

# SEKUENSING GENOM UNTUK KARAKTERISASI SUMBER DAYA GENETIK TANAMAN

*I Made Tasma, Sustipriatno, dan Mastur*

*Balai Besar Penelitian dan Pengembangan Bioteknologi dan Sumber Daya Genetik Pertanian*

*Jl. Tentara Pelajar 3A, Bogor 16111, Indonesia*

## PENDAHULUAN

Berkembangnya teknologi sekuensing modern yang dikenal dengan istilah *next generation sequencing* (NGS) telah merevolusi studi genomika dari berbagai spesies tanaman, hewan, dan mikroba pertanian. Teknologi ini mampu menghasilkan data sekuen dalam jumlah besar dalam waktu singkat dengan biaya murah (Ansorge 2009; Varshney *et al.* 2009) sehingga terjangkau oleh berbagai lembaga penelitian termasuk institusi penelitian di negara berkembang yang memungkinkan banyak lembaga dapat mengkarakterisasi sumber daya genetik (SDG) tanamannya dengan lebih komprehensif pada level sekuen genom.

Indonesia sangat kaya akan SDG tanaman yang perlu dikarakterisasi lebih detil dan dimanfaatkan semaksimal mungkin untuk program pemuliaan tanaman. Banyak spesies tanaman pertanian penting sumber keragamannya ada di Indonesia. Padi, ubi jalar, pisang, durian, manggis, berbagai spesies tanaman rempah seperti cengkeh dan pala, dan beraneka jenis tanaman

obat adalah beberapa contoh tanaman pertanian penting asli Indonesia. Keragaman genetik dari SDG tanaman penting pertanian khususnya yang asli Indonesia tersebut merupakan sumber daya pemuliaan bernilai tinggi, bahan baku industri dan merupakan aset nasional yang tidak ternilai harganya. Keragaman SDG ini sangat penting untuk perakitan dan pengembangan kultivar baru tanaman untuk pangan, pakan, sandang untuk mendukung ketahanan pangan, energi dan industri nasional.

Pesatnya perkembangan teknologi sekuensing seperti NGS akhir-akhir ini, mempercepat tersedianya sekuen genom rujukan berbagai spesies tanaman yang memfasilitasi resekuensing (menyekuen ulang anggota spesies) SDG tanaman pada level sekuen genom. Dengan NGS saat ini lebih dari 60 spesies tanaman pertanian penting telah tersedia sekuen genom rujukannya (Van *et al.* 2012; Tasma 2015). Dari spesies tanaman tersebut masih belum ada tanaman asli Indonesia kecuali padi yang telah disekuen lengkap genomnya. Banyak spesies tanaman unik Indonesia dimana Indonesia merupakan pusat asal (*center of origin*) seperti cengkeh, durian, manggis, dan kelapa belum tersedia sekuen rujukannya dan teknologi sekuensing modern ini akan memfasilitasi penyelesaian sekuen rujukan tanaman-tanaman asli Indonesia.

Ketersediaan sekuen genom rujukan sebagai sekuen kontrol/pembanding mempermudah karakterisasi SDG tanaman melalui resekuensing genom menggunakan NGS. Proyek resekuensing genom dilakukan untuk mengungkap informasi genom suatu sampel/koleksi SDG potensial dengan memanfaatkan sekuen rujukan yang telah ada sebelumnya (Tasma 2016a; 2016d). Informasi genotipe individu SDG memfasilitasi percepatan penemuan berbagai variasi genetik termasuk gen, alel, dan genotipe unggul untuk mendukung program pemuliaan tanaman yang terarah, lebih cepat, dan lebih akurat menggunakan metode pemuliaan paradigma baru, yaitu pemuliaan tanaman berbasis data geno-

mika. Program pemuliaan tanaman ke depan akan lebih berfokus pada komparasi komposisi genom individu plasma nutfah yang membuka peluang penggunaan kombinasi strategi baru pemetaan genetik dan analisis evolusi untuk penemuan dan pemanfaatan variasi genetik SDG tanaman dengan lebih optimal untuk menunjang program pemuliaan tanaman yang sinambung dan lestari. Disamping itu, hasil sekuensing genom juga dapat menghasilkan lokus unik untuk SDG tertentu sehingga dapat digunakan sebagai sidik jari (*fingerprint*) dari SDG penting yang dapat berfungsi untuk perlindungan varietas sekaligus alat konfirmasi kemilikan SDG.

Makalah ini membahas peran teknologi sekuensing genom untuk mengkarakterisasi SDG tanaman secara komprehensif pada level sekuen genom total sehingga informasi genotipe dari setiap aksesi plasma nutfah akurat dan komprehensif untuk mendukung program pemuliaan tanaman berbasis data genomika (*genomic-based breeding program*) dan juga untuk mendukung perlindungan varietas dan SDG tanaman.

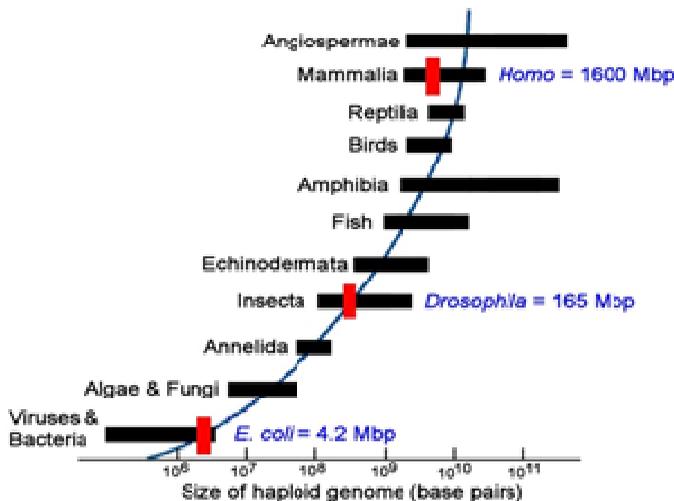
### ***Genom, Genomika, dan Pemuliaan Tanaman***

Genom adalah satu set kromosom beserta seluruh gen dan konstitusi genetik lainnya yang terkandung di dalam satu sel suatu organisme (Wetterstrand 2011). Genom juga dapat didefinisikan sebagai satu set lengkap cetak biru kehidupan yaitu satu set utuh semua DNA kromosom haploid pada satu inti sel suatu organisme. Genomika adalah studi tentang genom yang meliputi semua sekuen nukleotida termasuk struktur gen (*structural genes*), sekuen regulator (*regulatory sequences*), dan segmen DNA non koding (*non-coding DNA segments*) pada kromosom suatu organisme. Pemuliaan tanaman adalah seni dan pengetahuan untuk memanipulasi genetik tanaman untuk menghasilkan karakteristik yang diinginkan. Teknik pemuliaan tanaman sudah

lama dipraktekan untuk memilih individu tanaman yang mempunyai karakteristik yang ditargetkan seperti ukuran buah, tingkat kemanisan, dan warna yang menarik untuk buah-buahan; bulir yang panjang, rasa nasi enak, dan bergizi tinggi untuk padi; hingga produktivitas tinggi dan berkeragaan pendek pada gandum; serta kandungan protein tinggi dan produktivitas tinggi, tahan berbagai cekaman biotik di daerah tropis pada kedelai.

