawa

alel-alel yang bermanfaat untuk pemuliaan kakao. Hal ini didukung telah tersedianya sekuen rujukan kakao yang tersedia di publik yang memfasilitasi identifikasi karakter unggul pada genom kakao. Teknologi bioinformatika menjadi kata kunci dalam usaha identifikasi pengendali karakter unggul tersebut pada milyaran basa sekuen genom kakao. Penguatan sumberdaya manusia dalam bidang bioinformatika menjadi sangat penting untuk dapat memanfaatkan kemajuan teknologi genomik ini sebesar-besarnya untuk pemuliaan tanaman kakao yang efisien, efektif, dan berdaya guna.

Konstruksi pustaka genom merupakan langkah awal penggunaan NGS. Pustaka genom mengandung semua untaian DNA suatu genom. Pada pustaka genom tersimpan gen (*coding regions*) total yang dipunyai oleh suatu organisme dan daerah bukan penyandi protein (*non coding regions*). Pustaka genom sangat bermanfaat dalam usaha isolasi dan karakterisasi gen. Pemetaan gen secara fisik juga dapat dilakukan melalui konstruksi pustaka genom. Peta genetika tersebut sangat penting dalam program pemuliaan tanaman secara konvensional maupun secara molekuler (Suharsono, 2002).

Indonesia merupakan salah satu negara pembudidaya tanaman kakao paling luas di dunia dan termasuk negara penghasil kakao terbesar ketiga di dunia. Namun, produktivitas yang dihasilkan dari perkebunan kakao di Indonesia belum maksimal karena serangan hama dan penyakit yang menyebabkan menurunnya mutu produksi kakao. Usaha untuk meningkatkan produktivitas kakao dapat dilakukan melalui perbaikan genetika. Kendala lain yang dihadapi adalah terbatasnya informasi mengenai genetika dan studi tentang keseluruhan genom kakao. Oleh karena itu, konstruksi pustaka genom untuk sekuensi genom total kakao menggunakan *next generation sequencing* HiSeq2000 diharapkan dapat mempercepat peningkatan kemajuan pemuliaan kakao dan pemahaman mengenai informasi genetika serta genomik kakao.

Tujuan penelitian ini adalah mengkonstruksi pustaka genom tiga genotipe kakao untuk resekuensi genom total kakao menggunakan *next generation sequencer* HiSeq2000 dan mendapatkan data resekuensi genom total tiga genotipe kakao. Data resekuensi yang dihasilkan

digunakan untuk mengidentifikasi, mengisolasi, dan mengkoleksi marka SNP kakao genom total dengan menjajarkan data resekuen genotipe-genotipe kakao Indonesia dengan sekuen rujukan genom kakao yang sudah tersedia untuk publik.

BAHAN DAN METODE

Penelitian dilakukan di laboratorium Kelti Biologi Molekuler Balai Besar Penelitian Bioteknologi dan Sumber Daya Genetik Pertanian dari bulan Februari sampai dengan Juni 2012.

Bahan Tanaman

Bahan tanaman yang digunakan adalah tiga klon unggul kakao Indonesia yaitu ICCR02, ICCR04, dan SUL02. Bahan tanaman ini diperoleh dari Balai Penelitian Tanaman Industri dan Penyegar Pakuwon.

Isolasi, Uji Kualitas dan Kuantitas DNA

DNA genomik diisolasi dari tepung daun muda dari setiap varietas kakao (ICCR02, ICCR04, dan SUL02) menggunakan buffer CTAB dengan mengikuti metode Michiels *et al.* (2003) yang dimodifikasi. Protokol ini didesain khusus untuk mengisolasi DNA untuk tanaman bergetah. Urutan kerja lengkap dari isolasi DNA genomik ini telah dilaporkan sebelumnya (Satyawan dan Tasma, 2011). DNA hasil isolasi dilarutkan dalam 50 μ L bufer TE (10 mM Tris pH 8, 1 mM EDTA pH 7,5). Konsentrasi DNA diukur dengan spektrofotometer nanodrop (Thermo Scientific, USA). Pita DNA genomik yang dihasilkan di elektroforesis menggunakan 1% gel agarose (Sambrook *et al.*, 1989).

Konstruksi Pustaka Genom Total Kakao

Konstruksi pustaka genom kakao dilakukan dengan menggunakan protokol TruSeq DNA *low throughput protocol* (LT) dari Illumina (Illumina, 2010c). Pembentukan pustaka genom total menggunakan urutan kerja sebagai berikut: (1) fragmentasi DNA yang dilakukan dengan menggunakan metode nebulisasi, (2) modifikasi ujung fragmen menggunakan *Insert Modification Plate* (IMP) dikombinasikan dengan penggunaan *AMPure XP Beads* dan *magnetic stand*, dan (3) adenilasi ujung 3' DNA yang terdiri dari

penempelan adapter, purifikasi DNA hasil ligasi, amplifikasi PCR dan pemurnian pustaka genom total kakao, dan validasi pustaka genom total kakao.

