

# INTEGRASI PIPELINE DAN BASIS DATA SINGLE NUCLEOTIDE POLYMORPHISM UNTUK PEMULIAAN TANAMAN KEDELAI

## INTEGRATION OF PIPELINE AND SINGLE NUCLEOTIDE POLYMORPHISM DATABASE FOR SOYBEAN BREEDING

Miftakhul Huda<sup>1</sup>, M Abrar Istiadi<sup>1</sup>, Habib Rijzaani<sup>2</sup>, Wisnu Ananta Kusuma<sup>3\*</sup>

Departemen Ilmu Komputer, FMIPA Institut Pertanian Bogor, Bogor<sup>1,3</sup>  
ananta@ipb.ac.id<sup>3\*</sup>

Balai Besar Litbang Bioteknologi & Sumber Daya Genetik Pertanian<sup>2</sup>

### ABSTRACT

Plant breeding using conventional approaches is high cost and time consuming. Researchers start elaborating the opportunity of plant breeding based on genome analysis in order to obtain an efficient process. This approach uses *Single Nucleotide Polymorphism* (SNP) marker and their association with phenotypes. The SNP identification process actually can be conducted using application software such as SAMtools and VCFutils. However, the output of the software, such a VCF file, is not easy to be understood and managed by breeding researchers. This research aims to integrate software pipeline yield by previous research and database for storing SNP candidates and their supporting information such as flanking area-obtained from DNA sequence alignment and SNP's ID-which is automatically integrated to Soybean SNP reference from NCBI. Moreover, the system also provides a web based searching feature for helping researchers retrieving relevant SNP candidate information.

*Keywords: database, pipeline, SNP, plant breeding, soybean*

### ABSTRAK

Upaya penemuan varietas unggul dengan cara pemuliaan tanaman secara konvensional membutuhkan waktu yang lama dan biaya yang besar. Kendala tersebut mulai dapat diatasi dengan memanfaatkan informasi yang diperoleh dari analisis genom, yaitu berupa markah *Single Nucleotide Polymorphism* (SNP) beserta asosiasi dengan fenotipnya. Selama ini, identifikasi SNP dapat dilakukan menggunakan aplikasi SAMtools dan VCFutils. Keluaran dari aplikasi tersebut adalah informasi SNP yang disimpan dalam fail berformat *Variant Call Format* (VCF). Fail dengan format tersebut sulit dimengerti dan diolah oleh peneliti pemuliaan. Penelitian ini bertujuan mengintegrasikan perangkat lunak *pipeline* yang dihasilkan oleh penelitian sebelumnya dan basis data yang menyimpan informasi berupa SNP dan informasi pendukung lainnya yang terkait dengan identifikasi SNP, antara lain *flanking area* yang diperoleh dari *DNA sequence alignment* dan ID kandidat SNP yang terintegrasi dengan data referensi SNP kedelai dari NCBI. Selain itu, sistem ini juga dilengkapi fitur pencarian yang berbasis web untuk memudahkan peneliti pemuliaan menemukembalikan informasi kandidat SNP yang relevan.

*Katakunci: basis data, pipeline, SNP, pemuliaan tanaman, kedelai*

### PENDAHULUAN

Kedelai merupakan salah satu sumber protein nabati yang secara langsung dapat digunakan untuk pemenuhan kebutuhan gizi manusia maupun sebagai bahan penghasil produk berkualitas tinggi [1]. Di Indonesia kebutuhan tanaman kedelai mencapai 2,35 juta ton per tahun dan cenderung mengalami peningkatan setiap

---

tahunnya (Atman 2009; Purna et al. 2009). Peningkatan jumlah tersebut harus diikuti dengan penemuan varietas unggul agar kualitas kedelai bisa meningkat. Usaha pemuliaan tanaman kedelai merupakan salah satu cara untuk membantu menghasilkan varietas baru kedelai yang unggul.