Kromosom pada suatu genom sebetulnya adalah DNA. Struktur DNA adalah untai ganda (*double helix*) terdiri dari unsur gula (*sugar*), fosfat, empat nukleotida (adenin, timin, guanin, dan sitosin), dimana adenin berpasangan dengan timin dihubungkan oleh dua ikatan hidrogen (*hydrogen bond*) dan guanin berpasangan dengan sitosin dihubungkan oleh tiga ikatan hidrogen. DNA genomik terdiri dari area fragmen gen dan area fragmen non gen. Gen ditranskripsi menjadi RNA oleh RNA polimerase, dan RNA diproses untuk menghasilkan *messenger RNA* (mRNA) yang ditranslasi menjadi protein. Protein tunggal (produk satu gen) atau protein kombinasi (kombinasi protein dari produk beberapa gen) menentukan kenampakan fenotipe suatu karakter. Karakter superior menjadi target pemulia tanaman untuk digabungkan ke individu superior lain tetapi belum memiliki karakter target untuk menghasilkan kultivar dengan beberapa karakter unggul untuk digunakan oleh petani.

Ukuran genom dari berbagai organisme sangat bervariasi dari beberapa juta basa (MB) sampai dengan sekitar 16 milyar basa (GB) pada tanaman poliploid seperti gandum dan tebu (Gambar 1.1).



**Gambar 1.1.** Ukuran genom berbagai organisme sangat bervariasi dari organisme bersel tunggal (bakteri) sampai dengan organisme dengan genom kompleks angiospermae yang diantara anggotanya memiliki genom poliploid (Wetterstrand 2011)

Ukuran genom tanaman juga sangat bervariasi tergantung tingkat kompleksitas genom dari setiap spesies tanaman (Varsney *et al.* 2009). Kebanyakan spesies tanaman adalah diploid yaitu genomnya memiliki hanya dua set DNA genomik. Padi dan kakao sebagai contoh, memiliki genom diploid dengan ukuran genom relatif kecil, masing-masing sekitar 450 MB (Tabel 1.13). Beberapa spesies tanaman memiliki beberapa *copy* DNA kromosom yang sama (*multiple copy of the same chromosomal DNA*) atau yang lebih dikenal dengan genom poliploid. Tanaman gandum sebagai contoh, memiliki 6 *copy* genom (hexaploid) dengan ukuran genom besar sekitar 16 GB (Tabel 1.13). Semakin kompleks suatu genom semakin sulit untuk menangani dan menggabungkannya satu dengan yang lainnya. Kompleksitas dan ukuran genom besar menjadi tantangan tersendiri bagi peneliti genomika. Sangat sulit untuk melakukan analisis *assembly* (penyambungan

**Tabel 1.13.** Ukuran genom pada inti sel dari berbagai spesies tanaman pertanian (Dhanapal 2012)

Jenis tanaman	Nama latin	Ukuran genom (Mb) <sup>a</sup>
Gandum	<i>Triticum aestivum</i>	15.966
Bawang merah	<i>Allium cepa</i>	15.290
Garden pea	<i>Pisum sativum</i>	3.947
Jagung	<i>Zea mays</i>	2.292
Asparagus	<i>Asparagus officinalis</i>	1.308
Kedelai	<i>Glycine max</i>	1.110
Tomat	<i>Lycopersicum esculentum</i>	907
Sugarbeat	<i>Beta vulgaris</i>	758
Apel	<i>Malus X domestica</i>	743
Kacang buncis	<i>Phaseolus vulgaris</i>	637
Cantaloupe	<i>Cucumis melo</i>	434
Anggur	<i>Vitis vinifera</i>	483
Kakao	<i>Theobroma cacao</i>	430
Padi	<i>Oryza sativa</i>	450

<sup>a</sup>1 Mb = 1.000.000 bp

sekuen pendek menjadi sekuen dengan ukuran kromosom utuh) pada genom tanaman dengan ukuran dan kompleksitas yang tinggi tersebut.

### **Perkembangan teknologi sekuensing DNA**

Sekuensing DNA atau pengurutan basa DNA merupakan teknik kunci untuk berbagai perkembangan ilmu pengetahuan di antaranya genetika, bioteknologi, biologi molekular dan genomika (Franca *et al.* 2002). Sekuensing DNA adalah teknik untuk menentukan urutan basa nitrogen (adenin, guanin, sitosin dan timin) suatu sampel DNA. Salah satu contoh aplikasi ambisius teknologi sekuensing DNA yaitu pengurutan genom manusia melalui proyek yang dikenal *Human Genome Project*. Manfaat dari diketahuinya sekuen lengkap genom manusia, maka dapat didiagnosis dan selanjutnya dikembangkan pengobatan untuk penyakit genetik.

Awal tahun 1970 merupakan perkembangan awal dari sekuensing DNA dengan metode yang digunakan yaitu metode kromatografi. Perkembangan selanjutnya mulai diperkenalkan metode sekuensing DNA dengan menggunakan metode *dye based sequencing* (Olsvik 1993). Sampai dengan akhir abad ke 20, alat yang dapat membaca urutan basa DNA umumnya hanya mampu menentukan urutan basa DNA yang panjangnya kurang dari 1.000 basa, sehingga untuk proyek sekuensing genom, molekul DNA harus terlebih dulu dipotong menjadi ukuran yang lebih kecil dan dapat dibaca oleh alat. Data urutan DNA yang dihasilkan kemudian disusun ulang menggunakan program komputer sehingga urutan keseluruhan DNA dalam inti sel dapat diketahui (Gambar 1.2).

Sebelum ditemukannya alat yang mampu membaca urutan DNA secara paralel dalam jumlah besar, perunutan basa-basa DNA (*sequencing*) memakan waktu, tenaga, dan biaya yang sangat besar. Pemerintah Amerika Serikat menginvestasikan sekitar 3,8 milyar dollar (5,6 milyar dollar jika kita memperhitungkan inflasi) untuk mencari urutan basa DNA manusia yang besarnya sekitar 3 milyar pasang basa, yang baru dapat di selesaikan pada tahun 2003 atau sekitar 13 tahun dari waktu proyek tersebut mulai dicanangkan. Setelah ditemukannya alat pembaca DNA paralel generasi berikutnya, biaya serta waktu yang diperlukan untuk mendapatkan data yang sama dapat dipangkas menjadi beberapa bulan saja. Namun besarnya data yang diperoleh memerlukan fasilitas komputer dan tenaga ahli bioinformatika yang cukup signifikan untuk memprosesnya. Pusat-pusat penelitian genom terkemuka di dunia seperti Broad Institute, Washington University, maupun Beijing Genome Institute memiliki akses ke komputer yang berkapasitas besar serta ratusan ahli bioinformatika untuk mengelola dan memproses data sekuen DNA yang diperoleh.

Metode sekuensing DNA yang pertama kali digunakan (*first generation sequencing*) dikenal dengan metode Sanger (*Sanger dideoxy sequencing*). Metode ini menggunakan DNA templat dan memerlukan primer spesifik untuk reaksi sekuensing. Teknologi sekuensing dari Sanger ini dan modifikasinya mendominasi metode sekuensing selama 30 tahun (Sanger *et al.* 1997). Panjang sekuen yang dihasilkan 1000–1200 pasang basa (bp) dan tidak mampu mencapai lebih dari 2000 bp. Untuk mendapatkan sekuen lebih panjang dikembangkan metode sekuensing *shotgun* yang dikembangkan pada proyek sekuensing genom manusia. Pada metode ini templat DNA dipotong dengan enzim restriksi lalu fragmen DNA diklon pada vektor sekuensing dan individu fragmen DNA setiap klon disekuen terpisah. Hasil sekuen lengkap dari fragmen DNA yang panjang dapat diperoleh dengan mensejajarkan (*alignment*) dan menyambung (*assembly*) sekuen fragmen DNA berdasarkan bagian sekuen yang tumpang tindih (*overlap*) yang memungkinkan pemetaan genom manusia terselesaikan (Gambar 1.2).