Sekuensing Pustaka Genom Total Kakao pada Mesin Hiseq2000

Proses sekuensing genom total terdiri dari dua tahap yaitu klusterisasi pustaka genom total kakao pada mesin cBot dan sekuensing kluster DNA pustaka genom pada mesin NGS HiSeq2000. Klusterisasi DNA pustaka genom total dilakukan pada mesin cBot menggunakan protokol *cluster generation protocol* dari Illumina (Illumina, 2010b). Pembentukan kluster diawali dari preparasi sampel yang terdiri dari dua tahapan, yaitu denaturasi dengan menggunakan 2 N NaOH dan pelarutan DNA dalam buffer hibridisasi. Sampel diinkubasi selama 5 menit pada suhu ruang untuk mendenaturasi utas ganda DNA menjadi utas tunggal.

Tahapan preparasi sampel dilakukan dengan pengenceran DNA dalam buffer hibridisasi. Satu nM DNA hasil denaturasi diencerkan menjadi 5 pM, 6 pM, 7 pM dan 8 pM. Sebanyak 120 μ L *control library* (DNA phiX) dari Illumina disiapkan yang digunakan sebagai kontrol dalam diletakkan pada lajur 4 dari *flow cell*. Posisi sampel pada mesin *cBot Cluster Generation* dilakukan dengan mengikuti petunjuk pembentukan kluster DNA pada mesin cBot.

Sekuensing pustaka genom total kakao menggunakan mesin perunut urutan basa, Illumina HiSeq 2000 (Illumina, 2010a; Illumina, 2010d) terdiri dari beberapa tahap yaitu persiapan reagen, pemilihan parameter untuk sekuensing, pemasangan *flow cell*, sekuensing, dan *post-run sequencing (maintenance wash)*. Reagen yang digunakan selama sekuensing diantaranya SBS reagen (TruSeq SBS kit), *multiplexing reagen* dan *paired-end reagen*. Reagen-reagen disiapkan dengan mengikuti protokol dari Illumina dan dimasukkan ke dalam rak dan disimpan dalam ruang *reagen compartment* yang terdapat dalam mesin Illumina HiSeq 2000. Pemilihan parameter untuk sekuensing dilakukan melalui *Hiseq control software interface*. Parameter yang tersedia diantaranya *flow cell ID*, *experiment*, *user name*, *flow cell type*, *control lane*, *output folder*, *confirm first base*, *keep intensity*

files, existing recipe, save image dan align lanes. Tahapan selanjutnya yaitu pemasangan *flow cell* yang telah diklasterisasi melalui mesin *cBot cluster generation* pada *flow cell stage*. Sekuensing dilakukan satu arah (*single read*) yang dilakukan selama sekitar 11 hari dengan panjang pembacaan DNA 2 x 101 siklus. Tahapan terakhir setelah proses sekuensing selesai yaitu dilakukan *maintenance wash*. Tahapan *maintenance wash* mesin HiSeq2000 dengan menggunakan air dan NaOH dilakukan dengan mengikuti protokol *maintenance wash* dari Illumina.

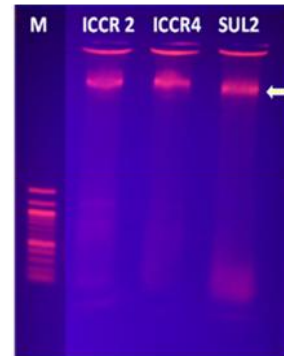
HASIL DAN PEMBAHASAN

Kualitas dan Kuantitas DNA Genomik Tiga Genotipe Kakao

Protokol isolasi DNA genomik tiga genotipe kakao menghasilkan DNA dengan kualitas dan kuantitas yang tinggi (Gambar 1 dan Tabel 1). DNA genomik yang dihasilkan utuh dengan pita yang besar dan terang (Gambar 1).

Konsentrasi DNA yang dihasilkan juga sangat tinggi pada ketiga genotipe kakao masing-masing 198,1 ng/ μ L (ICCR02), 259,5 ng/ μ L (ICCR04) dan 303,7 ng/ μ L (SUL02). Kualitas DNA yang dihasilkan sangat baik seperti ditunjukkan oleh nilai rasio A260/A280 yang berkisar 1,992 sampai dengan 2,05 (Tabel 1). Kisaran standar mutu DNA yang baik harus mempunyai rasio A260/A280 berkisar 1,8-2,0. Ini menunjukkan bahwa kontaminan protein (A280) pada sampel DNA sangat rendah dan sangat memenuhi syarat untuk digunakan sebagai materi konstruksi pustaka genom total. Demikian juga dengan nilai rasio A260/A230 yang menunjukkan

kontaminan lain seperti karbohidrat dan senyawa lainnya selain protein juga telah memenuhi syarat kualitas DNA dengan kriteria kualitas tinggi. Kualitas DNA yang dihasilkan sangat ideal untuk digunakan pada penelitian sekuensing dengan NGS. Kualitas DNA ini juga ideal untuk penelitian biologi molekuler lain yang memerlukan kualitas DNA tinggi seperti penelitian kloning gen.

