

## IDENTIFIKASI SNP GENOM PADA POPULASI *Elaeis guineensis* x *Elaeis oleifera*

### GENOME-WIDE SNP IDENTIFICATION OF THE *Elaeis guineensis* x *Elaeis oleifera* POPULATION

Sri Wening<sup>1</sup>, Heri Adriwan Siregar, Edy Suprianto, Dani Satyawan<sup>1</sup>, Hernawan Y Rahmadi, Retno Diah Setiowati, Mohamad Arif, Yurna Yenni, Nanang Supena, Sujadi, dan Abdul Razak Purba

**Abstrak** Usaha pencarian marka DNA yang berhubungan dengan sifat yang diinginkan pada *Elaeis oleifera* guna introgresi sifat tersebut ke genome *Elaeis guineensis* memerlukan marka DNA yang polimorfik. Untuk menghasilkan marka DNA yang polimorfik dengan jumlah banyak, identifikasi SNP genom dilakukan melalui pengurutan kembali (*resequencing*) 12 individu contoh populasi hibrida *E. guineensis* x *E. oleifera* (hibrida OxG), yaitu *E. oleifera* tipe liar, F1 hibrida interspesifik, *pseudo-backcross* dan material maju *E. guineensis*, menggunakan *next generation sequencing* (NGS). *Read* (urutan basa yang "dibaca"/merupakan keluaran mesin NGS) dari 12 contoh memiliki mutu yang baik dan 96% total *read* yang disaring dapat dilakukan demultipleks dan ditentukan pada contoh yang sesuai. Setelah proses penyaringan dan pemotongan, 84% *read* dapat digunakan untuk pemetaan genom dan menghasilkan 5,7X hingga 10,42X cakupan genom. Dari 34.410.224 SNP yang teridentifikasi, 98,7% diantaranya adalah varian *non-coding*, dan berdasarkan lokasi, 69,1% total SNP adalah SNP *intergenic*. Sebanyak 5.618 SNP dari total SNP yang dihasilkan dibuktikan menggunakan *targeted genotyping by sequencing* pada 500 individu contoh. Sebanyak 74% SNP yang digunakan bermutu tinggi yang dibaca pada setidaknya 95% contoh. *Principal*

*component analysis* menggunakan SNP tersebut mampu mengidentifikasi setiap latar belakang genetik contoh. Pembuktian tersebut menyimpulkan bahwa identifikasi SNP yang dilakukan melalui pengurutan kembali menghasilkan SNP bermutu tinggi yang dapat digunakan untuk pengembangan marka DNA yang dapat diperbantukan pada seleksi populasi pemuliaan *E. guineensis* x *E. oleifera*.

**Kata kunci:** SNP, pengurutan kembali, *Elaeis guineensis*, *Elaeis oleifera*, *targeted genotyping by sequencing*, *principal component analysis*

**Abstract** Efforts in finding DNA markers for favorable characters of *Elaeis oleifera* introgressed into *Elaeis guineensis* genome requires polymorphic DNA markers. To generate high number of polymorphic DNA markers, a genome-wide SNP identification was done by resequencing 12 individual samples using next generation sequencing (NGS) from an *E. guineensis* x *E. oleifera* (OxG hybrid) population comprising wild *E. oleifera* species, interspecific F1 hybrids, pseudo-backcrosses, and advanced *E. guineensis* samples. The reads generated from the 12 samples have good qualities and 96% of total filtered reads could be demultiplexed and assigned to the corresponding samples. After filtering and trimming, 84% of reads were retained for genome mapping and produced 5.7X to 10.42X genome coverage. Among the 34,410,224 identified SNPs, 98.7% of them were non-coding variants, and based on their location 69.1% of the total SNPs were intergenic SNPs. There were 5,618 SNPs out of the total generated SNPs that were validated using targeted genotyping by sequencing on 500 individuals, and 74% of the SNPs produced high quality calls in at least 95% of the samples. *Principal*

Penulis yang tidak disertai dengan catatan kaki instansi adalah peneliti pada Pusat Penelitian Kelapa Sawit

Sri Wening (✉)  
Pusat Penelitian Kelapa Sawit  
Jl. Brigjen Katamso No. 51 Medan 20158  
Email: sriwening.sw@gmail.com

<sup>1</sup>Balai Besar Penelitian dan Pengembangan Bioteknologi dan Sumber Daya Genetik Pertanian; Jl. Tentara Pelajar No. 3A, Bogor 16111

*component analysis using the SNPs correctly identified each sample's genetic background. Thus, the SNPs had high quality and could be used to develop marker-assisted selection of the E. guineensis x E. oleifera breeding population.*

**Keywords:** SNP, resequencing, *Elaeis guineensis*, *Elaeis oleifera*, targeted genotyping by sequencing, principal component analysis

## PENDAHULUAN

*Elaeis guineensis* Jacq. adalah spesies kelapa sawit yang tumbuh atau dibudidayakan pada daerah tropis khatulistiwa, yaitu di Afrika, Asia Tenggara, Amerika Tengah dan Amerika Selatan (Corley dan Tinker, 2016). Untuk melestarikan ekosistem kelapa sawit, usaha pemenuhan kebutuhan minyak sawit dan turunannya perlu dilakukan dengan meningkatkan produktivitas per hektar, daripada melalui cara ekstensifikasi lahan. Meskipun demikian, saat ini pasar tidak hanya memerlukan bahan tanaman yang bersifat unggul dalam hasil, tetapi juga memerlukan sifat yang lain seperti kecepatan meninggi yang rendah, serta mutu minyak yang tinggi. *Elaeis oleifera* (HBK) Cortes adalah spesies lain kelapa sawit yang berasal dari Amerika Tengah dan Amerika Selatan, yang memiliki pertumbuhan meninggi yang rendah serta kadar asam lemak tak jenuh yang tinggi (Corley dan Tinker, 2016). Introgresi sifat tersebut dari *E. oleifera* ke *E. guineensis* yang memiliki produktivitas tinggi, menjadi perhatian pemulia kelapa sawit. *E. guineensis* sering digunakan sebagai tetua penerima, sementara *E. oleifera* digunakan sebagai tetua donor, sehingga, *E. oleifera* disilangkan dengan *E. guineensis*, kemudian F1 yang dihasilkan disilangkan kembali dengan *E. guineensis* menghasilkan generasi Back Cross (Silang balik) 1 (BC1). Silang balik kemudian dilakukan berulang kali kepada tetua *E. guineensis*. Melalui proses ini, dilakukan seleksi individu hasil persilangan yang memiliki sifat yang diinginkan dari *E. oleifera* dan *E. guineensis* (Rahmadi *et al.*, 2006; Siregar *et al.*, 2018; Yaakub *et al.*, 2020; Ernayunita *et al.*, 2020). Tetapi metode tersebut membutuhkan areal penanaman yang luas, biaya mahal, tenaga banyak, dan memerlukan waktu yang lama karena kelapa sawit memiliki siklus hidup yang lama. *Marker-assisted selection* (MAS) dapat digunakan untuk

mempercepat program pemuliaan tersebut, karena melalui MAS, individu yang memiliki sifat tertentu dapat dipilah secara dini melalui penggunaan marka DNA (Jenkins *et al.*, 2012).