Generasi kedua teknologi sekuensing dikenal dengan istilah *next generation sequencing* (NGS) secara kolektif mendeskripsikan teknologi sekuensing selain teknologi sekuensing Sanger seperti yang diuraikan pada paragraf sebelumnya. Teknologi ini mempunyai potensi untuk mengurangi biaya sekuensing. Filosofi dasar pembentukan mesin NGS diadaptasi dari metode sekuensing *shotgun* (Venter *et al.* 2003; Shendure *et al.* 2008). Teknologi NGS membaca templat DNA secara acak (*random*) sepanjang seluruh genom. Ini dilakukan dengan membuat potongan-potongan pendek DNA kemudian menyambungkannya dengan adapter (potongan DNA pendek didesain khusus untuk tujuan ini) untuk dibaca oleh mesin NGS secara random selama proses sintesis DNA (*sequencing by synthesis*). Dengan demikian, teknologi NGS ini sering juga disebut dengan sekuensing paralel secara masif Panjang bacaan sekuen DNA yang dihasil-

kan dari mesin NGS jauh lebih pendek dibandingkan dengan yang dihasilkan oleh mesin sekuensing dengan metode Sanger. Saat ini NGS menghasilkan panjang sekuen DNA berkisar antara 50–500 bp. Karena panjang sekuen yang dihasilkan NGS pendek, sekuensing setiap fragmen DNA mesti dilakukan lebih dari sekali ukuran genom (*genome sequence coverage*). Sebagai contoh *sekuen coverage* suatu sekuensing genom 30 kali berarti fragmen-fragmen DNA pada genom tersebut disekuensing sebanyak 30 kali sehingga setiap fragmen DNA dalam genom dibaca mesin sebanyak 30 kali. Hal ini dilakukan untuk menjaga akurasi data hasil sekuensing. Inovasi teknologi NGS ini saat ini berkembang sangat cepat dan berbagai kemajuan telah dicapai.

Tiga teknologi NGS utama yang tersedia di pasar adalah *Roche/454 pyrosequencing platform* (2005), *Illumina Solexa polymerase sequencing platform* (2006), dan *ABI/SOLID ligase sequencing technology* (2007). Dibandingkan dengan metode sekuensing Sanger ketiga teknologi NGS ini menghasilkan data sekuen yang jauh lebih banyak pada sekali menjalankan alat. Dengan demikian alat ini dikenal dengan *high throughput sequencing platforms* (Ansorge 2009). Perbandingan karakteristik utama teknologi sekuensing beberapa teknologi NGS dan estimasi biaya melakukan sekuensing jika dibandingkan dengan metode sekuensing konvensional Sanger seperti disajikan pada Tabel 1.14. Alat Roche/454 menghasilkan data bacaan sekuen terpanjang tetapi menghasilkan kuantitas data sekuen terendah (*lowest throughput*) diantara ketiga alat. Jumlah sekuen basa yang dihasilkan Illumina/Solexa tertinggi (*highest throughput*) tetapi panjang bacaannya hanya sekitar 100 basa. Panjang bacaan yang dihasilkan ABI/SOLID terpendek, hanya sekitar 50 basa, akan tetapi tingkat kesalahan pembacaan DNA pada proses sekuensing (*error rate*) paling rendah. Teknologi NGS ini merevolusi metode sekuensing yang menghasilkan data sekuen luar biasa besar dalam waktu yang relatif singkat dibandingkan dengan metode sekuensing

**Tabel 1.14.** Perbandingan karakteristik utama teknologi sekuensing konvensional (Sanger) dan beberapa teknologi NGS serta estimasi biaya sekuensing (US \$ per Mbp) (Perez-de-Castro et al. 2012)

Teknologi sekuensing	Panjang bacaan ( <i>read length</i> , bp)	Mbp setiap menjalankan alat	Biaya (US \$/Mbp) <sup>a</sup>
Sanger	1000	0,001	3000,00
454 Roche	450	450,000	66,00
Illumina Hi-Seq2000	100	270.000,000	0,07
Solid 5500xl	50	270.000,000	0,07

Sanger. Dengan demikian diperlukan sumberdaya IT dan ahli Bioinformatika untuk menganalisis data yang dihasilkan sehingga menjadi bermanfaat untuk mendukung program pemuliaan komoditas tanaman utama Kementerian Pertanian.

Di samping alat sekuensing generasi kedua (NGS) telah dikenalkan juga alat sekuensing generasi ketiga (*third generation sequencing platforms*). Alat ini mampu menyekuen individu molekul DNA atau RNA. Yang termasuk kelompok ini antara lain *HeliScope*, *Ion Torrent*, *SMRT* (*single molecule real-time sequencing*), dan *Oxford Nanopore*. Keunggulan alat-alat ini adalah kemampuannya membaca urutan molekul DNA yang berasal dari molekul tunggal. Dengan demikian kebutuhan materi awal untuk sekuensing sangat sedikit (<1 ug) DNA atau RNA dan tidak diperlukan amplifikasi PCR sehingga dapat terhindar dari kesalahan (*bias*) yang disebabkan oleh amplifikasi PCR. Namun demikian, data sekuen yang dihasilkan (*sequence throughput*) dari alat sekuensing generasi ketiga tidak setinggi yang dihasilkan alat sekuensing dengan alat generasi kedua (NGS) dan tingkat kesalahan (*error rate*) yang dihasilkan dari metode sekuensing generasi ketiga masih tinggi. Dari keterbatasan tersebut, alat sekuensing generasi ketiga ini saat ini hanya digunakan untuk sekuensing organisme dengan ukuran genom kecil. Dengan demikian, dari aspek kuantitas data sekuen yang dihasilkan

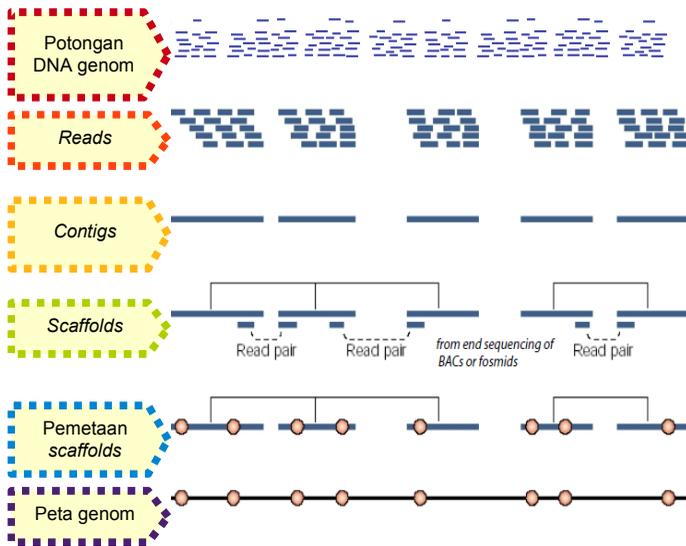
(*throughput*) dan biaya sekuensing, aplikasi alat sekuensing generasi ketiga ini saat ini masih belum dapat menandingi kehebatan alat sekuensing generasi kedua, NGS.