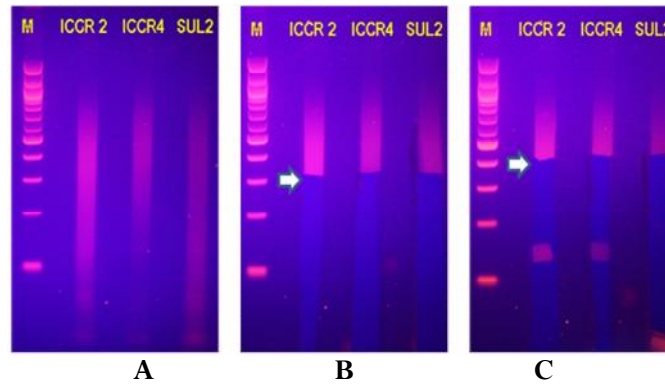


Gambar 1. DNA genomik kualitas tinggi tiga genotipe kakao (ICCR02, ICCR04, dan SUL02). DNA diisolasi menggunakan protokol Michiels *et al.* (2003), dielektroforesis pada 1% gel agarose. Tanda panah menunjukkan DNA genomik dengan berat molekul tinggi (*high molecular weight*) yang digunakan untuk penyediaan pustaka genom total kakao. M = 100 basa marka DNA

Figure 1. High quality genomic DNA of three cocoa genotypes (ICCR02, ICCR04, and SUL02). DNA was isolated based on the protocol of Michiels *et al.* (2003), electrophoresed in 1% agarose gel. The arrow indicates the genomic high molecular weight DNA used for genomic library preparation. M= 100 bp DNA ladder

Tabel 1. Kuantitas dan kemurnian DNA genomik total tiga genotipe kakao
 Table 1. Quantity and purity total genomic DNA isolated from three cocoa genotypes

Genotipe	[DNA] ng/ μ L	A260	A280	260/280	260/230
ICCR02	198,1	3,962	1,992	1,99	1,57
ICCR04	259,5	5,191	2,539	2,04	1,72
SUL02	303,7	6,075	2,968	2,05	1,61



Gambar 2. DNA genomik hasil nebulisasi yang sudah dimodifikasi dan diligasi dengan adapter disiapkan dari DNA genomik tiga genotipe kakao (ICCR02, ICCR04, dan SUL02). (A) DNA genomik hasil modifikasi dan penempelan dengan adapter, (B) DNA calon pustaka genom berukuran 300 bp dipotong dari gel, (C) DNA genomik calon pustaka genom berukuran 400 bp dipotong dari gel. Tanda panah menunjukkan fragmen DNA genomik berukuran 300 bp dan 400 bp dipotong dari gel, dipurifikasi untuk persiapan pustaka genom total tiga genotipe kakao. M = 100 basa marka DNA

Figure 2. Genomic DNA resulted from nebulization that were ligated with the adapter of three cocoa genotypes (ICCR02, ICCR04, and SUL02). (A) The banding pattern of DNA fragments after nebulisation and modification with appropriate steps and adapter, (B) the 300 bp DNA bands sliced from the gel, (C) the 400 bp DNA bands sliced from the gel. Arrows indicate the 300 and 400 bp DNA bands, respectively, used for genomic library preparation. M = 100 bp DNA ladder

Tabel 2. Kuantitas dan kemurnian DNA genom kakao hasil fragmentasi

Table 2. Quantity and purity the genomic DNA after nebulization

Genotype	[DNA] ng/ μ L	A260	A280	260/280	260/230
ICCR02	113,3	2,266	1,092	1,97	2,01
ICCR04	123,7	2,473	1,169	2,05	1,91
SUL02	145,1	2,903	1,372	1,83	1,98

Konstruksi Pustaka Genom Tiga Genotipe Kakao

Fragmen DNA genomik hasil nebulisasi

Dalam penelitian ini, tekanan gas yang digunakan selama proses nebulisasi yaitu sebesar 40 psi. Penggunaan tekanan gas tersebut menghasilkan fragmen DNA kurang dari 800 bp dengan intensitas DNA paling tinggi berada pada ukuran pita sekitar 500 bp. Secara visual, DNA genom setelah fragmentasi terlihat seperti *smear* (Gambar 2A). Hal tersebut menunjukkan bahwa DNA genom telah berhasil difragmentasi dan menghasilkan ukuran fragmen antara 100-800 bp. Hasil tersebut sesuai dengan ukuran fragmen DNA yang dibutuhkan dalam pembuatan pustaka genom untuk sekuensing dengan NGS HiSeq2000. Protokol *LT DNA preparation* dari Illumina (Illumina, 2010c) merekomendasikan bahwa ukuran fragmen DNA yang diperlukan untuk konstruksi pustaka genom untuk NGS HiSeq2000 berukuran kurang dari 800 bp.