Upaya penemuan varietas unggul dengan cara pemuliaan tanaman secara konvensional membutuhkan waktu yang lama dan biaya yang besar. Kendala tersebut mulai dapat diatasi dengan memanfaatkan informasi yang diperoleh dan analisis genom pada tanaman kedelai tersebut, yaitu berupa markah *Single Nucleotide Polymorphism* beserta asosiasi dengan fenolipenya. *Single Nucleotide Polymorphism* atau SNP (diucapkan "snips") didefinisikan sebagai variasi dalam urutan DNA yang terjadi ketika sebuah nukleotida tunggal A, T, C, G dalam genom berbeda antar anggota spesies (antar kromosom berpasangan dalam individu). Ketika SNP terjadi di dalam gen, mereka menciptakan varian yang berbeda (*alel*) pada gen tersebut. SNP adalah salah satu penanda molekuler yang dapat digunakan untuk mempercepat usaha pemuliaan tanaman. Selama ini, identifikasi SNP dapat dilakukan menggunakan aplikasi SAMtools dan VCFutils [2]. Proses ini terbagi atas beberapa tahapan, antar lain *Sequence alignment*, *Alignment post-process*, *Variant calling*, *Filtering SNP Candidates*, dan *Making Sense of SNP Data* [3]. Keseluruhan tahap identifikasi SNP sangat penting untuk membantu proses penyilangan menjadi lebih efektif dan efisien.

Pada penelitian sebelumnya telah dihasilkan sebuah sistem identifikasi SNP kedelai berbasis web yang diberi nama ISNIP [4]. Keluaran dari sistem tersebut adalah informasi SNP yang disimpan dalam fail berformat *Variant Call Format* (VCF). Fail dengan format tersebut sulit dimengerti dan diolah oleh peneliti pemuliaan. Oleh karena itu, diperlukan sebuah sistem yang mampu menyimpan informasi SNP kedelai dari fail VCF tersebut. Hal ini akan memberikan kemudahan peneliti pemuliaan untuk mengolah informasi SNP kedelai tersebut. Pada penelitian ini akan diintegrasikan perangkat lunak *pipeline* yang dihasilkan oleh penelitian sebelumnya dan basis data yang menyimpan informasi berupa SNP kedelai dan informasi pendukung lainnya yang terkait dengan identifikasi SNP kedelai, antara lain *flanking area* yang diperoleh dari *DNA sequence alignment* dan ID kandidat SNP yang terintegrasi dengan data referensi SNP kedelai dari NCBI. Selain itu, sistem ini juga dilengkapi fitur pencarian yang berbasis web untuk memudahkan peneliti pemuliaan menemukembalikan informasi kandidat SNP yang relevan.

## METODE PENELITIAN

### Data

Data yang digunakan pada penelitian ini terdiri dari tiga jenis, di antaranya :

1. Data dengan fail berformat VCF yang didapatkan dengan cara menjalankan semua proses *pipeline* pada sistem penelitian sebelumnya. Untuk percobaan, pada penelitian ini digunakan 4 varietas tanaman kedelai, di antaranya B3293 5M, Daforst, Grobogan, dan Malabar. Data ini merupakan informasi SNP yang berfungsi sebagai kunci yang diperlukan untuk proses penyimpanan ke dalam basis data.
2. Data dengan fail berformat FASTA yang didapatkan dari proses *sequence alignment* pada sistem penelitian sebelumnya. Data ini sebagai informasi pendukung, yaitu berupa *flanking area* untuk setiap SNP kedelai yang disimpan ke dalam basis data.
3. Data SNP kedelai yang diambil dari basis data situs resmi NCBI (*National Center for Biotechnology Information*), yang terdiri atas 20 kromosom dengan kurang lebih

12.000.000 kemunculan SNP. Data ini sebagai integrasi untuk kandidat SNP yang akan disimpan ke dalam basis data.

#### Metode

Penelitian ini dimulai dengan melakukan analisis untuk menyelesaikan permasalahan yang terjadi. Selain itu juga dilakukan analisis data dari fail berformat VCF yang digunakan dengan berkonsultasi kepada pakar Biologi Molekuler mengenai bagian-bagian penting dalam SNP kedelai, termasuk informasi pendukung lainnya. Setelah dilakukan analisis, maka didapatkan informasi untuk membangun basis data, termasuk atribut apa saja yang digunakan.

Basis data yang dibangun ini berfungsi untuk menyimpan informasi SNP kedelai yang sudah diproses dari *pipeline* pada penelitian sebelumnya. Untuk memproses data informasi SNP kedelai agar tersimpan di basis data, maka dilakukan integrasi *pipeline* data SNP kedelai dan basis data, termasuk juga harus integrasi data SNP kedelai dengan data SNP kedelai dari NCBI. Adapun penambahan *flanking area* diperoleh dengan mengolah data hasil Multiple Sequence Alignment.