### ***Sekuensing De Novo Genom Tanaman untuk Pembentukan Sekuen Genom Acuan***

Sekuensing *de novo* (*de novo sequencing*) yaitu sekuensing genom dan menyambungkannya menjadi urutan basa kromosom utuh pertama kali. Untuk mengeksplorasi kandungan genetik total suatu spesies tanaman dan potensi genetiknya secara utuh diperlukan melakukan pengurutan susunan basa genom total (*whole genome sequencing*). Langkah awal untuk identifikasi urutan basa komprehensif tersebut, diperlukan adanya peta genom acuan (*reference genome sequence*) untuk setiap spesies tanaman. Sekuen genom acuan ini menjadi pedoman untuk studi lanjutan dari berbagai individu aksesi/genotipe anggota spesies tanaman tersebut. Pembentukan peta genom acuan tanaman ini tidak mudah karena begitu data sekuen diperoleh kita perlu mengurutkannya dengan susunan yang benar pada setiap kromosom. Sebagai pegangan dalam pengurutan susunan basa ini menjadi kromosom yang utuh diperlukan cukup informasi genetik spesies yang dibuat peta genom rujukannya. Informasi genetik yang diperlukan misalnya peta genetik spesies yang berisi cukup marka DNA dan marka lainnya yang mendefinisikan setiap kromosom (*linkage group*) dari peta genetik, sekuen pustaka genom *bacterial artificial chromosome* (BAC), informasi genetik spesies yang berkerabat dengan spesies yang sedang dipelajari, dan data genetik lainnya terkait dengan spesies tersebut.

Untuk menghasilkan dan menyusun dalam urutan yang tepat, sekuen genom sebuah spesies tanaman pertama kali (*de novo sequencing and de novo assembly*), DNA kromosom dipotong-potong menjadi miliaran potongan DNA dengan panjang ber-

kisar antara 500–30.000 pasang basa (bp), dan pembentukan berbagai jenis set pustaka genom (*complex DNA libraries*) dihasilkan dalam beberapa bulan. Biaya tinggi pada sekuensing dengan bacaan lebih panjang (*high cost long read sequence data*) diproduksi dari *complex DNA libraries* dan pemetaan *long contiguous chromosomal regions* dikonstruksi menggunakan *de novo assembly computer*. Sekuen yang panjang selanjutnya diurut sambungkan (*assembly*) menggunakan suatu program yang memungkinkan untuk menyambungkan puluhan ribu sekuen berulang (*repetitive element*) hingga akhirnya menghasilkan *sequence scaffolds* yang lebih panjang (Gambar 1.2). *Sequence scaffolds* yang lebih panjang jauh lebih informatif bagi peneliti genomika dan aplikasinya untuk mendukung program pemuliaan tanaman dibandingkan dengan *scaffolds* yang ukurannya lebih pendek.



**Gambar 1.2.** Skema pembentukan peta sekuen genom acuan tanaman (Tasma et al. 2012; Tasma 2015)

Gambar 1.2 memperlihatkan proses pembacaan sekuen genom total dengan teknik *shotgun*. Analisis sekuen dimulai dari potongan kecil (100 bp) dibuat *contigs*, *scaffolds*, dan peta genom utuh. Proses ini memerlukan waktu panjang antara lain karena mengisi celah (*gaps*) antar *contigs* dan *scaffolds* yang memerlukan sekuensing dan analisis data sekuen berulang

Sekali sebuah *sequence scaffold* panjang telah diperoleh, *express regions* (area sekuen yang mengandung gen) dari pada genom dianotasikan dengan jutaan sekuen pustaka cDNA (mewakili komunitas mRNA dari gen-gen dalam genom) yang juga diperoleh menggunakan *long read sequencing technologies*. Sangat mirip dengan *de novo genome sekuencing*, *long read sequencing technologies* untuk sekuensing cDNA lebih disukai dari pada *NGS short read sequencing technologies* untuk konstruksi draf pertama dari *transcriptome*, karena sekuen yang lebih panjang ketika disambung (*assembly*) lebih besar kemungkinannya menghasilkan sekuen transkrip utuh yang akan jauh lebih bermanfaat untuk gen model.

Sekuen genom acuan berkualitas tinggi (ukurannya panjang mendekati ukuran genom sesungguhnya) dan akurat (dengan tingkat kesalahan rendah) jauh lebih mahal memproduksinya dibandingkan dengan membuat data resekuensing genom yang sama. Lebih jauh, Tim Peneliti sekuensing genom *de novo* suatu spesies tidak pernah menyelesaikan sekuen acuan, akan tetapi Tim peneliti sekuensing genom *de novo* ini akan menyempurnakan terus sekuen genom acuan yang sudah di publikasikan sebelumnya. Sebagai contoh, walaupun sekuen genom acuan manusia pertama kali sudah diselesaikan dan dipublikasikan pada tahun 2002, sekitar 300 peneliti terus mengedit sekuen acuan tersebut dan sekarang telah diperoleh sekuen genom acuan generasi ke 19 (*its 19<sup>th</sup> build*). Semakin akurat sekuen genom acuan suatu spesies tanaman, akan menghasilkan data resekuensing yang semakin perkasa (*powerful*) dan akurat.

Untuk sekuensing suatu spesies tanaman, peneliti kadang-kadang memerlukan membuat 3-5 sekuen genom acuan dari beberapa genotipe dari spesies yang sama atau spesies yang berkerabat sangat dekat dengan spesies tanaman yang sedang dipelajari. Sebagai contoh, pada jagung (*Zea mays*) sekuen genom acuan sudah dibuat dari genotipe B73, MO17, dan spesies *popcorn* (Schnable *et al.* 2009; Calzada *et al.* 2009). Pada padi (*Oryza sativa*), sekuen genom acuan dibuat dari Niponbare, Indica dan Javanica (Yu *et al.* 2002). Pada kakao (*Theobroma cacao*), sekuen genom acuan dibuat dari Belizean Criollo dan dari kultivar Matina 1-6 (Argout *et al.* 2011). Pada kelapa sawit (*Elaeis guineensis*), sekuen genom acuan dibuat dari genotipe Dura, Pisifera, dan spesies *E. oleifera* (Singh *et al.* 2013). Pada kedelai, sekuen acuannya dibuat dari kedelai varietas Willians 82, suatu varietas yang secara genetik genomnya sudah terkarakterisasi dengan baik, mempunyai keunggulan karakter yang bernilai ekonomi, genomnya terpakai luas sebagai komponen pembentukan varietas unggul, dan pertimbangan lainnya sehingga data sekuen yang dihasilkan akan bernilai tinggi dan aplikasinya sedapat mungkin sangat luas. Jumlah genotip yang dipakai untuk membuat sekuen genom rujukan tergantung dari penggunaan data sekuen yang dihasilkan untuk aplikasi teknis terutama dari aspek program pemuliaannya. Sedapat mungkin peta genom acuan dapat mewakili genom-genom yang digunakan dalam program pemuliaan pembentukan varietas unggul mengandung karakter-karakter penting yang menjadi target pemuliaannya.