Hasil uji kuantitatif DNA hasil fragmentasi menunjukkan bahwa DNA genom hasil fragmentasi

memiliki rasio A_{260}/A_{280} berkisar 1,83-2,05 (Tabel 2). Hasil tersebut menunjukkan bahwa DNA genom hasil fragmentasi memiliki tingkat kemurnian yang baik karena rasio A_{260}/A_{280} yang dimiliki masih termasuk ke dalam kisaran 1,8-2,0 (Walker dan Wilson, 2000). Konsentrasi DNA ketiga genotipe kakao setelah fragmentasi sangat tinggi, masing-masing 113,3 ng/ μ L (ICCR02), 123,7 ng/ μ L (ICCR04) dan 145,1 ng/ μ L (SUL02) (Tabel 2). Konsentrasi ini jauh dari cukup untuk persiapan pustaka genom total tanaman kakao.

Pustaka genom total tiga genotipe kakao

Tiga pustaka genom dikonstruksi menggunakan protokol TruSeq LT library preparation dari Illumina (Illumina Inc., USA). DNA genomik hasil fragmentasi mengalami serangkaian proses yang meliputi modifikasi ujung fragmen DNA untuk membuat ujung fragmen tumpul, adenilasi ujung 3' DNA, penempelan adapter pada ujung-ujung fragmen DNA, dan memisahkan fragmen DNA yang sudah memiliki adapter pada gel agarose. Pada setiap langkah pembuatan pustaka genom ini dilakukan

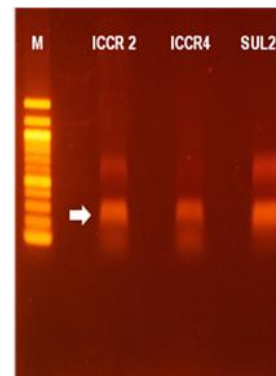
pengukuran kuantitas dan kualitas DNA yang dihasilkan pada setiap step penyiapan DNA pustaka dengan mengukurnya menggunakan spektrofotometer nanodrop (*Thermo Scientific, USA*). Dengan pengukuran ini diketahui konsentrasi DNA yang diperoleh dari setiap step penyiapan pustaka genomik kakao.

Pada penelitian ini telah dipilih fragmen DNA hasil fragmentasi yang berukuran 300 dan 400 bp. Pita DNA berukuran 300 dan 400 bp dipisahkan (*sliced*) dari gel dan dimurnikan menggunakan protocol dari Illumina.

Pita DNA berukuran 300 dan 400 bp diambil dari gel, dipurifikasi dengan kit Qiagen (*Qiagen Inc., USA*). DNA hasil purifikasi dari gel agarose diamplifikasi dengan PCR. Hasil PCR yang sudah dipurifikasi adalah pustaka genom total ketiga genotipe kakao. Pola pita pustaka genom total ketiga genotipe kakao yang dihasilkan pada penelitian ini disajikan pada Gambar 2. Pada penelitian ini hanya dipilih pustaka berukuran 300 bp yang digunakan pada penelitian sekuensing menggunakan HiSeq2000 (Gambar 3). Hasil tersebut sesuai dengan ukuran fragmen DNA yang dipilih pada saat tahap purifikasi fragmen DNA hasil ligasi saat preparasi konstruksi pustaka genom.

Konsentrasi DNA pustaka genom total kakao berkisar 12,9-15,2 ng/uL (Tabel 3). Konsentrasi tersebut memenuhi syarat untuk digunakan sebagai sekuensing genom total menggunakan NGS HiSeq2000 yang memerlukan konsentrasi pustaka genom setidaknya 10 nM. Kualitas DNA pustaka genom yang dihasilkan

(A260/A280) juga sangat baik berkisar 1,89-2,01 (Tabel 3) dibanding kualitas rujukan A260/A280 berkisar 1,8-2,0. Demikian juga kualitas DNA dilihat dari kontaminan lain selain protein yang ditunjukkan oleh nilai rasio A260/A230 yang masih pada kisaran rujukan. Dengan demikian pustaka genom ketiga genotipe kakao siap digunakan untuk penelitian sekuensing genom total kakao.



Gambar 3. Pustaka DNA genomik tiga genotipe kakao (ICCR02, ICCR04, dan SUL02). Pustaka DNA hasil purifikasi diamplifikasi dengan PCR dan dielektroforesis pada 1% gel agarose. Tanda panah menunjukkan fragmen DNA pustaka DNA genomik berukuran 300 bp yang digunakan untuk penelitian sekuensing. M = 100 basa marka DNA

Figure 3. Genomic DNA library of three cocoa genotypes (ICCR02, ICCR04, and SUL02). Purified genomic libraries were PCR amplified and electrophoresed in 1% agarose gel. The arrow indicates the genomic library of interest that was used in the sequencing experiments. M=100 bp DNA ladder

Tabel 3. Kuantitas dan kemurnian pustaka genom total tiga genotipe kakao

Table 3. The quantity and purity of genomic libraries derived from three cocoa genotypes

Genotipe	[DNA] ng/ μ L	A260	A280	260/280	260/230
ICCR02	14,7	0,147	0,073	2,01	2,39
ICCR04	15,2	0,127	0,067	1,89	1,98
SUL02	12,9	0,159	0,082	1,94	2,19

Tabel 4. Rangkuman hasil sekuensing 3 genotipe kakao pada penelitian sekuensing menggunakan mesin HiSeq2000

