

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu tanaman hortikultura atau tanaman yang dibudidayakan di Indonesia adalah cabai. Secara umum di Indonesia cabai merupakan salah satu bumbu yang telah lama dan banyak digunakan. Beberapa daerah bahkan menggunakan cabai sebagai bahan baku utama pada setiap masakan (Tempo Store, 2014). Pemanfaatan cabai begitu luas karena keanekaragamannya jenisnya dan salah satu manfaatnya yaitu digunakan sebagai bahan baku industri makanan. Hal tersebut mengakibatkan tanaman cabai memiliki nilai ekonomi yang tinggi.

Tingginya permintaan pasar terhadap cabai terkadang tidak diimbangi dengan ketersediaan produknya. Salah satu penyebab ketersediaan cabai tidak memenuhi adalah akibat daya tahan tanaman yang rendah terhadap hama, penyakit maupun cuaca sehingga tidak sedikit tanaman yang mati. Seperti kasus serangan virus gemini di Kabupaten Malang yang membuat petani cabai menjadi resah. Virus gemini atau biasa disebut dengan virus kuning ini dapat mengakibatkan tanaman cabai tidak berbuah. Faktor iklim dan cuaca memudahkan virus yang dibawa oleh serangga putih (kutu kebul) tumbuh subur sehingga menyerang tanaman cabai yang baru saja memulai masa tanamnya. Lahan cabai mereka banyak diserang virus kuning hingga jumlah produksinya menyusut sampai 5 ton dalam satu hektarenya (Aminudin, 2018).

Oleh karena itu perlu dilakukan upaya untuk meningkatkan produksi cabai dengan memproduksi varietas tanaman cabai yang baik. Salah satu pendekatan yang digunakan untuk mendukung upaya peningkatan produksi cabai dengan mutu yang baik adalah dengan menerapkan kegiatan pemuliaan tanaman. Pemuliaan tanaman merupakan suatu kegiatan yang dilakukan untuk memperbaiki karakter tanaman yang diwariskan pada suatu populasi baru dengan sifat genetik baru. Produk pemuliaan tanaman adalah kultivar dengan ciri-ciri khusus sesuai dengan yang diinginkan, seperti produksi tinggi, toleran terhadap kondisi-kondisi

lingkungan yang mariginal, resisten terhadap hama serta penyakit dan lain-lain (Nuraida, 2012).

Salah satu metode untuk mempercepat proses pemuliaan tanaman adalah dengan melakukan pendekatan berbasis marka molekuler. Penggunaan marka molekuler sangat menjanjikan dalam program pemuliaan tanaman, terutama dalam membantu penandaan gen-gen yang mengontrol karakter yang diinginkan, seperti produktivitas, ketahanan terhadap hama dan penyakit tanaman serta kondisi lingkungan yang ekstrem (Reflinur & Puji, 2015). Marka molekuler dapat diklasifikasikan menjadi tiga kategori. Pertama adalah marka berdasarkan hibridisasi, yaitu *restriction fragment length polymorphism* (RFLP). Kedua adalah marka berdasarkan *polymerase chain reaction* (PCR) seperti *random amplification of polymorphic DNA* (RAPD), *amplified fragment length polymorphism* (AFLP), *inter simple sequence repeats* (ISSR) dan *microsatellite* atau *simple sequence repeats* (SSR). Berikutnya yang ketiga merupakan marka berdasarkan sekuens, yaitu *single nucleotide polymorphism* (SNPs) (Varshney, et al., dalam Izzah, 2017).

Diantara berbagai macam jenis marka, SNPs merupakan marka molekuler terbaru yang ditemukan. SNPs merupakan variasi urutan DNA yang terjadi ketika sebuah nukleotida tunggal A, T, C, dan G dalam suatu genome dari aksesori cabai tertentu terdeteksi memiliki perbedaan dengan aksesori cabai lainnya. Artinya, terdapat perbedaan antara sekuens DNA tanaman cabai dengan *genom referencenya* pada marker tertentu. Pada prosesnya untuk menemukan *SNPs*, dapat diterapkan metode *genotyping by sequencing* (GBS). GBS merupakan metode untuk menemukan SNPs dalam studi genotip (Illumina, 2015). GBS menggunakan enzim restriksi untuk mengurangi kompleksitas genom dan beberapa contoh DNA genotip.