Berkembangnya teknologi sekuensing modern memacu penyelesaian peta sekuen acuan (peta rujukan) berbagai spesies tanaman pertanian penting di berbagai belahan dunia. Revolusi teknologi sekuensing DNA dengan teknologi NGS menurunkan biaya sekuensing genom yang memungkinkan menyekuen lebih banyak genom spesies tanaman dengan biaya yang lebih murah dan terjangkau (Varsney *et al.*, 2009; Tasma, 2015; Tasma, 2016a).

**Tabel 1.15.** Spesies tanaman pertanian penting Indonesia yang telah tersedia peta sekuen genom acuannya (Van *et al.* 2013; Tasma 2015)

Spesies tanaman	Nama latin	Ukuran genom (juta pasang basa, Mbp*)
Padi	<i>Oryza sativa</i> ssp <i>indica</i>	430,0
	<i>O. sativa</i> ssp <i>japonica</i>	430,0
Kedelai	<i>Glycine max</i>	1 115,0
Jagung	<i>Zea mays</i>	2 300,0
	<i>Zea mays</i> spp. <i>parviglumis</i>	2 100,0
Sorgum	<i>Sorghum bicolor</i>	730,0
Ubi kayu	<i>Manihot esculenta</i>	760,0
Pepaya	<i>Carica papaya</i>	372,0
Mentimun	<i>Cucumis sativus</i>	367,0
Kentang	<i>Solanum tuberosum</i>	844,0
Pisang	<i>Musa accuminata</i>	472,2
Apel	<i>Malus domestica</i>	742,3
Stroberi	<i>Fragaria vesca</i>	240,0
Anggur	<i>Vitis vinifera</i>	487,0
Semangka	<i>Citrullus lanatus</i>	450,0
Melon	<i>Cucumis melo</i> L.	430,0
Tomat	<i>Solanum lycopersicum</i>	900,0
Kacang buncis	<i>Phaseolus vulgaris</i>	486,9
Kakao	<i>Theobroma cacao</i>	430,0
Kapas	<i>Gossypium raimonddi</i>	750,0
Kelapa sawit	<i>Elaeis guineensis</i>	1.800,0
Jarak kepyar	<i>Ricinus communis</i>	350,0

\*1 Mbp (Mega base pair) = 10<sup>6</sup> base pairs

Saat ini setidaknya 60 spesies tanaman, kebanyakan tanaman budi daya bernilai ekonomi tinggi telah tersedia sekuen rujukan-nya (Danapal 2012; Van *et al.* 2013). Hal ini diawali dengan sekuen rujukan tanaman model *Arabidopsis thaliana* yang dipublikasi sekuen genomnya pada tahun 2000, diikuti oleh sekuen acuan padi pada tahun 2002, jagung pada tahun 2009, kedelai 2010, kakao 2010, kelapa sawit 2013 dan banyak lagi peta sekuen acuan berbagai spesies tanaman penting lainnya (Tabel 1.15). Penyelesaian peta genom rujukan suatu spesies tanaman biasanya dikerjakan oleh konsorsium internasional yang memacu *networking* dalam pembentukan dan pemanfaatan data genom yang dihasilkan oleh

anggota konsorsium secara bersama-sama. Pendanaannya juga biasanya melibatkan anggota konsorsium, umumnya konsorsium internasional.

### ***Resekuensing genom untuk identifikasi variasi genetik individu SDG tanaman***

Resekuensing artinya menyekuen ulang dalam konteks kita dapat menyekuen ulang genom tanaman anggota spesies tanaman yang genom rujukannya sudah tersedia. Sekuen genom rujukan ini digunakan sebagai kontrol/acuan untuk menganalisis data resekuen genom yang diperoleh berikutnya pada sampel tanaman kedua, ketiga, keempat, dan seterusnya. Data resekuen yang diperoleh dianalisis dengan cara disejajarkan dengan data sekuen genom rujukan untuk mengidentifikasi variasi genom pada beberapa sampel dari spesies tertentu yang dianalisis.

Dengan demikian, untaian sekuen pendek yang dihasilkan mesin NGS ( $\pm$  berukuran 55-150 bp) hasil resekuensing tidak perlu di-*assembly* satu dengan yang lainnya, akan tetapi dipetakan kembali ke sekuen genom acuan untuk mengidentifikasi perbedaan genetik yang terdapat antara sekuen acuan dan sekuen dari sampel individu kedua, ketiga dan seterusnya. Dengan demikian sejumlah perbedaan dapat didokumentasikan dan perbedaan genetik ini digunakan untuk menjelaskan perbedaan antar genotipe (varietas) dalam sebuah spesies yang berkaitan dengan karakter penting, seperti daya hasil, ketahanan hama dan penyakit, waktu pembungaan, toleransi kekeringan dan banjir, dan sebagainya.

Jenis perbedaan variasi genetik pada level genom yang umumnya diamati pada penelitian resekuensing antara lain variasi *single nucleotide polymorphism* (SNP), variasi *insertion* dan variasi *deletion*. Dua variasi yang terakhir sering disingkat dengan istilah Indel (*insertion and deletion*). SNP adalah sebuah perubahan komposisi nukleotida di sekuen DNA pada suatu posisi ter-

tentu pada sekuen tersebut. Inseri adalah terjadinya penyisipan satu jenis basa tertentu pada lokasi tertentu pada sekuen DNA genom. Sedangkan delesi adalah terjadinya penghapusan satu basa tertentu pada lokasi tertentu pada sekuen DNA genom tanaman.

Frekuensi terjadinya variasi genom tersebut pada genom tanaman bervariasi tergantung jenis tanaman dari aspek *sexual mode* (menyerbuk sendiri atau menyerbuk silang) dan jarak genetik genotipe tanaman yang disejajarkan (*aligned*) sekuennya dengan sekuen genom acuan. DNA pada tanaman menyerbuk silang biasanya memiliki frekuensi variasi genom yang lebih tinggi dibandingkan dengan tanaman menyerbuk sendiri. Penyejajaran sekuen genom genotipe tanaman dengan jarak genetik jauh menghasilkan frekuensi variasi genom yang lebih tinggi dibandingkan dengan penyejajaran data sekuen genom yang dihasilkan dari genotipe-genotipe tanaman berjarak genetik dekat.

Untuk itu pada penelitian dengan target penemuan SNP peneliti genomika akan menresekuen genotipe-genotipe yang jarak genetiknya jauh untuk mendapatkan cukup banyak varian SNP dari data resekuen yang dihasilkan. Di samping itu, semakin banyak jumlah genotipe yang diresekuen untuk penemuan variasi SNP maka semakin besar peluangnya mendapatkan lebih banyak variasi SNP. Sebagai contoh digunakan 20 genotipe padi dengan variasi luas (jarak genetik jauh) dan mengandung variasi fenotipe luas diantara genotipe terpilih untuk digunakan pada resequensing untuk mendapatkan 160,000 lokus SNP (Thompson 2014).

## **PERKEMBANGAN PENELITIAN RESEKUENSING GENOM SDG TANAMAN DI INDONESIA**

Penelitian resequensing genom telah dilakukan di Balai Besar Biogen, Balitbangtan sejak tahun 2010 hingga sekarang.