Selain itu, untuk membantu peneliti pemuliaan atau pengguna mengolah dan menemukembalikan data yang sudah tersimpan di basis data, maka dibangun fitur pencarian berbasis web. Setelah sistem diimplementasikan, dilakukan pengujian dengan metode Black Box. Tujuan pengujian ini adalah untuk mengidentifikasi adanya kesalahan pada implementasi integrasi *pipeline*, yaitu dengan melakukan semua kemungkinan percobaan beberapa data yang berbeda.

## HASIL DAN PEMBAHASAN

Permasalahan pada penelitian ini adalah peneliti pemuliaan sulit untuk memahami dan mengolah fail berformat VCF yang diperoleh dari proses *pipeline* pada penelitian sebelumnya yang diberi nama "ISNIP" serta kurangnya informasi pendukung untuk membantu identifikasi SNP tanaman kedelai. Untuk itu, perlu adanya solusi untuk mengatasi masalah tersebut yaitu dengan mengintegrasikan *pipeline* dan basis data yang menyimpan informasi berupa SNP kedelai dan informasi pendukung lainnya. Proses tersebut akan diimplementasikan dalam antarmuka berbasis web untuk memudahkan peneliti pemuliaan melakukan proses tersebut.

#### Analisis Data

Langkah awal pada penelitian ini adalah menganalisis data yang digunakan. Data yang digunakan terdiri dari tiga jenis fail berformat VCF yang dihasilkan pada proses *pipeline* penelitian sebelumnya. Fail tersebut merupakan informasi utama mengenai SNP. Format data bisa dilihat pada gambar 1.

```
##format=VCFv4.1
##contig=chr1
##fileformat=VCFv4.1
##INFO=ID=DP,Number=1,Type=Integer,Description="Raw read depth"
##INFO=ID=MQ,Number=1,Type=Integer,Description="Mean square mapping quality of all reads mapped"
##INFO=ID=MQ0,Number=1,Type=Integer,Description="Number of reads with zero mapping quality"
##INFO=ID=POS,Number=1,Type=String,Description="Position of the variant site"
```

Gambar 1 Format Data SNP Fail berformat VCF

Dari Gambar 1 dapat dilihat bahwa fail tersebut bisa dikelompokkan menjadi dua bagian, yaitu:

1. *Header* yang berisikan info dan format penjelasan mengenai isi data SNP. Pada bagian tersebut diawali dengan tanda '#' untuk setiap barisnya. Pada penelitian ini, sebagian besar bagian tersebut tidak diproses atau digunakan untuk proses integrasi *pipeline* dan basis data. Hanya pada baris terakhir bagian tersebut memberikan informasi *field-field* yang akan digunakan, yang berisikan antara lain : *chrom* (kromosom), *pos* (posisi), *id*, *ref* (*reference*), *alt* (alternatif), *qual* (kualitas), *filter*, info dan format.
2. Informasi SNP yang dihasilkan dari proses *pipeline*. Bagian ini untuk setiap barisnya memiliki pemisah yang dipisahkan oleh 'tab', artinya untuk setiap bagian yang dipisahkan oleh 'tab' mengandung informasi di setiap field-nya. Sebagai contoh:

```
10 1417490 . C T 9.31
```

- 10: mengandung informasi *field chrom*.
- 1417490: mengandung informasi *field pos*.
- .: mengandung informasi *field id*, tanda "." artinya informasinya kosong.
- C: mengandung informasi *field ref*.
- T: mengandung informasi *field alt*.
- 9.31: mengandung informasi *field qual*.

Pada penelitian ini, bagian ini yang nantinya akan diseleksi untuk proses disimpannya ke dalam basis data.

Data kedua yang digunakan adalah fail berformat FASTA (.fa atau .fq). Data tersebut diperoleh dari proses *sequence alignment* pada penelitian sebelumnya. Format data bisa dilihat pada Gambar 2.

```
>gi|9626243|ref|NC_001416.1| Enterobacteria phage lambda,
complete genome

GGGCGGCGACCTCGCGGGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCT
TCTTCG

TCATAACTTAATGTTTTTATTTAAAATACCCTCTGAAAAGAAAGGAAACGACAGGTGCTGAAAG
CGAGGC

TTTTTGGCCTCTGTCGTTTCCTTTCTCTGTTTTTGTCCGTGGAATGAACAATGGAAGTCAACAA
AAAGCA

GCTGGCTGACATTTTCGGTGCGAGTATCCGTACCAATCAGAAGTGGCAGGAACAGGGAATGCC
GTTCTG

CGAGGCGGTGGCAAGGGTAATGAGGTGCTTTATGACTCTGCCGCCGTCATAAAATGGTATGCCG
AAAGGG

ATGCTGAAATGAGAACGAAAAGCTGCCCGGGAGGTTGAAGAAGTGGCCAGGCCAGCGAGGC
AGATCT
```