Table 4. Data summary of three Indonesian cocoa genotypes resulted from the sequencing experiment using the HiSeq2000 platform

Yield Total (Gbp) ^a	Projected Total Yield (Gbp) ^a	Yield Perfect (Gbp)	Yield ≤ 3 errors (Gbp)	% Perfect [Num cycles]	% ≤ 3 errors [Num cycles]
52.9	52.9	4.6	7.1	86.2 [100]	86.6 [100]

Keterangan: ^aData pada kolom 1 dan 2 adalah data sekuen seluruh lajur dari *flow cell*. Data pada kolom lainnya(3-6) adalah data PhiX terletak pada lajur 4 pada *flow cell* yang berfungsi sebagai kontrol selama proses sekuensing

Notes : ^aData of the first and second columns are the whole lane data of the flow cell. Data of other columns (3-6) are the PhiX sequence data run in lane 4 of the flow cell functions as a sequence control during the sequencing process.

Tabel 5. Densitas dan kualitas pustaka genom hasil sekuensing menggunakan mesin sekuensing HiSeq2000
Table 5. Density and quality of the genomic libraries resulted from sequencing of the genomic libraries using HiSeq2000 platform.

Lajur	Densitas (K/mm ²)	Klaster PF (%)	Phasing (%)	Prephasing (%)	Reads (Mbp)	Read PF (Mbp)	%>=30
ICCRI02 8 pM	366	89,01	0,476	0,503	59,95	59,95	71,6
ICCRI04 8 pM	314	87,90	0,500	0,602	50,50	50,50	73,1
SUL02 8 pM	312	89,30	0,516	0,545	51,39	51,39	71,9
Kontrol (PhiX)	527	88,15	0,330	0,395	85,67	85,67	84,0
Jamur BLB 12 pM	519	88,00	0,339	0,337	84,30	84,30	81,9
ICCRI02 12 pM	458	86,60	0,346	0,488	73,40	73,40	82,5
ICCRI04 12 pM	416	86,94	0,429	0,580	63,60	63,60	72,8
SUL02 12 pM	403	86,80	0,510	0,592	59,61	59,61	71,7

Data Hasil Sekuensing Pustaka Genom Total Tiga Genotipe Kakao

Hasil penelitian sekuensing pustaka genom tiga genotipe kakao menghasilkan data sekuen sebanyak $52,9 \times 10^9$ basa (Tabel 4). *Yield perfect* yang diperoleh selama proses sekuensing yaitu 4,6 Gbp (Tabel 4). Nilai tersebut menunjukkan jumlah basa yang secara akurat melewati *passing filter*. *Passing filter* adalah sejenis algoritma yang dipakai oleh komputer untuk menentukan keakuratan pembacaan basa saat sekuensing. *Passing filter* yang digunakan berdasarkan data sekuen PhiX yang berperan sebagai kontrol pada tahapan sekuensing menggunakan NGS HiSeq 2000. Hasil ini juga terkait dengan *percent perfect* sebesar 86,2% untuk data kontrol pada lajur 4 menggunakan DNA PhiX. Semua data di atas menunjukkan bahwa pustaka DNA dari ketiga genotipe kakao sangat ideal digunakan dalam penelitian sekuensing menggunakan HiSeq2000.

Hasil sekuensing menunjukkan bahwa densitas klaster pustaka genom yang dihasilkan berkisar antara 312-527 K/mm² (Tabel 5). Dari keenam lajur *flow cell* yang digunakan untuk kakao empat (50%) lajur *flow cell* termasuk kategori ideal yaitu lajur ICCR02 12 pM, ICCR04 12 pM, SUL02 12 pM, dan kontrol dengan densitas klaster di atas 400 K/mm² (Tabel 5). Densitas klaster pustaka genom termasuk kategori ideal jika klaster pustaka genom memiliki densitas berkisar 400-600 K/mm² (Illumina, 2009). Lajur lainnya (semuanya dengan konsentrasi 8 pM menunjukkan densitas lebih dari 300 K/mm² namun kurang dari 400 K/mm² (Tabel 5) yang masih cukup ideal untuk sekuensing menggunakan HiSeq2000. Densitas klaster pustaka

genom yang rendah dapat disebabkan karena konsentrasi pustaka genom yang digunakan dalam proses klasterisasi lebih rendah dari konsentrasi yang seharusnya (8 pM atau 12 pM). Hal ini menunjukkan bahwa konsentrasi 8 pM dari penelitian ini belum cukup untuk menghasilkan kepadatan klaster yang ideal. Hal tersebut dapat terjadi kemungkinan karena kesalahan dalam pengukuran konsentrasi pustaka genom sebelum klasterisasi (Quail *et al.*, 2008). Dengan meningkatkan konsentrasi pustaka menjadi 12 pM menghasilkan jumlah klaster/mm² yang ideal untuk sekuensing menggunakan HiSeq2000.