Resekuensing genom dilakukan pada beberapa SDG penting, seperti kedelai, jagung, cabai, pisang, kentang, kelapa sawit, kakao dan jarak pagar. Tujuan dari penelitian resekuensing ini adalah untuk mendapatkan variasi genom (SNP/Indel) pada genom SDG prioritas nasional yang selanjutnya dapat dimanfaatkan sebagai marka seleksi. Di antara penelitian resekuensing tersebut telah dilakukan pada lima genotipe kedelai varietas Indonesia terpilih dan lima varietas terpilih kakao asal Indonesia. Resekuensing dilakukan menggunakan mesin NGS HiSeq. Data resekuensing genom dari lima varietas kedelai Indonesia ini kemudian dipetakan ke peta sekuen genom acuan yaitu peta genom kedelai varietas Williams 82 (Schmutz *et al.* 2010). Sedangkan data resekuensing lima varietas kakao Indonesia juga dipetakan ke peta sekuen genom acuan kakao varietas Criollo (Argout *et al.* 2011).

Dari pemetaan data resekuensing lima varietas kedelai Indonesia tersebut telah diidentifikasi lebih dari 3,15 juta variasi genom (Tabel 1.16) yang terdiri dari sekitar 2,9 juta *single nucleotide polymorphism* (SNP) dan sekitar 250 ribu *insertion* dan *deletion* (indel) (Satyawan *et al.* 2014; Tasma *et al.* 2015a).

Secara rata-rata diidentifikasi satu variasi genom (SNP atau Indel) pada setiap 312 basa pada genom kedelai (Tabel 1.16). Di samping SNP dan Indel juga diidentifikasi variasi genom lain seperti simple sequence repeat (SSR) yang jumlahnya bisa mencapai beberapa ribu SSR. Berdasarkan peta sekuen genom acuan sebagai rujukan, data resekuensing ini dapat disejajarkan dengan data resekuensing hasil penelitian lain untuk mendapatkan lebih banyak variasi genom (SNP/Indel) sebagai sumber variasi (marka DNA) mendukung program pemuliaan tanaman kedelai.

Dari pemetaan data resekuensing lima varietas kakao Indonesia diidentifikasi lebih dari 2,68 juta variasi genom (Tabel 1.17) yang terdiri dari sekitar 2,33 juta SNP, dan sekitar 362 ribu indel (Tasma *et al.* 2016). Secara rata-rata diidentifikasi satu variasi

**Tabel 1.16.** Variasi genom yang terdeteksi pada setiap kromosom kedelai hasil analisis penyejajaran sekuen genom lima varietas kedelai Indonesia dengan sekuen genom rujukan kedelai varietas Williams 82 (Satyawan et al. 2014; Tasma et al. 2015a)

Kromosom*	Panjang kromosom (pb)	Jumlah basa yang berubah (pb)	Laju perubahan ( <i>change rate</i> )**
Gm01	55.915.595	164.050	340
Gm02	51.656.713	140.745	367
Gm03	47.781.076	182.784	261
Gm04	49.243.852	122.041	403
Gm05	41.396.504	102.278	410
Gm06	50.722.821	207.015	245
Gm07	44.683.157	137.522	324
Gm08	46.995.532	156.487	300
Gm09	46.843.750	152.358	307
Gm10	50.959.635	91.056	559
Gm11	39.172.790	115.306	339
Gm12	40.113.140	99.504	403
Gm13	44.408.971	168.210	264
Gm14	49.711.204	169.969	292
Gm15	50.939.160	218.143	233
Gm16	37.397.385	194.427	192
Gm17	41.906.774	164.865	254
Gm18	62.308.140	280.791	221
Gm19	50.589.441	148.520	340
Gm20	46.773.167	111.122	420
Total	972.068.482	3.150.869	312

\*Gm = *Glycine max* (L) Mer. \*\*Laju perubahan rata-rata = 312, artinya rata-rata ditemukan satu variasi DNA pada setiap 312 basa pada genom kedelai hasil penelitian ini. pb = pasang basa

genom (SNP atau Indel) pada setiap 121 basa pada genom kakao (Tabel 1.17).

Frekuensi kemunculan SNP/Indel pada kakao lebih tinggi dari pada kedelai karena kakao adalah tanaman menyerbuk silang sedangkan kedelai termasuk kelompok tanaman menyerbuk sendiri sehingga variasi sekuen DNA pada kedelai lebih rendah dibandingkan dengan variasi sekuen DNA kakao. Akan tetapi dari jumlah variasi yang dideteksi, resequencing lima varietas

**Tabel 1.17.** Variasi genom yang terdeteksi pada setiap kromosom kakao hasil analisis peninjauan sekuen genom lima varietas kakao Indonesia dengan sekuen genom rujukan kakao varietas Criollo (Tasma et al. 2016)

Kromosom*	Panjang kromosom (pb)	Jumlah basa berubah (pb)	Laju perubahan ( <i>change rate</i> )**
Tc00	108.886.888	835.679	130
Tc01	31.268.538	254.505	122
Tc02	27.754.001	232.265	119
Tc03	25.475.297	198.549	128
Tc04	23.504.306	166.738	140
Tc05	25.651.337	208.396	123
Tc06	15.484.475	138.079	112
Tc07	14.169.093	106.745	132
Tc08	11.535.834	111.952	103
Tc09	28.459.094	284.016	100
Tc10	15.164.258	151.245	100
Total	327.353.121	2.688.169	121

\*Tc = *Theobroma cacao*. \*\*Laju perubahan rata-rata = 121, artinya rata-rata ditemukan satu variasi DNA pada setiap 121 basa pada genom kakao hasil penelitian ini. pb = pasang basa

kedelai menghasilkan lebih banyak SNP dan Indel dibandingkan dengan resekuensing lima varietas kakao Indonesia. Hal ini karena materi genetik yang digunakan pada resekuensing mempunyai jarak genetik yang lebih besar pada kedelai dibandingkan dengan jarak genetik antar varietas kakao yang digunakan pada penelitian resekuensing.

Dari jutaan variasi genom yang diidentifikasi dari penelitian resekuensing di atas, sebagian besar variasi yang diamati berupa SNP dan sebagian kecil berupa Indel. SNP/Indel yang dideteksi dapat berlokasi dalam gen (*genic SNP/Indel*) atau di luar gen (*non genic SNP/Indel*). Dari hasil penelitian penemuan SNP/Indel pada kedelai dan kakao kebanyakan SNP/Indel yang dideteksi berlokasi di luar gen yaitu, di hulu (*upstream*), hilir (*downstream*) gen, dan di daerah antar gen (*intergenic regions*). Hanya sebagian kecil saja (berkisar sekitar 2–2,5%) dari SNP/Indel yang diidentifikasi berada pada *coding region* (*exon*). *Coding region* adalah *protein coding region*. SNP/Indel yang ada pada *exon* ini menjadi per-

hatian khusus dan target khusus dari peneliti genomika karena SNP/Indel ini mempengaruhi jenis protein yang dihasilkan oleh gen yang berlokasi pada SNP tersebut. Tergantung jenis basa yang berubah, SNP yang dimaksud dapat merubah susunan asam amino pada protein yang dihasilkan (*non synonymous SNP*) atau dapat juga tanpa merubah susunan susunan asam amino protein yang dihasilkan (*synonymous SNP*).