Gambar 2 Format Data Fail berformat FASTA

Data tersebut pada intinya mengandung informasi DNA tanaman kedelai tersebut. Data ini digunakan untuk mendapatkan *flanking area*. *Flanking area* adalah informasi

DNA yang berada di sekitar SNP. Manfaatnya adalah untuk memudahkan peneliti pemuliaan menganalisa SNP lebih lanjut. Data ini nantinya akan terintegrasi pada posisi SNP yang diperoleh pada data sebelumnya dan disimpan ke dalam basis data sesuai posisi. Menurut pakar, idealnya untuk memudahkan peneliti menganalisa adalah mengambil 200-300 pasang basa.

Data ketiga yang digunakan merupakan data SNP kedelai dengan fail yang diperoleh dari basis data NCBI. Fail ini berformat VCF. Format data bisa dilihat pada Gambar 3.

```
##fileformat=VCFv4.0
##source=dbSNP
##reference=GCF_000004515.1
##INFO=<ID=RV,Number=0,Type=Flag,Description="RS orientation is reversed">
```

Gambar 3 Format Data SNP Kedelai dari NCBI

Data ini bentuk maupun formatnya memiliki kesamaan dengan data pertama informasi SNP yang diperoleh pada penelitian sebelumnya. Hanya *field* chrom, pos dan ID yang digunakan untuk proses integrasi dengan data dari basis data SNP kedelai dari NCBI ini. Apabila data hasil percobaan proses *pipeline* pada penelitian sebelumnya memiliki chrom dan pos yang sama dengan chrom dan pos pada data ini, maka ID pada data hasil percobaan akan diisi oleh ID data dari basis data SNP NCBI. Tujuannya untuk bisa mengintegrasikan dan membandingkan antara data SNP dari percobaan *pipeline* penelitian sebelumnya dan data pada basis data SNP kedelai dari NCBI.

**Pembangunan Basis Data SNP**

Setelah dilakukannya analisis terhadap data yang digunakan. Selanjutnya adalah membangun basis data terkait data tersebut. Perlu diperhatikan bahwa yang akan disimpan dalam basis data adalah informasi SNP dan informasi pendukung lainnya. Untuk keperluan penyimpanan informasi SNP ini dibutuhkan sebuah, yaitu tabel SNP. Tabel-tabel lain bersifat sebagai pendukung, antara lain tabel Peneliti, Institusi, dan Komoditi. Pada tulisan ini hanya difokuskan pada tabel SNP. Atribut dan tipe data dalam tabel SNP dapat dilihat pada tabel 1. Tabel SNP ini sudah termasuk informasi pendukung lainnya, yaitu ID, flank\_left dan flank\_right. Tabel ini yang menyimpan informasi SNP yang diperoleh melalui proses *pipeline* pada penelitian sebelumnya

Tabel 1 Rincian tabel SNP

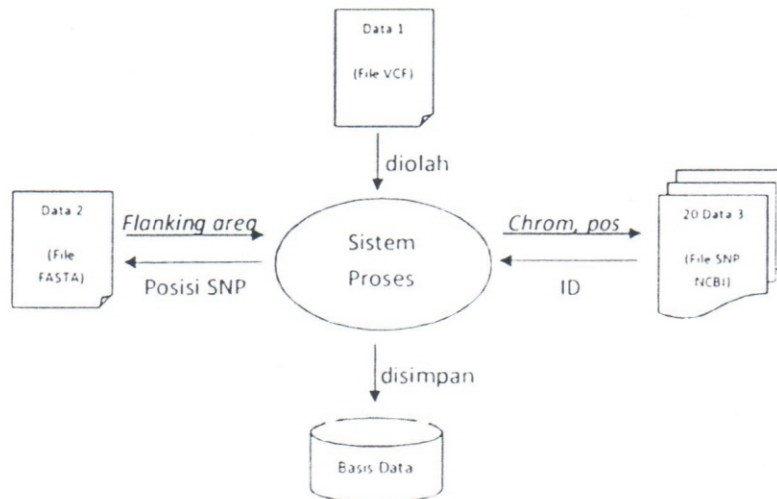
| ID_field | Deskripsi                     |
|----------|-------------------------------|
| chrom    | field yang berisi kromosom    |
| pos      | field yang berisi posisi SNP  |
| Id       | field yang berisi ID SNP      |
| ref      | field yang berisi reference   |
| alt      | field yang berisi alternative |
| qual     | field yang berisi kualitas    |
| filter   | field yang berisi filter SNP  |
| info     | field yang berisi informasi   |