Marka DNA seperti AFLP (*Amplified Fragment Length Polymorphism*) (Wang *et al.*, 2011; Zhang *et al.*, 2015) dan SSR (*Simple Sequence Repeat*) (Jenkins *et al.*, 2012; Lebedev *et al.*, 2020) dapat digunakan dalam MAS. SNP (*Single Nucleotide Polymorphism*) yang memiliki keunggulan karena kehandalan, kemudahan untuk otomatisasi, dan jumlah pada genome dapat juga digunakan dalam MAS (Singh dan Singh, 2015). SNP telah banyak digunakan pada analisis genetik seperti analisis keragaman genetik, pemetaan genetik dan *association mapping* (Kumar *et al.*, 2012; Riahi *et al.*, 2013). *Next Generation Sequencing* (NGS) yang bersifat *high-throughput* (Behjati dan Tarpey, 2013) memungkinkan menghasilkan marka SNP dalam waktu singkat dan dengan biaya yang rendah (Cortés *et al.*, 2011; Van *et al.*, 2013). Teknologi tersebut dapat memfasilitasi usaha menghasilkan marka dengan jumlah banyak pada genom kelapa sawit yang polimorfik pada suatu populasi, yang memungkinkan identifikasi marka yang berhubungan dengan sifat yang menguntungkan secara agronomi, melalui Genome-Wide Association Study (GWAS) (Brachi *et al.*, 2011; Kumar *et al.*, 2012; Luo *et al.*, 2019). Identifikasi marka tersebut dapat digunakan pada seleksi individu yang memiliki sifat yang diinginkan pada *E. oleifera* dan *E. guineensis* pada suatu program pemuliaan kelapa sawit.

Karena GWAS memerlukan data genotipik dan fenotipik yang bersegregasi, informasi situs SNP yang polimorfik di antara individu yang memiliki sifat fenotipik yang berbeda menjadi sangat penting. Pengurutan (*sequencing*) *de novo* kelapa sawit telah dilakukan pada spesies kelapa sawit (*E. guineensis* dan *E. oleifera*) (Singh, 2013; Jin *et al.*, 2016), yang menghasilkan genom rujukan untuk analisis genetik kelapa sawit. Genom rujukan yang dikombinasikan dengan data yang dihasilkan dari pengurutan kembali (*resequencing*) individu-individu perwakilan pada suatu program pemuliaan kelapa sawit, dapat digunakan untuk mengenali marka DNA yang polimorfik dan handal, yang dapat digunakan untuk tujuan seleksi. Pengurutan kembali kelapa sawit menggunakan

perwakilan suatu program pemuliaan telah dilakukan oleh Kwong *et al.* (2016). Keeratan hubungan dan keterwakilan individu yang diurutkan dengan populasi program pemuliaan sangat penting untuk meningkatkan peluang terkenalnya marka DNA yang handal. Selain itu, cakupan pengurutan yang tinggi pada genom kelapa sawit akan meningkatkan peluang mengidentifikasi marka DNA yang mewakili semua kromosom genom kelapa sawit. Penelitian ini bertujuan untuk mengidentifikasi marka SNP genom, melalui pengurutan kembali perwakilan suatu populasi program pemuliaan hibrida *E. guineensis* x *E. oleifera*, yang akan berguna untuk pengembangan kegiatan MAS populasi tersebut.

## BAHAN DAN METODE

### Material genetik, ekstraksi DNA dan persiapan pustaka DNA

DNA diekstraksi dari daun muda 12 individu tanaman contoh (Tabel 1), menggunakan DNeasy<sup>®</sup> Plant Mini Kit (Qiagen) sesuai dengan tata cara dari produsen. DNA yang memiliki mutu baik dipakai pada persiapan pustaka DNA untuk pengurutan genom. Persiapan pustaka DNA dilakukan oleh Genting Laboratory Services Sdn. Bhd. (Malaysia) menggunakan TruSeq<sup>®</sup> DNA PCR-Free Library Prep (Illumina, Inc.) sesuai tata cara dari produsen. Pustaka DNA yang dihasilkan kemudian dibuktikan dan dikuantifikasi menggunakan qPCR dan Bioanalyzer.

### Pengurutan Seluruh Genom

Pengurutan seluruh genom dilakukan oleh Genting Laboratory Services Sdn. Bhd. (Malaysia). Gabungan pustaka genom dari 12 contoh diurutkan menggunakan HiSeq 2500 (Illumina, Inc.), melalui pengurutan dua arah (*paired-end reads*), dengan panjang pembacaan DNA 2x100 pasang basa. Proses demultipleks dilakukan, kemudian *read* (urutan basa yang “dibaca”/merupakan keluaran HiSeq 2500) ditempatkan kepada contoh awal. *Read* yang tidak dapat ditentukan contohnya kemudian dikelompokkan sebagai *read* yang tidak dapat ditentukan.

Semua *read* asli dipetakan ke genom PhiX174 genome (accession no. NC\_001422.1) dengan

bantuan Bowtie version 0.12.7 untuk menghilangkan *read* yang berasal dari pustaka PhiX174, yang digunakan untuk *quality control* (Langmead, 2010). Kemudian, *read* disaring pada nilai mutu 25 ( $\geq Q25$ ) dan dipangkas hingga panjang *read* minimal 72 bp ( $\geq 72$  bp) untuk menghilangkan artefak dan *read* dengan mutu rendah. Mutu *read* dikaji menggunakan FastQC version 0.10.1 (Andrews, 2010). *Read* yang disaring dan dipangkas kemudian dipetakan ke genom *E. guineensis* dan *E. oleifera* (Tabel 2) dengan bantuan Burrows-Wheeler Aligner (BWA) version 0.5.9rc1 (Li dan Durbin, 2009). Cakupan basa pada *read* yang dipetakan dari pengurutan menggunakan HiSeq 2500 diduga dan digambarkan menggunakan Qualimap version 2.2.1 (Okonechnikov *et al.*, 2015). Persentase pemetaan terhadap genom rujukan dihitung menggunakan SAMtools version 0.1.16 dengan opsi *flagstat* (Li *et al.*, 2009).

### Penamaan dan pemilihan SNP

Penamaan SNP dilakukan pada *read* yang dipetakan terhadap genom rujukan MPOB EG5 dan MPOB EO8 (Tabel 2) menggunakan Genome Analysis Tool Kit (GATK) version 3.3 (McKenna *et al.*, 2010). Penamaan SNP dibatasi untuk lokus dengan *depth*  $\geq 4$  *read* (yang dipetakan) dan mutu skor basa  $\geq Q25$ . Semua SNPs yang dihasilkan dengan menggunakan genom rujukan MPOB EG5 dianotasi dan pengaruhnya diduga menggunakan SnpEff version 3.6c (Cingolani *et al.*, 2012). Pengaruh SNP yang dihasilkan melalui penggunaan genom rujukan MPOB EO8 tidak ditentukan karena kurangnya informasi anotasi genom. Pengaruh SNP dikelompokkan berdasarkan kategori dampak, kelas fungsional, serta tipe dan lokasi SNP.