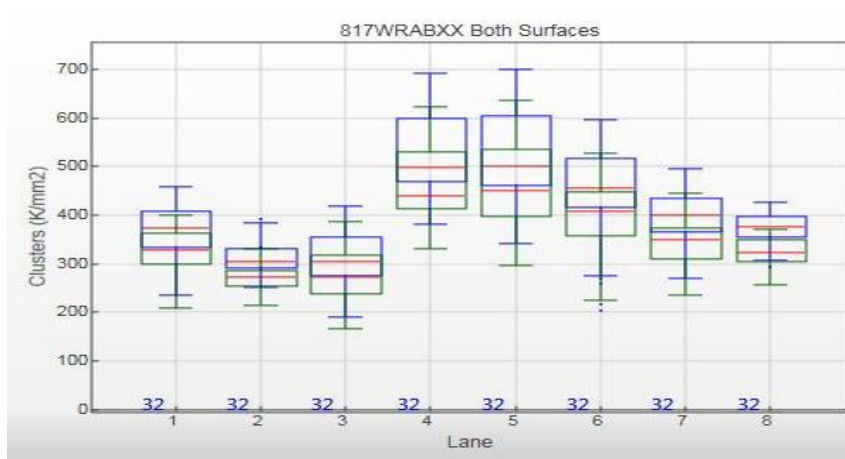
Kualitas urutan basa hasil sekuensing dapat diketahui melalui nilai *Q scores*. Nilai *Q scores* didefinisikan sebagai hubungan logaritmik antara proses *base calling* selama sekuensing dengan kemungkinan tingkat kesalahan yang dihasilkan dalam proses tersebut (Ewing dan Green, 1998). Voelkerderling *et al.* (2009) menyatakan bahwa *base calling* diartikan sebagai proses konversi data hasil pencitraan pendaran fluorescen selama proses sekuensing menjadi basa-basa yang berurutan. Nilai *Q scores* yang dihasilkan pada penelitian ini menunjukkan nilai *Q scores* 30 dengan nilai persentase diatas 70% pada setiap lajurnya (Tabel 5). Nilai *Q score* 30 berarti bahwa dari 1000 basa yang dibaca hanya ada satu basa yang salah pembacaan (Ewing and Green, 1998; Voelkerderling *et al.*, 2009). Secara keseluruhan, nilai *Q scores* 30 yang dihasilkan selama proses sekuensing yaitu 75,0% dengan jumlah basa sebanyak 40,1 Gbp atau $40,1 \times 10^9$ bp (Gambar 4).

Selama proses sekuensing, sekuen yang dihasilkan dilajur kontrol (PhiX) dibandingkan

dengan sekuen PhiX yang telah tersedia di database. Hal tersebut bertujuan untuk mengetahui tingkat kesamaan antara sekuen yang dihasilkan pada lajur kontrol dengan sekuen PhiX pada database. Nilai kesamaan (*aligned*) yang diperoleh selama proses sekuensing yaitu 95,5% (Tabel 6).

Tingkat kesalahan selama proses sekuensing juga dapat diketahui dari nilai % *error rate*. Nilai % *error rate* menunjukkan persentase

perbandingan kesalahan pembacaan basa antara pustaka genom dengan kontrol yang dibaca pada beberapa siklus sekuensing yaitu pada siklus 35, 75, dan 100. Secara keseluruhan, nilai *error rate* yang dihasilkan selama proses sekuensing pada penelitian ini yaitu 1,47% (Tabel 6). Hasil tersebut menunjukkan bahwa sekuensing yang telah dilakukan berlangsung dengan baik dengan tingkat kesalahan yang kecil.



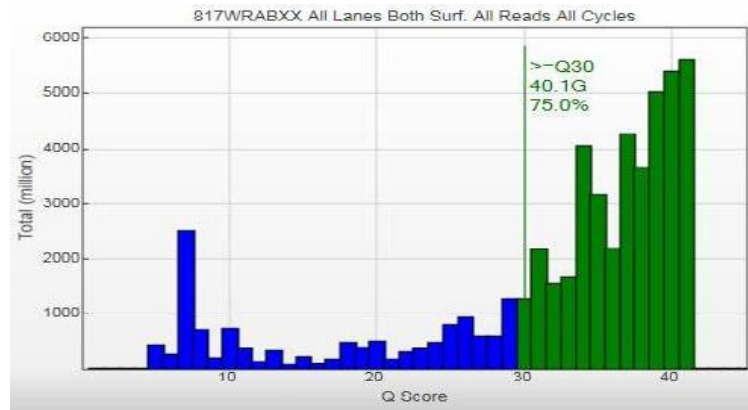
Gambar 4. Kurva jumlah densitas kluster setiap lajur dari 8 lajur flow cell percobaan sekuensing 3 genotipe kakao. Lajur 1, ICCR02 8 pM; lajur 2, ICCR04 8 pM; lajur 3, SUL02 8 pM; lajur 4, control phix; lajur 5, BLB patogen; lajur 6, ICCR02 12 pM; lajur 7, ICCR04 12 pM; dan lajur 8, SUL02 12 pM

