IMPLEMENTASI *PACKAGE BIOCONDUCTOR* PADA *SOFTWARE* R UNTUK OLAH DATA HASIL *GENOTYPING BY SEQUENCING* PADA TANAMAN CABAI (*Capsicum annuum*)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Jurusan Statistika

Hawila Sonya Savitri
14 611 176

JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2018

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Implementasi *Package Bioconductor* pada *Software* R untuk Olah Data Hasil *Genotyping by Sequencing* pada Tanaman Cabai (*Capsicum annuum*).
Nama Mahasiswa	:	Hawila Sonya Savitri
Nomor Mahasiswa	:	14 611 176

TUGAS AKHIR TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Yogyakarta, 1 Agustus 2018

Pembimbing I	**Pembimbing II**
Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.	**Muryanto, S.P., M.Si.**

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI *PACKAGE BIOCONDUCTOR* PADA *SOFTWARE* R UNTUK OLAH DATA HASIL *GENOTYPING BY SEQUENCING* PADA TANAMAN CABAI (*Capsicum annuum*)

Nama Mahasiswa : Hawila Sonya Savitri

Nomor Mahasiswa : 14 611 176

TUGAS AKHIR INI TELAH DIAJUKAN

PADA TANGGAL 1 AGUSTUS 2018

Nama Penguji	Tanda Tangan
1. Husna Nugrahapraja, Ph.D.	: .....
2. M. Hasan Sidiq Kurniawan, S.Si., M.Sc.	: .....
3. Muryanto, S.P., M.Si	: .....
4. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.	: .....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

UNIVERSITAS ISLAM INDONESIA YOGYAKARTA FAKULTAS MIPA

Prof. Riyanto, S.Pd., M.Si., Ph.D

KATA PENGANTAR

Assalaamu'alaikum warahmatullaahi wabarakaatuh

Alhamdulillahirobbil alamin, puji syukur senantiasa penulis panjatkan atas kehadirat Allah SWT yang telah memberikan rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "Implementasi *Package Bioconductor* pada *Software* R untuk Olah Data Hasil *Genotyping by Sequencing* pada Tanaman Cabai (*Capsicum annuum*)". Shalawat serta salam tak lupa penulis haturkan kepada junjungan Nabi Besar Muhammad SAW, karena berkat perjuangan beliau kita dapat merasakan nikmat Islam yang begitu luar biasa serta indahnya ilmu pengetahuan.

Penelitian ini dilakukan sebagai salah satu persyaratan yang harus dipenuhi untuk mencapai gelar sarjana statistika strata satu di Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Penulis menyadari bahwa dalam penyusunan tugas akhir ini tidak lepas dari bimbingan, dorongan, dan bantuan dari berbagai pihak, sehingga kegiatan yang telah direncanakan dapat terlaksana dengan baik dan terselesaikan pada waktu yang tepat. Oleh karena itu, perkenankan penulis menghaturkan terima kasih kepada :

1. Mama Nurul tercinta sebagai penyemangat terbesar bagi penulis yang luar biasa memberikan kasih sayang serta mendoakan yang terbaik untuk penulis dan bekerja keras tanpa kenal lelah demi kelancaran studi penulis. Bude Aniek, Bude Inien, Om Yoyok dan seluruh keluarga penulis yang telah memberikan semangat kepada penulis, memberikan dukungan baik moril maupun materiil.
2. Bapak Fathul Wahid, S.T., M.Sc., Phd, selaku Rektor Universitas Islam Indonesia.
3. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia, Yogyakarta.