Kandidat marka untuk *targeted genotyping by sequencing* dipilih dari SNP yang diperoleh melalui rujukan genom MPOB EG5. SNP disaring dan dipilih berdasarkan kriteria: tidak ada data genotipe hilang untuk dua contoh *E. oleifera* dan dua contoh *E. guineensis*, hanya bi-alel SNP yang dipilih, dan SNP setidaknya memenuhi satu dari kriteria yang terdapat pada Tabel 8. SNP yang terseleksi selanjutnya disaring lebih lanjut hingga tinggal sekitar 5.000, dengan memperhatikan dampaknya. Sekuen genom pengapit di sebelah kiri dan kanan pada SNP

Tabel 1. Daftar individu contoh untuk pengurutan kembali  
Table 1. List of individual samples for resequencing

Nomor contoh	Latar belakang genetik
1	<i>E.o.</i> Brazil (tipe liar)
2	<i>E.o.</i> Suriname (tipe liar)
3	F1 - <i>E.o.</i> Brazil x <i>E.g.</i> Afrika
4	F1 - <i>E.o.</i> Brazil x <i>E.g.</i> Afrika
5	F1 - <i>E.o.</i> Suriname x <i>E.g.</i> Afrika
6	F1 - <i>E.o.</i> Suriname x <i>E.g.</i> Afrika
7	BC1 - ( <i>E.o.</i> Brazil x <i>E.g.</i> Afrika) x <i>E.g.</i> Deli
8	BC1 - ( <i>E.o.</i> Brazil x <i>E.g.</i> Afrika) x <i>E.g.</i> Deli
9	BC1 - ( <i>E.o.</i> Suriname x <i>E.g.</i> Afrika) x <i>E.g.</i> Afrika
10	BC1 - ( <i>E.o.</i> Suriname x <i>E.g.</i> Afrika) x <i>E.g.</i> Afrika
11	<i>E.g.</i> (grup A RRS)
12	<i>E.g.</i> (grup B RRS)

Keterangan: *E.o.*: *E. oleifera*, *E.g.*: *E. guineensis*, RRS: populasi *Reciprocal Recurrent Selection*

Tabel 2. Informasi genom rujukan kelapa sawit (Singh, 2013)  
Table 2. Information of oil palm reference genomes (Singh, 2013)

Nama	Spesies	Ukuran Genom (bp)	Jumlah Scaffolds	Anotasi	Versi
MPOB EG5	<i>E. guineensis</i>	1.535.150.282	16 kromosom, 40.056 scaffold	Tersedia	EG5 (28 Juni 2013)
MPOB EO8	<i>E. oleifera</i>	1.402.851.853	26.769 scaffold	Tidak tersedia	EO8 (24 Juli 2013)

yang terseleksi kemudian digunakan untuk mendesain primer pada tahap *targeted genotyping by sequencing*.

### **Targeted Genotyping by Sequencing**

Sebanyak 5.618 SNP yang dipilih digunakan untuk *genotyping* populasi pemuliaan *E. guineensis* x *E. oleifera*, menggunakan jasa *targeted genotyping by sequencing* SeqSNP dari LGC Group (UK). Informasi sekuen pengapit SNP yang dipilih ditabulasi dalam bed.file sesuai tata cara dari LGC Group (Biosearch Technologies, 2020a). *Genotyping* dilakukan pada 500 individu contoh dari populasi pemuliaan *E. guineensis* x *E. oleifera*. Contoh daun dari individu tersebut disiapkan sesuai tata cara yang diberikan oleh LGC Group (UK), dan dikirim ke institusi tersebut untuk dilakukan analisis SeqSNP (Biosearch Technologies, 2020b).

### **Sebaran SNP pada Genom**

Untuk visualisasi sebaran SNP genom, setiap kromosom dipisah ke dalam bagian-bagian yang masing-masing sepanjang 900.000 bp, dan jumlah SNP di tiap bagian tersebut dihitung. Data tersebut kemudian diplot dalam histogram menggunakan Circos (Krzywinski *et al.*, 2009).

### **Principal component analysis**

Data hasil analisis SeqSNP (dalam *variant call format*) yang belum disaring, kemudian digunakan sebagai data untuk *Principal Component Analysis* (PCA) untuk menghitung dekomposisi *eigenvalue* dari matriks *covariance* pada TASSEL 5.2.51 (Bradbury *et al.*, 2007). *Eigenvalue* dari *principal components* kemudian dipetakan menggunakan CurlyWhirly 1.17.08.31 (<https://lics.hutton.ac.uk/curlywhirly/>) (Milne *et al.*, 2014).

## **HASIL DAN PEMBAHASAN**

### **Seleksi Genotipe dan Pengurutan**

Pada deteksi hubungan marka dan fenotipe dalam kajian *association mapping*, pemilihan contoh plasma nutfah merupakan hal yang penting, selain mutu data genotipe dan fenotipe, pemilihan analisis statistika

yang sesuai, serta verifikasi asosiasinya (Ibrahim *et al.*, 2020). Pada kajian ini, contoh yang dipilih terdiri dari tetua *E. oleifera* tipe liar dari Suriname dan Brazil, individu *E. guineensis* maju, dan berbagai hasil persilangan meliputi F1 hibrida interspesifik dan *pseudo-backcross* (Tabel 1). Populasi hibrida yang digunakan merupakan keturunan *E. oleifera* Brazil dan Suriname, sehingga tetua *E. oleifera* tipe liar dari kedua daerah tersebut digunakan sebagai contoh pada pengurutan kembali. Digunakannya kedua spesies tipe liar tersebut juga berdasarkan hasil penelitian Faizah *et al.* (2016), yang menunjukkan adanya kelompok genetik pada kedua orijin *E. oleifera* pada populasi pemuliaan kelapa sawit hibrida OxG. Penggunaan kedua tipe liar, hibrida interspesifik dan *pseudo-backcross* akan meliputi hampir semua alel yang ada pada populasi pemuliaan, dengan fenotipe yang bersegregasi yang disebabkan oleh introgresi dari beberapa alel *E. oleifera* ke dalam genom *E. guineensis*. Diharapkan, semua keragaman genetik teridentifikasi pada pengurutan kembali menggunakan *next generation sequencing* (Jiang *et al.*, 2016), dan pemilihan contoh plasma nutfah tersebut akan mendukung teridentifikasinya SNP.

### **Pengurutan Seluruh Genom dan Sifat SNP**

Dua belas contoh yang diurutkan menggunakan HiSeq 2500 menghasilkan 275 Gb keluaran data mentah (Tabel 3). Proporsi data urutan yang bermutu baik cukup tinggi, karena 86,9% dari data mentah tersebut memiliki skor mutu setidaknya 30. Proporsi read dengan Q<sub>≥</sub>30 juga dilaporkan pada pengurutan *Beta vulgaris* dan *Arabidopsis thaliana* menggunakan HiSeq 2000 (Minoche *et al.*, 2011).

Sebanyak 96% dari keseluruhan *read* yang tersaring dapat ditentukan jenis contohnya (Tabel 4). Tabel 4 juga menunjukkan jumlah kuantitas keluaran pengurutan oleh HiSeq 2500, sebelum dan sesudah tahap penyaringan dan pemangkasan, dimana terdapat 84% data yang dapat digunakan dari keluaran HiSeq 2500. Setelah tahap demultipleks, dimana setiap *read* dapat diurutkan dan ditentukan contohnya, jumlah *read* yang tersaring dan terpangkas pada tiap contoh berkisar dari 154.152.050 hingga 202.091.495.