Figure 5. Cluster density curve of the 8-flow cell lane used in the resequencing experiment of three cocoa genotypes. Lane 1, ICCR02 8 pM; lane 2, ICCR04 8 pM; lane 3, SUL02 8 pM; lane 4, control phix; lane 5, BLB pathogen; lane 6, ICCR02 12 pM; lane 7, ICCR04 12 pM; and lane 8, SUL02 12 pM

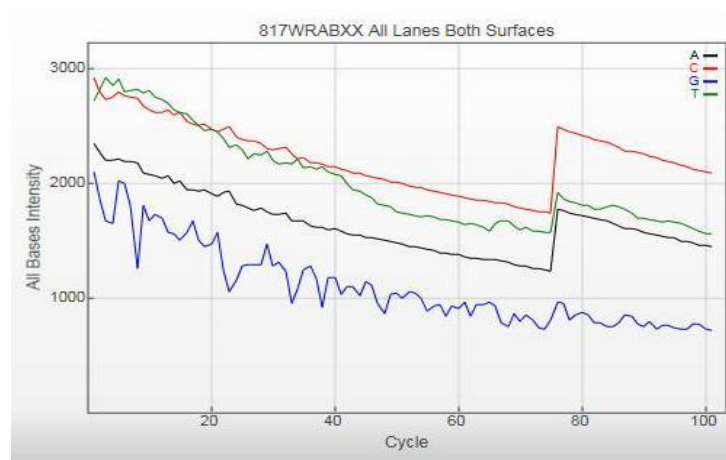
Tabel 6. Tingkat kesalahan (*error rate*), kesamaan sekuen (*sequence aligned*), dan intensitas basa hasil sekuensing pustaka genom tiga genotipe kakao menggunakan mesin HiSeq2000

Table 6. Sequencing data error rates, aligned, and base intensity of three Indonesian cocoa genotypes resulted from sequencing experiments using the HiSeq2000 platform

Lane	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
ICCR02 8 pM	0	0	0	0	0	2363	81.1
ICCR04 8 pM	0	0	0	0	0	2472	79.4
SUL02 8 pM	0	0	0	0	0	2471	80.3
Kontrol (PhiX)	95.5	1.47	0.19	0.88	0	2304	78.5
Jamur BLB 12 pM	0	0	0	0	0	2449	79.8
ICCR02 12 pM	0	0	0	0	0	2337	83.6
ICCR04 12 pM	0	0	0	0	0	2387	80.4
SUL02	0	0	0	0	0	2501	81.6



Gambar 5. Nilai Q -score diatas 30 (terdapat hanya satu kesalahan dalam setiap 1000 basa) hasil sekuensing tiga genotipe kakao
Figure 5. Q -scores values >30 (only one error occurs in a total of 1000 base) sequence resulted from the sequencing experiment of three cocoa genotypes



Gambar 6. Intensitas basa-basa nitrogen selama proses sekuensing tiga genotipe kakao Indonesia menggunakan 100 siklus sekuensing
Figure 6. Base intensities of all bases shown across 100-sequencing cycles during the sequencing process of the tree cocoa genotype whole genomes

Basa adenin, timin, guanin dan sitosin yang dihasilkan selama proses sekuensing dihitung intensitas setiap siklusnya. Intensitas keempat basa tersebut dideteksi melalui hasil pendaran *fluorescen* yang dihasilkan dari sintesis basa selama proses sekuensing. Intensitas basa yang dihasilkan setiap lajur *flow cell* pada siklus pertama memiliki nilai lebih dari 2000 (Tabel 6). Intensitas basa kemudian dihitung kembali pada siklus ke 20 dan menghasilkan persen intensitas sekitar 80% dari intensitas basa pada siklus pertama.

Secara keseluruhan, intensitas keempat basa setiap siklus dapat dilihat pada gambar 6. Keempat basa yang dihasilkan memiliki intensitas lebih dari 1000 kecuali intensitas beberapa basa setelah siklus 100. Hal tersebut terjadi karena kemungkinan pendaran *fluorescen* yang dideteksi setelah siklus 100 bukan berasal dari fragmen DNA melainkan berasal dari label adapter yang menempel pada *flow cell*. Kircher *et al.* (2011)

menjelaskan bahwa rendahnya intensitas dapat disebabkan oleh beberapa hal, di antaranya rendahnya kluster pustaka genom yang dihasilkan selama proses *bridge amplification*, ukuran pustaka genom yang terlalu panjang, penempelan primer yang tidak efisien, aktivitas enzim polimerase yang tidak optimal dan terjadi gangguan dalam pendeteksian pendaran *fluorescen*. Namun, secara umum intensitas keempat basa yang dihasilkan selama proses sekuensing termasuk kedalam kategori ideal. Illumina (2009) menyatakan bahwa nilai intensitas ideal yang dihasilkan selama proses sekuensing yaitu sebesar 1000.