Saat ini GBS dianggap semakin penting karena menjadi alat dengan penggunaan biaya yang efektif atau hemat dan unik dalam bidang pemuliaan tanaman (He, et al., 2014). Data olahan proses GBS hasilnya akan mengeluarkan sekuens atau urutan DNA yang disimpan dalam format *fasta/fastq*. Data sekuens tersebutlah yang kemudian dilakukan *alignment* atau disejajarkan dengan *genome*

reference agar ditemukan lokasi SNPs. Hasil dari GBS tersebut perlu direpresentasi dengan menerapkan *variant call format* (VCF). *Variant call* adalah kesimpulan bahwa ada perbedaan nukleotida dengan referensinya (pada posisi tertentu) dalam genom individual atau transkriptom (Lawrence, 2014).

Pada kasus ini akan digunakan *software* R dalam melakukan analisis. Dimulai dari membaca data dalam format *fastq* dan menampilkan kualitas masing-masing basa hingga menemukan lokasi SNPs. Dibutuhkan beberapa *package* dari *Bioconductor* maupun *package* yang ada di dalam R agar dapat menampilkan hasil analisisnya. *Package* yang digunakan tentunya yang memiliki fungsi untuk bekerja dengan *file fasta/fastq*. Selain itu digunakan pula *package* yang berfungsi untuk menampilkan grafik kualitas sekuens DNA dan memiliki fungsi *alignment* guna menampilkan lokasi SNPs.

Jika SNPs telah ditemukan, maka hasil letak SNPs dapat dibandingkan dengan *genom reference*-nya. Hasil perbandingan tersebut digunakan sebagai studi pendahuluan untuk mengetahui perbedaan SNPs antar aksesori tanaman cabai yang memiliki karakteristik yang berbeda (berbeda secara penampakan agronomis, ketahanan penyakit tertentu, dan lain-lain). Data SNPs tersebut kemudian akan divalidasi menggunakan populasi uji tertentu untuk melihat tingkat akurasi/kecocokan antar kandidat SNPs tertentu dengan penampakan aslinya (data fenotipik) di lapangan.

1.2 Rumusan Masalah

Berdasarkan dengan latar belakang yang telah dituliskan, maka rumusan masalah yang akan dikaji dalam penelitian ini adalah sebagai berikut :

1. Bagaimana membaca data sekuens DNA dengan format *fastq* menggunakan *package Bioconductor* di *software* R?
2. Bagaimana implementasi penggunaan *package Bioconductor* di *software* R guna menampilkan kualitas data sekuens?
3. Bagaimana langkah-langkah preparasi data untuk *SNP Calling* di *software* R?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah :

1. Data yang digunakan adalah data sekuens DNA tanaman cabai yang diperoleh dari PT. East West Seed Indonesia.
2. Bahasa pemrograman dan *software* yang digunakan dalam penelitian ini yaitu *software* R versi 3.5 dengan beberapa *package* dari *Bioconductor* yaitu *ShortRead*, *Rqc*, *Rbowtie*, dan *Rsamtools*.
3. Hasil akhir berupa *bam file* dimana didalamnya berisi sekuens DNA yang terdapat SNPs.

1.4 Tujuan Penelitian

Adapun beberapa hal yang ingin dituju pada penelitian ini adalah:

1. Untuk mengetahui cara membaca data sekuens DNA dengan format *fastq* menggunakan *package* dari *software* R.
2. Untuk mengetahui implementasi penggunaan *package* *Bioconductor* di *software* R guna menampilkan kualitas data sekuens.
3. Untuk mengetahui langkah-langkah preparasi data untuk *SNP Calling* di *software* R.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Manfaat bagi penulis ialah mendapatkan pengetahuan baru terkait ilmu bioinformatika serta dapat menerapkan ilmu statistika yang telah diperoleh sebelumnya.
2. Penelitian ini dapat digunakan untuk membantu proses pemuliaan tanaman dengan cara menyeleksi hasil olahan data GBS menjadi hasil akhir yang dapat berguna bagi kegiatan pemuliaan tanaman. Hasil kualitas basa maupun sekuens *reads* dapat digunakan oleh peneliti dibidang pertanian untuk menilai valid tidaknya sekuens yang dihasilkan oleh mesin sekuensing. Selanjutnya dari data SNPs yang diperoleh penulis dapat

digunakan sebagai studi pendahuluan untuk membedakan karakteristik tanaman cabai dari berbagai aksesori yang ada berdasarkan posisi SNPs markernya.