Penyaringan dan pemangkasan urutan basa menghasilkan *read* yang bermutu baik, dan sesuai



untuk penjajaran ke genom rujukan. Data penjajaran menunjukkan bahwa cakupan *read* terendah adalah sebesar 5,7X untuk *read* dari contoh *E. guineensis* (ID contoh nomor 12, Tabel 1) yang dipetakan ke genom MPOB EO8, sementara cakupan *read* tertinggi adalah sebesar 10,42X untuk *read* dari contoh *E. guineensis* (ID contoh nomor 11, Tabel 1) yang dipetakan ke genom MPOB EG5. Persentase GC berkisar dari 36,86% untuk *read* dari contoh *E. guineensis* (ID contoh nomor 12, Tabel 1) yang dipetakan ke genom MPOB EO8, hingga 38,98% untuk *read* dari *E. oleifera* tipe liar (ID contoh nomor 1, Tabel 1) yang dipetakan ke genom MPOB EO8. *Read* dari dua contoh *E. guineensis* memiliki persentase pemetaan yang paling tinggi pada genom *E. guineensis* (MPOB EG5), diikuti berturut-turut oleh contoh BC1, F1 dan *E. oleifera*. Sebaliknya, dua contoh *E. oleifera* tipe liar memiliki persentase pemetaan yang tertinggi pada genom *E. oleifera* (MPOB EO8), diikuti berturut-turut oleh contoh F1, BC, dan *E. guineensis*. Perbedaan nilai

cakupan (5,7X hingga 10,42X) disebabkan karena perbedaan genom spesies yang digunakan dalam pemetaan. Hasil ini hampir serupa dengan hasil pengurutan kembali pada 17 individu kelapa sawit, yang menghasilkan *sequencing depth* sebesar 5 hingga 8 (Jin *et al.*, 2016). Sebagai perbandingan, pengurutan tomat menggunakan Illumina GAllx menghasilkan 95,4% hingga 98,8% *read* yang dipetakan dengan nilai cakupan berkisar antara 6,7X hingga 16,6X, dengan rerata cakupan sebesar 11,2X (Causse *et al.*, 2013), sementara pengurutan kembali gabungan dari 132 individu kelapa sawit menghasilkan cakupan sekitar 40X pada genom kelapa sawit (Kwong *et al.*, 2016). Adanya *read* yang tidak dapat dipetakan dapat disebabkan karena ketidakakuratan pada perakitan genom rujukan, kesalahan pengurutan, atau genom rujukan yang tidak lengkap (Lunter *et al.*, 2008). Penjajaran *read* kepada genom rujukan menggunakan BWA, sebuah algoritma/pendekatan yang sering digunakan (Hatem *et al.*, 2013), yang

Tabel 3. Ringkasan Keluaran Pengurutan Seluruh Genom Menggunakan HiSeq 2500  
Table 3. Output summary from the whole-genome sequencing runs menggunakan HiSeq

Parameter	Keluaran
Hasil Total (gigabasa)	275
Kerapatan Kelompok (K/mm <sup>2</sup> )	570 +/- 68 hingga 699 +/- 93
Kelompok Lolos Penyaringan (%)	85,0 +/- 3,2 hingga 91,3 +/- 1,4
<i>Read</i> (juta)	1.484,79
<i>Read</i> Lolos Penyaringan (juta)	1.334,76
Persentase ≥ Q30	86,9

telah menghasilkan hasil yang baik pada pemetaan *read* kepada genom rujukan (Kumar *et al.*, 2012). Mutu BWA pada penjajaran urutan basa juga telah dilaporkan oleh Wang *et al.* (2011).

Jumlah total SNP yang diidentifikasi dari semua *read* yang dipetakan terhadap MPOB EG5 dan MPOB EO8 adalah masing-masing sebanyak 34.410.224 dan 5.811.378. Jumlah SNP yang diidentifikasi melalui pemetaan *read* terhadap

MPOB EG5 jauh lebih banyak dibandingkan yang diperoleh Kwong *et al.* (2016), yang telah mengidentifikasi lebih dari 7 juta SNP dari pengurutan kembali 132 individu kelapa sawit. Perbedaan tersebut kemungkinan disebabkan oleh perbedaan keragaman alel pada contoh yang digunakan serta perbedaan pada cakupan genom kelapa sawit. Kwong *et al.* (2016) melakukan pengurutan kembali yang lebih dalam, dengan cakupan genom kelapa sawit sebesar 40X.

Pengaruh setiap SNP hanya dapat dianalisis berdasarkan rujukan MPOB EG5, karena kurangnya anotasi genom pada genom rujukan MPOB EO8. Beberapa tipe pengaruh SNP ditunjukkan pada Tabel 5 hingga Tabel 7, berdasarkan kriteria dampak, kelas fungsional, serta tipe dan lokasi SNP genom rujukan MPOB EG5. Lebih dari 98% SNP merupakan *non-coding variants* (“*modifier*” effect). Sekitar 0,7% SNP memiliki dampak signifikan pada fungsi gen melalui perubahan *non-disruptive* yang mungkin dapat merubah efektivitas ekspresi protein (“*moderate*” impact). Sebagian kecil SNP memiliki

dampak pada protein yang diekspresikan, seperti menginduksi hilangnya fungsi melalui pembentukan stop kodon secara premature pada gen yang terpengaruh (“*high*” impact). SNP lainnya diduga tidak merubah protein (“*low*” impact). *Rasio missense* terhadap mutasi silent adalah sebesar 1,3737 dan 69,1% dari total SNP teridentifikasi berlokasi pada daerah *intergenic* (*non-coding*) (Tabel 7). Pada kajian yang dilakukan oleh Jin *et al.* (2016), dilaporkan bahwa sekitar 98% SNP yang ditemukan pada pengurutan kembali kelapa sawit berada pada daerah *non-coding*. Daerah *intergenic* mungkin berevolusi lebih cepat

Tabel 4. Keluaran Pengurutan HiSeq 2500 Sebelum dan Sesudah Penyaringan dan Pemangkasan  
Table 4. HiSeq 2500 sequencing output before and after filtering and trimming

Parameter	Sebelum penyaringan dan pemangkasan	Sesudah penyaringan dan pemangkasan:
Jumlah <i>read</i> yang ditentukan contohnya	2.559.833.790	2.150.173.493
Jumlah <i>read</i> yang tidak dapat ditentukan contohnya	109.704.374	83.443.904
Total jumlah <i>read</i>	2.669.538.164	2.233.617.397
% <i>read</i> yang dapat ditentukan contohnya / total <i>read</i> <sup>1</sup>	96%	96%
Cakupan contoh <sup>2</sup>	~10,36 – 13,42X	~8,56 – 11,22X

<sup>1</sup> Jumlah *read* yang ditentukan contohnya adalah *read* yang berhasil dilakukan *de-multiplex*

<sup>2</sup> Perkiraan cakupan contoh adalah berdasarkan ukuran genom *E. guineensis* (rata-rata sebesar 1,8 Gbp)

dibandingkan daerah *conserved* (Guo *et al.*, 2007) yang menyebabkan polimorfisme yang tinggi. Adanya variasi fungsional intergenik yang ditemukan oleh Rodgers-Melnick *et al.* (2016) menunjukkan potensi penemuan marka yang berhubungan dengan sifat, pada daerah tersebut.