|             |   |
|-------------|---|
| GT          | <i>field</i> yang berisi genotipe SNP                                     |
| PL          | <i>field</i> yang berisi <i>List of Phred-scaled genotype likelihoods</i> |
| GQ          | <i>field</i> yang berisi kualitas genotipe                                |
| flank_left  | <i>field</i> yang berisi <i>flanking area</i> sebelumnya posisi SNP       |
| flank_right | <i>field</i> yang berisi <i>flanking area</i> setelah posisi SNP          |

### Integrasi Pipeline dan Basis Data

Untuk mengintegrasikan *pipeline* dan basis data, maka diperlukan alur kerja proses tersebut. Alur kerja proses integrasi antar data bisa dilihat pada gambar 4.

Data dari fail VCF diolah untuk setiap barisnya dengan mengabaikan baris yang diawali simbol '#'. Dalam pengolahan data tersebut, yang diperhatikan antara lain: dipisahkan oleh "tab" untuk setiap *field*-nya, dan memperhatikan kualitas serta nukleotida tunggal. Kemudian sistem mengambil posisi untuk setiap *field*-nya dan digunakan untuk mendapatkan *flanking area* yang diperoleh dari data pada fail FASTA. Selain itu juga mengambil chrom dan pos untuk setiap *field*-nya dan digunakan untuk mendapatkan ID yang sama dengan data yang berupa fail VCF dari NCBI, yang berupa data dari basis data SNP kedelai NCBI. Hingga akhirnya data yang juga saling terintegrasikan akan disimpan ke dalam basis data.



Gambar 4 Alur Kerja Sistem

Alur tersebut diimplementasikan dalam sebuah sistem berbasis web dengan bahasa pemrograman PHP. Sistem ini memiliki dua fitur, yaitu pemrosesan data dan pencarian. Untuk melakukan pemrosesan data yang harus dilakukan peneliti pemuliaan adalah memilih fail VCF yang akan diolah dan memasukkan batas minimal kualitas yang akan disimpan dalam basis data SNP kedelai. Setelah dilakukan proses tersebut, maka data yang sesuai kriteria tersebut akan tersimpan ke dalam basis data dan sebagai mengalami perubahan ID karena telah terintegrasikan dengan data basis data SNP kedelai dari NCBI. Salah satu contoh data bagian yang berubah sebelum dan sesudah diproses dapat dilihat pada gambar 5.

| <u>Sebelum</u>  |
|---|
| <pre>id = flank_area =</pre>  |
| <u>Sesudah</u>  |
| <pre>id = rs124907313 flank_area = AAACCCCGGGGGTTTTACGTACGTAAAA AAAAACCCCGGGGGTTTTAAAAACCCCC GGGGTTTTTAAAAACCCCGGGGGTTTT AAAAACCCCGGGGGTTTTAAAAACCCCC GGGGTTTTTACGTACGTAAAAAAAACCC CCGGGGTTTTTAAAAACCCCGGGGGTTT TTAAAAACCCCGGGGGTTTTAAAAACCC CCGGGGTTTTTAAAAACCCCGGGGGTTT TTACGTACGTAAAAAAAACCCCGGGGGT TTTTAAAAACCCCGGGGGTTTTAAAAAC</pre> |

Gambar 5 Contoh Hasil Pemrosesan Data

Pada gambar 5 terlihat ada perubahan pada *field* ID dan penambahan flank\_area. Hal tersebut terbukti telah terjadi integrasi antar data pada sistem yang telah dibangun.

#### Fasilitas Pencarian

Selain fitur pemrosesan data juga ada fitur yang membantu peneliti pemuliaan mengolah data SNP kedelai tersebut, yaitu fitur pencarian. Fitur pencarian ini menggunakan teknik kueri tradisional yang sudah tersedia dalam *mysql*. Data yang akan dicari memiliki field yang banyak, sehingga fitur pencarian ini dibuat selain kueri tunggal juga bisa *advanced search* dengan multi-kueri, dalam arti pengguna bisa memasukkan kueri pencarian lebih dari satu sesuai *field* yang diinginkan. Antarmuka fitur pencarian SNP bisa dilihat pada gambar 6.