4. Dr. RB. Fajriya Hakim, S.Si., M.Si., selaku Ketua Jurusan Statisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia beserta jajarannya, terima kasih atas kerja keras bapak ibu dalam memberikan ilmu yang manfaat.
5. Ibu Dr. techn. Rohmatul Fajriyah, S.Si., M.Si., selaku dosen pembimbing pertama yang telah membimbing penulis, meluangkan waktu dan tenaga, memberikan keyakinan kepada penulis, memberikan semangat, arahan, saran serta ilmu yang bermanfaat sehingga tugas akhir ini dapat terselesaikan dengan baik.
6. Bapak Muryanto, S.P., M.Si. selaku pembimbing kedua, Bapak Boy Valenza Damiri, Mas Sandi Arya Rumintang, Ibu Anisa beserta jajaran staf divisi Bioteknologi PT East West Seed Indonesia yang telah mengenalkan ilmu pengetahuan baru dibidang bioinformatika, memberikan saran maupun nasihat, memberikan semangat, serta membimbing penulis dalam memahami setiap detail ilmu bioinformatika.
7. Kawan-kawan seperjuangan *Bioinformatics* Ika, Nurlina, Ummi, Nanda, Husain, Ari, Leni, Himelda, Maidah, Shodiq, Gilang dan Aat yang sudah banyak memberikan kenangan dan tak lupa saling mendukung serta menguatkan satu sama lain karena kita luar biasa.
8. Teman-teman, kakak-kakak dan adik-adik Statistika UII yang sudah memberikan dukungan, bantuan, serta berbagi ilmu dan pengalaman yang berharga kepada penulis selama menjalani kuliah di Universitas Islam Indonesia.
9. Semua pihak yang telah memberikan semangat, dukungan, doa, dan bantuan kepada penulis baik secara langsung maupun tidak langsung yang tidak dapat penulis sebutkan satu-persatu.

Semoga bantuan dan dukungan yang telah diberikan mendapat balasan dari Allah SWT. Penulis menyadari sepenuhnya bahwa dalam penyusunan laporan penelitian ini masih banyak kekurangan, oleh sebab itu penulis mengharap kritik dan saran yang bersifat membangun dalam pengembangan di masa mendatang

dan bermanfaat bagi yang membaca serta penulis khususnya. Semoga Allah SWT selalu melimpahkan rahmat dan hidayah-Nya kepada kita semua, Aamiin.

Wassalamualaikum warahmatullaahi wabarakatuh

Yogyakarta, Juli 2018

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 31 Juli 2018

Penulis

IMPLEMENTASI *PACKAGE BIOCONDUCTOR* PADA *SOFTWARE* R UNTUK OLAH DATA HASIL *GENOTYPING BY SEQUENCING* PADA TANAMAN CABAI (*Capsicum annuum*)

Oleh: Hawila Sonya Savitri

Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Islam Indonesia

INTISARI

Penelitian ini menerapkan *package* dari *Bioconductor* pada *software* R dalam menangani data sekuens DNA cabai yang disimpan dengan format *fastq*. Data tersebut akan dicari lokasi SNPs dengan cara *alignment* data *fastq* dengan referensi genom cabai Zunla yang disimpan dalam format *fasta*. Sebelumnya terlebih dahulu digunakan *package ShortRead* dan *Rqc* untuk melihat kualitas data sekuens DNA yang dihasilkan oleh mesin sekuensing. Hasil sekuens dapat dikatakan sempurna karena nilainya mendekati 100%. Selanjutnya dilakukan *alignment* data *fastq* dan *fasta* yang hasilnya disimpan dalam *sam file* dengan bantuan *package Rbowtie*. Kemudian data dalam format *sam* dikonversi menjadi *bam* dengan menggunakan *package Rsamtools*. Hasilnya yaitu dari dua belas data yang dimiliki, data *bam* A2_142804-1_1 memiliki jumlah lokasi sekuens DNA terbanyak yang didalamnya mengandung SNPs, yaitu sebanyak 95.349 *ranges*.

Kata Kunci: *Alignment*, Cabai, *Fastq*, Referensi Genom, Sekuens DNA

IMPLEMENTATION OF BIOCONDUCTOR'S PACKAGE IN R SOFTWARE FOR DATA PROCESSING OF GENOTYPING BY SEQUENCING ON HOT PEPPER (Capsicum annuum)

Hawila Sonya Savitri

Department of Statistics, Faculty of Matematics and Natural Sciences

Islamic University of Indonesia

ABSTRACT

This study applied package from Bioconductor in software R to handling data of hot pepper (Capsicum annuum) DNA sequence which stored with fastq format. These data will be searched the location of SNPs by alignment the fastq file with genom reference of pepper Zunla (stored in fasta format). Previously we used ShortRead and Rqc packages to quality assessment of the DNA sequence data generated by the sequencing machine. The DNA sequence can be considered perfect because the value is close to 100%. The next step is alignment of fastq and fasta file and the result is stored in sam file by using Rbowtie package. Then, sam data is converted into bam using Rsamtools package. From twelve fastq datas, A2_142804-1_1 has the most number location of DNA sequence which contain SNPs. There are 95,349 ranges.

Keyword: *Alignment, Hot pepper, Fastq, Genome Reference, DNA Sequence*