Semua SNP kemudian dipilih berdasarkan delapan kriteria pada Tabel 8, yang memiliki peluang tinggi sebagai polimorfik pada populasi hibrida OxG. Tabel 8 menguraikan jumlah SNP yang memiliki

masing-masing kriteria, dimana terdapat total sejumlah 112.296 SNP sesuai dengan satu dari delapan kriteria yang ada.

Pada program hibrida OxG, marka DNA sangat penting untuk mempercepat pengembangan introgresi sifat superior dari *E. oleifera* kepada *E. guineensis*. Untuk mencapai hal ini, marka yang berasosiasi dengan sifat tersebut perlu diidentifikasi melalui pendekatan-pendekatan genetika, seperti GWAS, yang biasanya memerlukan SNP *bi-allelic*,

dengan tanpa data hilang pada populasi pemetaan. Kwong *et al.* (2016) juga melakukan strategi yang serupa saat menyeleksi kandidat SNP dari varian

yang ditemukan. Mereka menghilangkan varian yang dikategorikan sebagai *indels* dan *non-biallelic*. SNP bermutu tinggi yang teridentifikasi dengan

Tabel 5. Jumlah SNP sesuai dampaknya, untuk SNP yang diidentifikasi dari *read* yang dipetakan terhadap genom MPOB EG5

Table 5. The number of SNPs according to their impact, for SNPs called from mapped reads against MPOB EG5 genome

Tipe	Jumlah SNP	Persentase Jumlah SNP (%)
<i>High</i>	15.628	0,036
<i>Low</i>	247.174	0,576
<i>Moderate</i>	302.953	0,706
<i>Modifier</i>	42.345.004	98,682

Tabel 6. Jumlah SNP sesuai kelas fungsionalnya, untuk SNP yang diidentifikasi melalui *read* yang dipetakan terhadap genom MPOB EG5

Table 6. The number of SNPs according to their functional class, for SNPs called from mapped reads against MPOB EG5 genome

Tipe	Jumlah SNP	Persentase Jumlah SNP (%)
<i>Missense</i>	306.188	57,026
<i>Nonsense</i>	7.840	1,46
<i>Silent</i>	222.896	41,514

kriteria tersebut akan sesuai untuk berbagai analisis genetik, seperti analisis keragaman genetik, GWAS dan kajian evolusi (Xu dan Bai, 2015). Pada kajian ini, kriteria tambahan ditetapkan untuk memaksimalkan peluang untuk menemukan marka DNA yang polimorfik dan bersegregasi di populasi pemuliaan.

antaranya ditentukan genotipenya dengan  $\geq 8$  *reads* yang disejajarkan per contoh, setidaknya pada 95% dari 500 contoh, menurut analisis *quality control* pada laboratorium dan secara *in silico*. Tingkat *genotyping* yang tinggi (74%) menunjukkan bahwa tujuan untuk menemukan marka DNA yang polimorfik dan bersegregasi di populasi pemuliaan melalui kriteria saat seleksi SNP telah tercapai.

#### Targeted Genotyping by sequencing

Dari 112.296 SNP yang memenuhi kriteria polimorfisme seperti yang diuraikan pada Tabel 8, sejumlah 5.618 dipilih berdasarkan dampak SNP dan kemudian digunakan untuk *genotyping* menggunakan SeqSNP. Terdapat 4.969 dari 5.618 SNP tersebut (88%) yang dapat digunakan sebagai target pada *genotyping by sequencing* dan 3.690 (74%) di

Gambar 1 menunjukkan sebaran 3.690 SNP dengan mutu yang tinggi pada genom *E. guineensis*. Beberapa daerah kromosom memiliki kerapatan SNP yang lebih tinggi daripada lainnya, tetapi secara umum seluruh genom telah terwakili oleh SNP yang dipilih.

#### Principal component analysis

Diagram PCA menunjukkan bahwa *E. oleifera* tipe liar



Tabel 7. Jumlah SNP sesuai tipenya, untuk SNP yang diidentifikasi melalui reads yang dipetakan terhadap genom MPOB EG5

Table 7. The number of SNPs according to their type, for SNPs called from mapped reads against MPOB EG5 genome

Tipe	Jumlah SNP	Persentase Jumlah SNP (%)
<i>Downstream</i>	4.148.847	9,669
<i>Intergenic</i>	29.668.062	69,139
<i>Intron</i>	4.182.958	9,748
<i>Nonsynonymous coding</i>	302.953	0,706
<i>Nonsynonymous start</i>	125	0
<i>Splice site acceptor</i>	2.321	0,005
<i>Splice site donor</i>	2.357	0,005
<i>Splice site region</i>	24.153	0,056
<i>Start lost</i>	1.216	0,003
<i>Stop gained</i>	7.840	0,018
<i>Stop lost</i>	1.894	0,004
<i>Synonymous coding</i>	222.286	0,518
<i>Synonymous stop</i>	610	0,001
<i>Upstream</i>	4.345.137	10,126

dari Suriname, *E. oleifera* tipe liar dari Brazil, dan *E. guineensis* terpisah secara jelas pada kelompok yang berbeda (Gambar 2). Populasi F1 Brazil dan F1 Suriname berkelompok di antara *E. oleifera* tipe liar dan *E. guineensis*, sementara populasi BC1 lebih dekat secara genetik terhadap kelompok *E. guineensis* daripada terhadap populasi F1. Populasi F1 memiliki keragaman genetik tertinggi dibandingkan dengan populasi lain.

#### Pembuktian SNP

*Targeted genotyping by sequencing* menggunakan analisis SeqSNP (LGC Group, UK) dan PCA dapat dianggap sebagai tahap pembuktian SNP yang diidentifikasi melalui pengurutan kembali genom. Hasil menunjukkan bahwa secara umum SNP yang

dikembangkan memiliki mutu yang tinggi. Hal tersebut dapat dibuktikan dari tingginya persentase SNP yang berhasil digunakan dalam *genotyping* menggunakan SeqSNP (sekitar 74%), yang hasilnya lebih tinggi dibandingkan dengan tingkat yang diperoleh Islam *et al.* (2015) yaitu sebanyak 66,7%. Salah satu kunci sukses pada penentuan genotipe melalui *targeted genotyping by sequencing* adalah mutu urutan basa pada potongan DNA yang digunakan untuk mengenali SNP yang nyata. Daerah pengapit situs SNP digunakan untuk merancang primer (Islam *et al.*, 2015) yang kemudian digunakan pada tahap amplifikasi dan penentuan alel daerah target. Pembuktian SNP melalui pendekatan amplifikasi potongan DNA yang terdapat SNP juga telah

dilakukan oleh Hazzouri *et al.* (2015) dan Van *et al.* (2013).

