

**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*