KESIMPULAN

Pustaka genom tiga genotipe kakao (ICCR02, ICCR04, dan SUL02) yang berhasil dikonstruksi berukuran 300 bp masing-masing dengan konsentrasi 14,70 ng/ μ L (ICCR02), 15,20

ng/ μ L (ICCR04), dan 12,90 ng/ μ L (SUL02). Ukuran dan konsentrasi pustaka genom tersebut sangat ideal untuk sekuensing menggunakan NGS HiSeq2000. Jumlah basa total yang dihasilkan dari sekuensing pustaka genom tersebut 52,9 x 10⁹ bp. Kluster pustaka genom yang dihasilkan memiliki nilai *Q scores* diatas 30 (75,0%) dengan tingkat kesalahan pembacaan basa hanya 1,47%. Nilai densitas kluster, persen kluster PF, intensitas basa, persen *phasing*, dan persen *prephasing* yang dihasilkan menunjukkan kualitas kluster pustaka genom ketiga genotipe kakao termasuk kategori ideal. Data sekuen yang dihasilkan juga sangat ideal untuk identifikasi marka SNP genom kakao.

UCAPAN TERIMA KASIH

Tim penulis mengucapkan terima kasih kepada Saudara Andi Kosasih atas bantuan dan keterlibatannya pada penelitian ini. Penelitian ini didanai dari APBN BB Biogen tahun anggaran 2011.

DAFTAR PUSTAKA

- Alvim, P. T., A. D. Machado, and F. Vello. 1975. Physiological responses of cacao to enviroent factors. *J Revista Theobroma* 4 (3):3-12.
- Argout, X., J. Salse, J. M. Aury, M. J. Guiltinan, G. Droc, J. Gouzy, M. Allegre, C. Chaparro, T. Legavre1, S. N. Maximova, M. Abrouk, F. Murat, O. Fouet, J. Poulain, M. Ruiz, Y. Roguet, M. Rodier-Goud, J. F. Barbosa-Neto, F. Sabot, D. Kudrna, J. S. S. Ammiraju, S. C. Schuster, J. E. Carlson^{12,13}, Erika Sallet⁸, Thomas Schiex¹⁴, Anne Dievart¹, Melissa Kramer¹⁵, Laura Gelley¹⁵, Zi Shi, A. Bérard, C. Viot, M. Boccara, A. M. Risterucci, V. Guignon, X. Sabau, M. J. Axtell *et al.* 2011. The genome of *Theobroma cacao*. *Nature Genetics* 4 (2): 101-108.
- Departemen Perindustrian. 2007. Gambaran Sekilas Industri Kakao. Jakarta, Pusat Data dan Informasi Departemen Perindustrian.
- Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8: 186-194.
- Illumina. 2009. *Sequencing Analysis Software : User Guide*. San Diego: Illumina Inc.
- Illumina. 2010a. HiSeq Sequencing Systems. San Diego: Illumina Inc.
- Illumina. 2010b. cBot Quick Reference Guide. San Diego: Illumina Inc.
- Illumina. 2010c. Truseq DNA Sample Preparation Guide. San Diego: Illumina Inc.
- Illumina. 2011d. HiSeq2000 Quick Reference Guide. San Diego: Illumina Inc.
- Michiels, A. N., E. W. Van Den, M. Tucker, R. Liesbet, and L. Van Andre. 2003. Extraction of high quality genomic DNA from latex containing plants. *Analytical Biochemistry* 315: 85-89.
- Patterson, E., J. Lundeberg, and A. Ahmadian. 2009. Generations of sequencing technologies. *Genomics* 93: 105-111.
- Pusat Penelitian Kopi dan Kakao. 2010. Budidaya Kakao. Jakarta, Agromedia Pustaka.
- Quail, S. 2008. A large genome center's improvements to the illumina sequencing system. *Nature Methods* 5: 1005-1010.
- Sambrook, J., E. F. Fritch, and T. Maniatis. 1989. *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory Press.
- Satyawan, D. and I. M. Tasma. 2012. DNA markers applicable for genetic mapping of *J. curcas* genome. *Buletin RISTR I* 2 (3): 411-419.
- Schuster, S. C. 2008. Next generation sequencing transform today's biology. *Nature Method* 5: 16-18.
- Suharsono. 2002. Konstruksi Pustaka Genom Kedelai Kultivar Slamet. *Hayati* 9 (2): 67-70.
- Voelkerderling, K. V., S. H. Dames, and J. D. Durtschi. 2009. Next generation sequencing : from basic research to diagnostics. *Clinical chemistry* 55: 41-47.
- Wahyudi, T.R. Panggabean, dan Pujiyanto. 2008. *Panduan Lengkap Kakao*. Jakarta : Penebar Swadaya.
- Walker, J. M., and K. Wilson. 2000. *Principles and Techniques of Practical Biochemistry*. UK, Cambridge University Press.
- Wardojo, S. 1992. Major pest and diseases of cocoa in Indonesia In Keane P.J. and C.A.J. Petter (Eds). *Cocoa Pest and Disease Management In Southeast Asia and Australia* FAO. *Plant Production* 112: 63-67.
- Widyanthi, Y. 2010. Kloning gen proteinase inhibitor dari kulit buah kakao pada vektor ekspresi dengan metode gateway. Skripsi Fakultas Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Zhang, J., R. Chiodini, A. Badr, and G. Zhang. 2011. The impact of next generation sequencing on genomics. *J. Genet. Genomics* 38: 95-109.