SNP yang diidentifikasi melalui pengurutan kembali memiliki mutu yang tinggi, yang juga dibuktikan berdasarkan lebih tingginya polimorfisme SNP yang ditemui di populasi hibrida OxG saat dilakukan *genotyping* menggunakan SeqSNP, dimana jumlah contoh yang dianalisis lebih besar daripada jumlah contoh yang digunakan pada tahap pengurutan kembali.

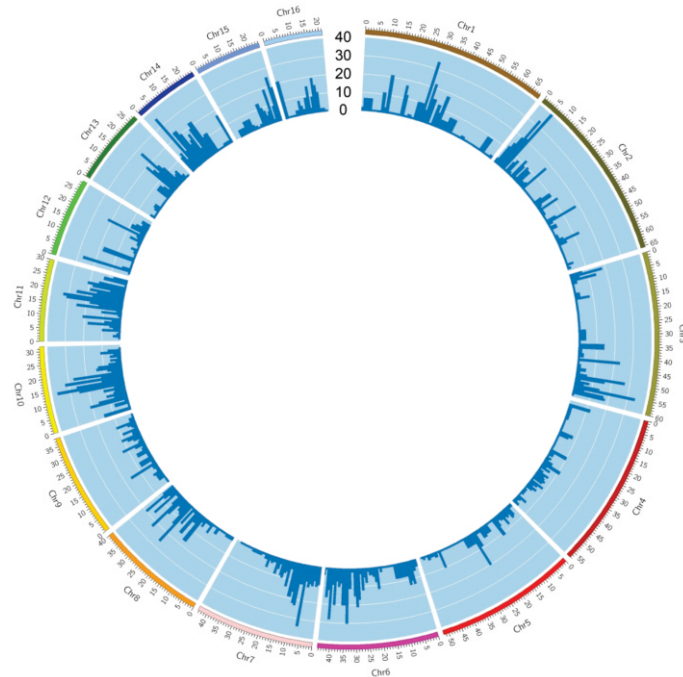
Walaupun digunakan hasil SeqSNP yang belum dilakukan penyaringan, hasil PCA yang menggunakan

data tersebut sesuai dengan informasi silsilah populasi hibrida OxG. *E. guineensis* terpisah dengan *E. oleifera* (sesuai dengan spesiasi), sementara *E. oleifera* tipe liar dari Suriname dan *E. oleifera* tipe liar dari Brazil masing-masing berada pada kelompok yang terpisah, yang sesuai dengan letak geografis asal usulnya. Sesuai dengan latar belakang persilangan, populasi F1 berada di antara *E. oleifera* dan *E. guineensis*, sementara populasi BC1 berada di antara contoh F1 dan *E. guineensis*. Walaupun SNP yang dipilih adalah yang polimorfik antara kedua *E. o.* dan polimorfik antara *E. g.* dan *E. o.*, tetapi hasil tersebut sesuai dengan hasil kajian Faizah *et al.* (2016) yang

Tabel 8. Kriteria yang digunakan untuk pemilihan SNP dan jumlah SNP yang dipilih  
 Table 8. Criteria used in shortlisting SNP and Number of Selected SNP

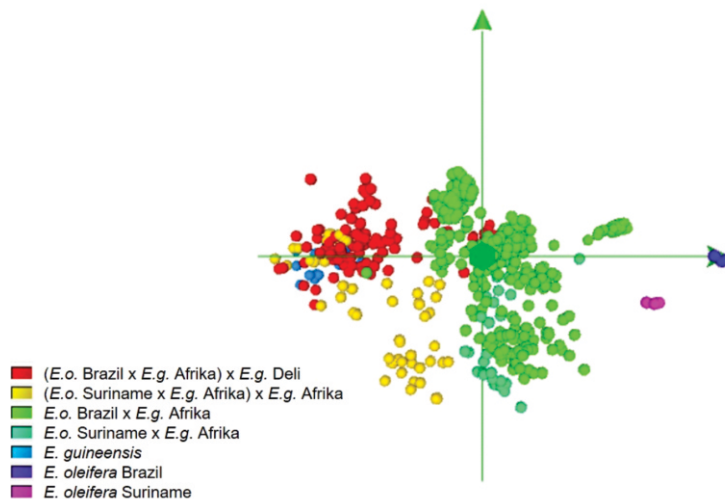
No.	Kriteria	Jumlah SNP
1	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> . Tidak polimorfik di dalam <i>E.o.</i> dan <i>E.g.</i>	49.774
2	Polimorfik antara contoh <i>E.o.</i> tanpa memperhatikan genotipe <i>E.g.</i>	5.018
3	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> . Polimorfik dalam <i>E.g.</i> , tidak polimorfik dalam <i>E.o.</i> (homozigot).	57.504
4	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> . Polimorfik dalam <i>E.o.</i> , tidak polimorfik dalam <i>E.g.</i> (homozygot).	8
5	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> . Polimorfik dalam <i>E.g.</i> dan tidak polimorfik dalam <i>E.o.</i> (heterozigot).	0
6	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> . Polimorfik dalam <i>E.o.</i> , tidak polimorfik dalam <i>E.g.</i> (heterozigot).	32
7	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> ; heterozigot pada tetua <i>E. oleifera</i> (WTB0 dan WTS0): AG-AG; genotipe <i>E.g.</i> tidak menjadi pertimbangan	8
8	Polimorfik antara <i>E.o.</i> dan <i>E.g.</i> ; polimorfik antara WTB0 dan WTS0: AA-GG atau AG-GG; genotipe <i>E.g.</i> tidak menjadi pertimbangan	40

Keterangan: *E.o.*: *Elaeis oleifera*, *E.g.*: *Elaeis guineensis*, WTB0: *E.o.* Brazil, WTS0: *E.o.* Suriname



Gambar 1. Histogram sebaran SNP yang dipilih untuk *targeted genotyping by sequencing* (SeqSNP), pada genom kelapa sawit. Setiap batang berwarna biru tua menunjukkan jumlah SNP setiap interval 900.000 bp pada kromosom

Figure 1. Histogram of the distribution of SNPs selected for *targeted genotyping by sequencing* (SeqSNP) across the oil palm genome. Each dark-blue bar indicates the number of SNPs in a 900,000 bp interval in a chromosome



Gambar 2. Plot 2 Dimensi PCA Populasi Pemuliaan Hibrida OxG

Keterangan: Merujuk ke tabel 1

Figure 2. Two dimensional PCA plot of OxG hybrid breeding population

Remarks: Refer to Table 1

melakukan kajian keragaman genetik populasi hibrida OxG menggunakan marka SSR, tanpa informasi polimorfisme di antara contoh sebelum dilakukan analisis. Informasi pengelompokan genetik tersebut sangat penting sebelum melakukan analisis GWAS, untuk menghindari *false positive* pada saat mengenali marka (Ong *et al.*, 2018).

*Genotyping by sequencing* telah digunakan pada berbagai kajian genom (Huang *et al.*, 2014). Dengan fasilitas *next generation sequencing* yang mampu melakukan analisis yang bersifat *high throughput*, *genotyping by sequencing* merupakan salah satu pendekatan yang dapat digunakan untuk kajian *genome-wide association* karena cakupan genom yang tinggi dan keefektifan dalam penggunaan biaya (Deschamps *et al.*, 2012).