SNP yang merubah susunan asam amino dan produk protein yang dihasilkan oleh gen (*non synonymous SNP*) di atas, peluangnya besar akan merubah fenotipe tanaman karena perbedaan protein yang mungkin mengakibatkan proses fisiologi tanaman yang berbeda. Misalnya merubah reaksi tanaman dari peka penyakit tertentu menjadi tahan terhadap penyakit tersebut. Analisis lanjut dari SNP semacam ini sangat diperlukan yang dapat dilakukan menggunakan analisis genomika fungsional (*functional genomics*) untuk menguji hipotesis dan fungsi gen yang mengalami perubahan susunan DNA (SNP) tersebut. Gen yang ada pada SNP dianalisis mendalam sampai pembuktian fenotipe dari gen tersebut. Salah satu pendekatannya ialah dengan mentransformasikan fragmen gen pada SNP target ke genotipe *wild type* (tanpa SNP) untuk melihat fungsi gen terhadap fenotipe. Dengan demikian, pada penelitian penemuan SNP ini juga akan berujung pada isolasi gen (gen unggul) untuk digunakan pada perbaikan materi genetik tanaman melalui teknologi rekayasa genetika tanaman ataupun seleksi individu pembawa karakter target dengan teknik MAS.

Sampai dengan saat ini, secara keseluruhan sebanyak 8 spesies tanaman (kedelai, jagung, pisang, cabai merah, kakao, kelapa sawit, kentang, dan jarak pagar) telah dikarakterisasi variasinya pada level genom (Tabel 1.18). Jumlah variasi DNA yang diperoleh dari setiap spesies tanaman berkisar antara 92,007 (pada jarak pagar) sampai dengan 26,077,774 pada cabai dengan total variasi keseluruhan untuk delapan spesies tanaman sebanyak

**Tabel 1.18.** Variasi genetik hasil penjarangan data resekuen genom dengan sekuen genom rujukan dari setiap spesies tanaman yang diteliti (Tasma et al. 2014; Satyawan et al. 2014; Tasma et al. 2015b)

Spesies tanaman	Ukuran genom (Mb) <sup>*</sup>	Jumlah genotipe yang disequenced	Rataan <i>genome coverage</i> <sup>**</sup>	Frekuensi variasi DNA yang diperoleh <sup>**</sup>	Variasi DNA total yang dihasilkan		Total
					SNP	INDEL	
Kedelai ( <i>Glycine max</i> )	1,115	5	33	288	2,690,000	459,000	3,150,000
Jagung ( <i>Zea mays</i> )	2,300	4	27	634	2,805,145	240,526	3,247,037
Pisang ( <i>Musa accuminata</i> )	472.9	14	47	57	5,159,450	571,885	5,731,335
Kakao ( <i>Theobroma cacao</i> )	430	5	77	121	2,326,088	362,081	2,688,169
Kelapa sawit ( <i>Elaeis guineensis</i> )	1,800	3	109	197	3,032,200	303,109	3,335,331
Cabai merah ( <i>Capsicum annuum</i> )	2,649	6	69	101	24,760,787	1,316,987	26,077,774
Kentang ( <i>Solanum tuberosum</i> )	844	6	27	160	4,166,472	339,372	4,505,844
Jarak pagar ( <i>Jatropha curcas</i> )	450	3	125	1,185	92,007	nd	92,007
Total dan nilai kisaran		58	27-125	57-1,185	65,012,138	4,512,297	48,625,109

<sup>\*</sup>Mb = Mega base pairs (10<sup>6</sup> bp); <sup>\*\*</sup>Jumlah genom yang disequenced per spesies tanaman; <sup>\*\*\*</sup> One Satu variasi DNA diperoleh pada setiap fragmen (bp) pada genom; nd, belum tersedia data

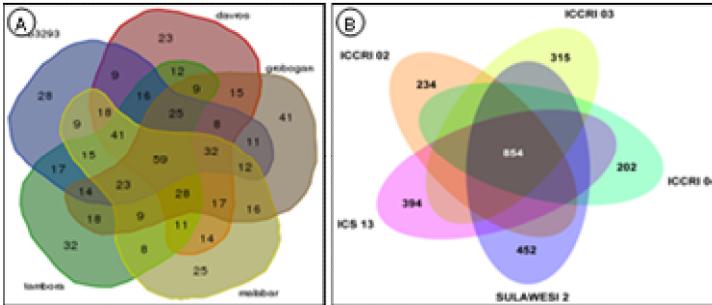


**Gambar 1.3.** Data base data genom yang bersisi jutaan variasi genetik hasil karakterisasi berbagai spesies tanaman prioritas nasional menggunakan NGS Hiseq2000 (Rijzaani et al. 2016)

48.625.109 variasi genetik (SNP dan INDEL). Data variasi genetik ini bernilai sangat tinggi yang sudah dikompilasi dalam bentuk database data genom dengan alamat web <http://genom.litbang.pertanian.go.id> (Gambar 1.3). Data base ini dapat diakses oleh publik.

Resekuensing berbagai galur unggul (*superior lines*) dan individu aksesi SDG mengidentifikasi keunikan dari setiap genotipe SDG tanaman. Hal ini akan menjadi identitas dari varietas, galur, maupun individu aksesi sebagai sidik jari dari individu tersebut.

Contoh dari keunikan ini disajikan pada Gambar 1.2 (Tasma *et al.* 2016; Tasma *et al.* 2018). Sekuensing 5 genotipe kedelai lokal Indonesia menghasilkan sekitar 2,6 juta marka SNP. Sebagian besar SNP dideteksi di luar gen. SNP di luar gen ini merupakan salah satu sumber daya pemuliaan bernilai tinggi diantaranya untuk pembentukan *system genotyping* berkapasitas tinggi (*high throughput genotyping*) seperti 50K SNP chip kedelai yang mengandung lebih dari 50.000 marka SNP. Hanya sebagian kecil marka SNP dideteksi pada gen. Dari *sampling* 337 SNP yang ada di exon, ditemukan 59 SNP yang dapat membedakan kelima varietas dari varietas acuan (Williams 82), dan beberapa alel SNP yang hanya ditemukan pada satu varietas kedelai saja dan berpotensi digunakan sebagai sidik jari untuk varietas tersebut. Ditemukan marka SNP umum dan marka SNP unik pada 5 genotipe kedelai. Genotipe B3293, Davros, Grobogan, Malabar, dan Tambora memiliki genotipe yang sama di 59 lokus SNP, yang berbeda dengan Williams82. Ada 10–16 SNP unik untuk setiap varietas yang dapat berfungsi sebagai sidik jari dari setiap varietas (Gambar 1.4). Hasil yang mirip juga ditemukan pada penelitian resekuensing 5 genotipe kakao lokal Indonesia. Kelima klon kakao yang diresekuensing memiliki variasi DNA yang sama di 854 lokus SNP, yang berbeda dengan DNA genom acuan Criolo. Ditemukan variasi DNA yang unik pada satu varietas saja, 234 lokus SNP untuk ICCRI 02, 315 lokus SNP untuk ICCRI 03, 202 lokus SNP untuk ICCRI 04, 452 lokus SNP untuk Sulawesi 2, 394 lokus SNP untuk ICS 13 (Tasma *et al.* 2012; Tasma *et al.* 2016). Marka SNP yang unik pada satu varietas dapat digunakan untuk karakterisasi dan sidik jari untuk proteksi varietas tersebut.