## KESIMPULAN

Pengurutan kembali perwakilan populasi pemuliaan hibrida OxG telah berhasil mengidentifikasi SNP genom kelapa sawit. Pembuktian melalui *targeted genotyping by sequencing* menunjukkan tingginya persentase jumlah SNP yang dapat digunakan, adanya polimorfisme SNP pada contoh serta hasil PCA yang sesuai dengan informasi silsilah dan letak geografis asal usul contoh. Hal tersebut merupakan bukti bahwa pengurutan kembali telah menghasilkan SNP genom kelapa sawit yang bermutu tinggi. SNP yang telah teridentifikasi dapat digunakan untuk pengembangan marka DNA, guna diperbantukan dalam program pemuliaan populasi tersebut.

## UCAPAN TERIMA KASIH

Penelitian ini didanai oleh Badan Pengelola Dana Perkebunan Kelapa Sawit. Penulis mengucapkan terima kasih kepada ACGT Sdn. Bhd. atas analisis pengurutan kembali dan analisis bioinformatika selama identifikasi SNP. Penulis juga mengucapkan terima kasih kepada Balai Besar Penelitian dan Pengembangan Bioteknologi dan Sumber Daya Genetik Pertanian, serta Pusat Penelitian Kelapa Sawit, atas dukungan yang diberikan selama penelitian.

## DAFTAR PUSTAKA

- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Behjati, S., P.S.Tarpey. 2013. What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice* 98: 236–238. <http://dx.doi.org/10.1136/archdischild-2013-304340>
- Biosearch Technologies. 2020a. SeqSNP service guidance notes. <https://biosearchassets.blob.core.windows.net/assetsv6/guidance-notes-seqsnp-service.pdf>
- Biosearch Technologies. 2020b. Targeted genotyping by sequencing service - SeqSNP. <https://www.biosearchtech.com/services/sequencing/targeted-genotyping-by-sequencing-seqsnp>
- Brachi, B. G.P. Morris, J.O. Borevitz. 2011. Genome-wide association studies in plants: the missing heritability is in the field. *Genome biology* 12: 232. <https://doi.org/10.1186/gb-2011-12-10-232>
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Causse, M., N. Desplat, L. Pascual L, M.-C, Le Paslier, C. Sauvage, G. Bauchet, A. Bérard, R. Bounon, M. Tchoumakov, D. Brunel, J.-P Bouchet. 2013. Whole genome resequencing in tomato reveals variation associated with introgression and breeding events. *BMC Genomics* 14: 791. <https://doi.org/10.1186/1471-2164-14-791>
- Cingolani, P., A. Platts, I.L. Wang, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, D.M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6: 80–92. <https://doi.org/10.4161/fly.19695>
- Corley, R.H.V., P.B. Tinker. 2016. The oil palm (5<sup>th</sup>

- edition). John Wiley & Sons Ltd., Chichester, UK.
- Cortés, A., M. Chavarro, M. Blair. 2011. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 123: 827–845. <https://doi.org/10.1007/s00122-011-1630-8>
- Deschamps, S, V. Llaca, G.D. May. 2012. Genotyping-by-Sequencing in Plants. *Biology (Basel)* 1(3): 460–483. <https://doi.org/10.3390/biology1030460>
- Ernayunita, Taryono, P. Yudono. 2020. Histochemistry of backcross 1 of oil palm seeds at different storage periods. *Ilmu Pertanian (Agricultural Science)* 5(1): 52–57. <https://doi.org/10.22146/ipas.51776>
- Faizah R, S. Wening, H.Y. Rahmadi, A.R. Purba. 2016. Keragaman genetik populasi *E. oleifera* dan populasi *E. guineensis* x *E. oleifera* pada koleksi plasma nutfah PPKS. *Jurnal Penelitian Kelapa Sawit* 24(1): 13–23. <https://doi.org/10.22302/iopri.jur.jpks.v24i1.3>
- Guo, X, Y. Wang, P.D. Keightley, L. Fan. 2007. Patterns of selective constraints in noncoding DNA of rice. *BMC Evolutionary Biology* 7(1) 208. doi: 10.1186/1471-2148-7-208
- Hatem, A., D. Bozdağ, A.E. Toland, Ü.V. Catalyürek. 2013. Benchmarking short sequence mapping tools. *BMC Bioinformatics* 14. 184. <https://doi.org/10.1186/1471-2105-14-184>
- Hazzouri, K.M., J.M. Flowers, H.J. Visser, H.S.M. Khierallah, U. Rosas, G.M. Pham, R.S. Meyer, C.K. Johansen, Z.A. Fresquez, K. Masmoudi, N. Haider, N.E. Kadri, Y. Idaghdour, J.A. Malek, D. Thirkhill, G.S. Markhand, R.R. Krueger, A. Zaid, M.D. Purugganan. 2015. Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Communications* 6(1): 1–11. doi: 10.1038/ncomms9824
- Huang, Y.F., JA. Poland, C.P. Wight, E.W. Jackson, N.A. Tinker. 2014. Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. *PLoS One* 9(7): e102448. <https://doi.org/10.1371/journal.pone.0102448>
- Ibrahim, A.K., L. Zhang, S. Niyitanga, M.Z. Afzal, Y. Xu, L. Zhang, L. Zhang, J. Qi. 2020. Principles and approaches of association mapping in plant breeding. *Tropical Plant Biology* 13. 212–224. <https://doi.org/10.1007/s12042-020-09261-4>
- Islam, M.S., G.N. Thyssen, J.N. Jenkins, D.D. Fang. 2015. Detection, validation, and application of genotyping-by-sequencing based single nucleotide polymorphisms in upland cotton. *Plant Genome* 8(1): 1–10. DOI:10.3835/plantgenome2014.07.0034
- Jenkins, J.N., J.C. McCarty, M.J. Wubben, R. Hayes, O. A. Gutierrez, F. Callahan, D. Deng. 2012. SSR markers for marker assisted selection of root-knot nematode (*Meloidogyne incognita*) resistant plants in cotton (*Gossypium hirsutum* L.). *Euphytica* 183:49–54. DOI:10.1007/s10681-011-0470-6
- Jiang, Z., H. Wang, J.J. Michal, X. Zhou, B. Liu, L.C.S. Woods, R.A. Fuchs. 2016. Genome Wide Sampling Sequencing for SNP Genotyping: Methods, Challenges and Future Development. *International Journal of Biological Science* 12(1): 100–108. doi: 10.7150/ijbs.13498. eCollection 2016
- Jin, J., M. Lee, B. Bai, Y. Sun, J. Qu, Rahmadsyah, Y. Alfiko, C.H. Lim, A. Suwanto, M. Sugiharti, L. Wong, J. Ye, N.-H. Chua, G.H. Yue. 2016. Draft genome sequence of an elite Dura palm and whole-genome patterns of DNA variation in oil palm. *DNA Research* 23: 527–533. doi: 10.1093/dnares/dsw036
- Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra. 2009. Circos: an Information Aesthetic for Comparative Genomics. *Genome Research*. 19:1639–1645. doi: 10.1101/gr.092759.109
- Kumar, S., T.W. Banks, S. Cloutier. 2012. SNP Discovery through Next-Generation Sequencing and Its Applications. *International Journal of Plant Genomics Article ID 831460*. <https://doi.org/10.1155/2012/831460>
- Kumar, S., F.M. You, S. Cloutier. 2012. Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *BMC Genomics* 13: 684. <https://doi.org/10.1186/1471-2164-13-684>



- Kwong, Q.B., C.K. The, A.L. Ong, H.Y. Heng, H.L. Lee, M. Mohamed, J.Z. Low, S. Apparow, F.T. Chew, S. Mayes, H. Kulaveerasingham, M. Tammi, D.R. Appleton. 2016. Development and Validation of a High-Density SNP Genotyping Array for African Oil Palm. *Molecular Plant* 9 : 1 1 3 2 - 1 1 4 1 .  
<https://doi.org/10.1016/j.molp.2016.04.010>
- Langmead, B. 2010. Aligning short sequencing reads with Bowtie. *Current Protocols Bioinformatics* 32 : 11.17.11-11.17.14. doi: 10.1002/0471250953.bi1107s32
- Lebedev., V.G., N.M. Subbotina, O.P. Maluchenko, T.N. Lebedeva, K.V. Krutovsky, K.A. Shestibratov. 2020. Transferability and Polymorphism of SSR Markers Located in Flavonoid Pathway Genes in *Fragaria* and *Rubus* Species. *Genes* 11:11. <https://doi.org/10.3390/genes11010011>
- Letunic, I., P. Bork. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research* 47(W1): W256-W259. doi: 10.1093/nar/gkz239
- Li, H., R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25: 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, Subgroup 1000 GPD. 2009. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 25: 2078-2079. doi: 10.1093/bioinformatics/btp352
- Lunter, G., A. Rocco, N. Mimouni, A. Heger, A. Caldeira, J. Hein. 2008. Uncertainty in homology inferences: assessing and improving genomic sequence alignment. *Genome Research* 18 : 298 - 309 . doi : 10.1101/gr.6725608
- Luo, Z., P. Tomasi, N. Fahlgren, H. Abdel-Haleem. 2019. Genome-wide association study (GWAS) of leaf cuticular wax components in *Camelina sativa* identifies genetic loci related to intracellular wax transport. *BMC Plant Biology* 19(1): 187. <https://doi.org/10.1186/s12870-019-1776-0>
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernysky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M.A. DePristo. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297-1303. DOI: 10.1101/gr.107524.110
- Milne, I., G. Stephen, M. Bayer, P.D. Shaw, S. Raubach, S. Hearne, S. Singh, P. Wenzl, D. Marshall. 2014. Graphical applications for visualization and analysis of genotype data sets. *Plant and Animal Genome XXII*, San Diego, 2014.
- Minoche, A.E., J.C. Dohm, H. Himmelbauer. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biology* 12:R112. <https://doi.org/10.1186/gb-2011-12-11-r112>
- Okonechnikov, K., A. Conesa, F. García-Alcalde. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32: 292-294. DOI: 10.1093/bioinformatics/btv566
- Ong, P.W., I. Maizura, M. Marhalil, N. Rajanaidu, N.A.P. Abdullah, M.Y. Rafii, L.C.L. Ooi, E.T.L. Low, R. Singh. 2018. Association of SNP markers with height increment in MPOB-Angolan natural oil palm populations. *Journal of Oil Palm Research* 30(1): 61-70. DOI:10.21894/jopr.2018.0003
- Rahmadi, Y.H., N. Supena, H.A. Siregar, Sujadi, A.R. Purba. 2006. Keragaan vegetatif, kualitas minyak dan karakter tandan hasil silang balik BC-1 antara *Elaeis oleifera* x *Elaeis guineensis*. *Jurnal Penelitian Kelapa Sawit* 14(3): 171- 182.
- Riahi, L., N. Zoghalmi, A. Fournier-Level, A. Dereeper, L. Le Cunff, V. Laucou, A. Mliki, P. This. 2013. Characterization of single nucleotide polymorphism in Tunisian grapevine genome and their potential for population genetics and evolutionary studies. *Genetic Resources and Crop Evolution* 60: 1139-1151. DOI:10.1007/s10722-012-9910-y
- Rodgers-Melnick, E., D.L. Vera, H.W. Bass, E.S. Buckler. 2016. Open chromatin reveals the functional maize

- genome. Proceedings of the National Academy of Sciences 113(22), E3177–E3184. <https://doi.org/10.1073/pnas.1525244113>
- Singh, B.D., A.K. Singh. 2015. Marker-Assisted Plant Breeding: Principles and Practices. Springer.
- Singh, R. 2013. Oil Palm Genome Programme - An Update MPOB Oil Palm Genome Programme. PIPOC, Kuala Lumpur. 2013.
- Siregar, H.A., H.Y. Rahmadi, S. Wening, E. Suprianto. 2018. Komposisi asam lemak dan karoten kelapa sawit *Elaeis oleifera*, interspesifik hibrida, dan pseudo-backcross pertama di Sumatera Utara, Indonesia. Jurnal Penelitian Kelapa Sawit 26 (2): 91-101.
- Van, K., Y.J. Kang, K.-S. Han, Y.-H. Lee, J.-G. Gwag, J.-K. Moon, S.-H. Lee. 2013. Genome-wide SNP discovery in mungbean by Illumina HiSeq. Theoretical and Applied Genetics 126(8): 2017-2027. DOI: 10.1007/s00122-013-2114-9
- Wang, D.-W., Z.-Q. Li. 2011. Identification of a Male-Specific Amplified Fragment Length Polymorphism (AFLP) and a Sequence Characterized Amplified Region (SCAR) Marker in *Eucommia ulmoides* Oliv. International journal of molecular sciences 12(1): 857-864. DOI: 10.3390/ijms12010857
- Wang, W., Z. Wei, T.-W. Lam, J. Wang. 2011. Next generation sequencing has lower sequence coverage and poorer SNP-detection capability in the regulatory regions. Scientific Reports 1(55): 1-7. DOI:10.1038/srep00055
- Xu, X., G. Bai. 2015. Whole-genome resequencing: changing the paradigms of SNP detection, molecular mapping and gene discovery. Molecular Breeding 35(1): 33. DOI:10.1007/s11032-015-0240-6
- Yaakub, Z., K. Kamaruddin, R. Singh, S. Mustafa, M. Marjuni, N.-C. Ting, M.D. Amiruddin, L.E.-T. Leslie, O.L. Cheng-Li, K. Sritharan, R. Nookiah, J. Jansen, M.O. Abdullah. 2020. An integrated linkage map of interspecific backcross 2 (BC2) populations reveals QTLs associated with fatty acid composition and vegetative parameters influencing compactness in oil palm. BMC Plant Biology 20. 356. <https://doi.org/10.1186/s12870-020-02563-5>
- Zhang, P., H. Zhou, C. Lan, Z. Li, D. Liu. 2015. An AFLP marker linked to the leaf rust resistance gene LrBi16 and test of allelism with Lr14a on chromosome arm 7BL. The Crop Journal 3 (2): 152-156. DOI:10.1016/j.cj.2014.11.004