**Gambar 1.4.** Marka SNP umum dan unik untuk setiap varietas kedelai (Gambar A) dan SNP umum dan unik untuk setiap varietas kakao (Gambar B). Sumber: Tasma et al. 2016; Tasma et al. 2018)

## KESIMPULAN

Berkembangnya teknologi sekuensing modern seperti NGS memfasilitasi pembentukan sekuen genom rujukan berbagai spesies tanaman. Tersedianya peta genom acuan tersebut memungkinkan penelitian resekuensing untuk mengerti lebih baik terhadap susunan genom individu koleksi plasma nutfah. Sebanyak delapan spesies tanaman telah dikarakterisasi variasinya pada level genom menggunakan teknologi NGS. Informasi genotipe individu plasma nutfah hasil karakterisasi ini memfasilitasi percepatan penemuan berbagai variasi genetik untuk mendukung program pemuliaan tanaman yang terarah, lebih cepat dan lebih akurat menggunakan metode pemuliaan paradigma baru yaitu pemuliaan tanaman berbasis data genomika. Program pemuliaan tanaman ke depan akan lebih berfokus pada komparasi komposisi genom individu plasma nutfah yang membuka peluang penggunaan kombinasi strategi baru pemetaan genetik dan analisis evolusi untuk penemuan dan pemanfaatan variasi genetik SDG tanaman dengan lebih optimal dan lebih

komprehensif untuk menunjang program pemuliaan tanaman yang sinambung dan lestari.

## DAFTAR PUSTAKA

- Ansorge WJ. 2009. Next-generation DNA sequencing techniques. *Nature Biotechnol.* 25:195-203.
- Argout X, Salse J, Aury JM, Guiltinan MJ, *et al.* [61 authors]. 2011. The genome of *Theobroma cacao*. *Nature Genet.* 43:101-108. doi:10.1038/ng.736.
- Calzada JPV, de la Vega OM, Hernandez-Guzman G, Ibarra-Laclette E, Alvarez-Mejia C, Vega-Arreguin JC, Jimenez-Moraila B, Fernandez-Cortes A, Corona-Armenta G, Herrera-Estrella L, Herrera-Estrella A. 2009. The Palomero genome suggests metal effects on domestication. *Science.* 326:1078.
- Dhanapal AP. (2012) Genomics of crop plant genetic resources. *Adv Biosci Biotechnol.* 3:378-385.
- Franca LTC, Carrilho E, Kist TBL. 2002. A review of DNA sequencing techniques. *Quarterly Rev Biophys.* 35:169-200.
- Kling J. 2005. The search for sequencing thoroughbred. *Nature Biotechnol.* 23:1333-1335. doi:10.1038/nbt1105-1333.
- Metzker M. 2010. Sequencing technologies—The next generation. *Nat Rev Genet.* 11:31-46. doi:10.1038/nrg2626.
- Olsvik O, Whalberg J, Petterson B, Uhlen M, Popovic T, Wachmuth IK, Fields PL. 1993. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *J Clin Microbiol.* 31:22-25.

- Pérez-de-Castro AM, Vilanova S, Canizares J, Pascual L, Blanca JM, Diez MJ, Prohens J, Pico B. 2012. Application of genomic tools in plant breeding. *Curr Genom.* 13 (3):179-195.
- Rijzaani H, Lestari P, Priyatno TP, Tasma IM. 2016. Pusat genom komoditas pertanian Indonesia. *Warta Balitbangtan.* 38(4):2-6.
- Sanger F, Nicklen S, Coulson AR. 1997. DNA sequencing with chain-terminating inhibitors. *Proc Nat Acad Sci. U.S.A.* 74:5463-5467. doi:10.1073/pnas.74.12.5463.
- Satyawan D, Rijzaani H, Tasma IM. 2014. Characterization of genomic variation in Indonesian soybean (*Glycine max*) varieties using next-generation sequencing. *Plant Genet Res.* 12:S109-S113. DOI: <http://dx.doi.org/10.1017/S1479262114000380>.
- Schmutz J, Cannon SB, Schlueter J, Ma J, *et al.* [44 authors]. 2010. Genome sequence of the paleopolyploid soybean. *Nature.* 463:178-183.
- Schnable PS, Ware D, Fulton RS, Stein JC, *et al.* [152 authors]. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science.* 326:1112-1115.
- Shendure J, Ji H. 2008. Next-generation DNA sequencing. *Nat Biotech.* 26:1135-1145.
- Singh R, Ong-Abdullah METL, Low MAA, Manaf R, Rosli R, Nookiah LCL, Ooi SE, Ooi KL, Chan MA, *et al.* [87 authors]. 2013. Oil palm genome sequence reveals divergence of interfertile species in old and new worlds. *Nature.* 500:335-339. doi: 10.1038/-nature12309.
- Singh R, Noorhariza MZ, Ting NC, Rozana R, Tan SG, Low LET, Ithnin M, Cheah SC. 2008. Exploiting an oil palm EST the development of gene-derived and their exploitation for assessment of genetic diversity. *Biologia.* 63:1-9.

- Tasma IM, Satyawan D, Rijzaani H, Utami DW, Lestari P, Rosdianti I. 2012. Pembentukan empat peta genetik sawit, jarak pagar, padi, dan kedelai, serta identifikasi marka SNP kakao dan sapi. Laporan Akhir Penelitian APBN 2012. BB Biogen, Bogor. 67 hlm.
- Tasma IM. 2014. *Single nucleotide polymorphism* (SNP) sebagai marka DNA masa depan. *Warta Biogen*. 10(3):7-10.
- Tasma IM, Satyawan D, Rijzaani H. 2015a. Pembentukan pustaka genom, resequensing, dan penemuan SNP berdasarkan sekuen genom total varietas kedelai Indonesia. *J AgroBiogen*. 11:7-16.
- Tasma IM, Rijzaani H, Satyawan D, Lestari P, Utami DW, Rosdianti I, Purba AR, Mansyah E, Sutanto A, Kirana R, Kusmana, Anggraeni A, Pabendon M, Rubiyo. 2015b. Nextgen-based DNA marker development of several important crop and animal species. Manuscript Presented at the 13th SABRAO Congress and International Conference, 14–16 September 2015, Bogor, Indonesia. 8 pp.
- Tasma IM, Satyawan D, Rijzaani H, Rosdianti I, Lestari P, Rubiyo. 2016. Genomic variation of five Indonesian cacao (*Theobroma cacao* L.) varieties based on analysis using next generation sequencing. *Indonesian J Agric Sci*. 17:57-64.
- Tasma IM, Satyawan D, Rijzaani H. 2018. Characterization of 337 exon-based single nucleotide polymorphisms (SNPs) unique to the Indonesian varieties. Presented at the 1<sup>st</sup> International Conference on Genetic Resources and Biotechnology (Information system and exchanges of genetic resources for effective crop improvement), Bogor, Indonesia 20–21 August 2018.

- Thompson MK. 2014. High throughput SNP genotyping to accelerate plant breeding. *Plant Breed. Biotech.* 2:195-212.
- Van K, Rastogi K, Kim K-H, and Lee S-H. 2013. Next-generation sequencing technology for crop improvement. *SABRAO J Breed Genet.* 45:84-99.
- Varshney RK, Nayak SN, May GD, Jackson SA. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 9:522-530.
- Wetterstrand, K.A. 2011. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Available at: <http://www.genome.gov/sequencingcosts/>. [Internet] (Accessed March 1, 2011).
- Yu J, S Hu, J Wang, Wong GK, *et al.* [85 authors]. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science.* 296:79-92.