The image shows a search interface with three input fields, each containing the text 'All Field'. Below the third field is a 'Search' button. The interface is simple and functional, designed for entering search criteria.

Gambar 6 Antarmuka Advanced Search

Hasil pencarian yang diperoleh adalah data SNP yang relevan dengan kueri yang dimasukkan. Data yang ditampilkan dalam hasil pencarian berupa data singkat SNP, bisa dilihat secara keseluruhan maupun dilakukan penghapusan data.

#### Pengujian Sistem

Untuk mengetahui performa sistem, maka dilakukan pengujian pada sistem tersebut. Hal ini dilakukan dengan menguji beberapa data 1 (varietas kedelai) yang berbeda, untuk mengetahui adanya kesalahan dan waktu eksekusi sampai proses tersimpannya ke dalam basis data. Setelah dilakukan pengujian, tidak ada terjadi kesalahan dan diperoleh hasil waktu eksekusi yang berbeda-beda. Hal tersebut bisa dilihat pada tabel 2.

Tabel 2 Pengujian Data Varietas Kedelai

| Nama Varietas Kedelai | Banyak Baris | Waktu Eksekusi |
|-----------------------|--------------|----------------|
| B3293 5M              | 1875         | 2452.51 second |
| Daforst               | 8027         | Over           |
| Grobogan              | 3031         | 3369.82 second |
| Malabar               | 10391        | Over           |

Pada table 2 terlihat waktu eksekusi yang lama dan ada yang mengalami *over*. Hal tersebut terjadi karena data yang diolah sangat banyak termasuk data dari NCBI yang digunakan untuk integrasi.

Selain itu, juga dilakukan pengujian kueri untuk pencarian data informasi SNP kedelai (Gambar 6).

**Search Fitur**

| No | Chrom | Pos     | ID       | Ref | Alt | Qual | Filter | GT | CT | Detail |
|----|-------|---------|----------|-----|-----|------|--------|----|----|--------|
| 1  | 1     | 1000000 | rs123456 | A   | G   | 100  |        | AA | 1  | View   |
| 2  | 2     | 2000000 | rs234567 | C   | T   | 100  |        | CC | 1  | View   |

Gambar 6 Hasil Pencarian Informasi SNP Kedelai

Pada gambar 6 ditunjukkan hasil pencarian dengan multi-kueri dan terbukti data hasil pencarian relevan dengan yang dicari dan peneliti pemuliaan bisa melihat *detail* data maupun membandingkan dengan data NCBI.

#### Kontribusi Sistem yang Dibangun

Keuntungan dengan adanya sistem ini, antara lain :

1. Peneliti pemuliaan dapat mengolah data SNP kedelai untuk pemuliaan kedelai.
2. Dengan adanya fitur pencarian, maka peneliti pemuliaan akan lebih mudah untuk menemukembalikan informasi SNP kedelai yang relevan.
3. Dengan adanya informasi pendukung, maka informasi peneliti pemuliaan tidak terbatas hanya mengidentifikasi SNP dari percobaan yang dilakukan, melainkan mampu membandingkan dengan data SNP kedelai dari NCBI yang terbukti baik kualiitas SNP-nya dan mampu melihat nukleotida di sekitar SNP (*flanking area*).

#### KESIMPULAN

Penelitian ini menghasilkan sebuah sistem berbasis web untuk melakukan proses integrasi *pipeline* dan basis data SNP. Sistem yang dihasilkan mampu mengekstrak informasi SNP kedelai dari fail VCF dan mengintegrasikan dengan informasi data pendukung lainnya baik yang berasal dari fail FASTA maupun fail VCF yang berasal dari NCBI. Sistem ini juga mampu membantu peneliti pemuliaan untuk mengolah dan menemukembalikan data informasi SNP yang relevan.

#### PUSTAKA



- [1] Tiwari S, Shanker P, Tripathi M. 2004. *Effects of genotype and culture medium on in vitro androgenesis in soybean*. Indian Journal of Biotechnology. 3: 441-444.
  - [2] H. Li, B. Handsaker, A. Wysoker, T. Fenell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. *The sequence alignment/map format and SAMtools*. Bioinformatics. 25:2078–2079
  - [3] Altmann, A. et al., 2012. *A beginners guide to SNP calling from high-throughput DNA-sequencing data*. Human Genetics, 131(10):1541–54.
  - [4] Wisnu Ananta Kusuma, I Made Tasma, Agus Buono, Mukhlis Hidayat, Habib Rijzani, Toto Haryanto, M. Abrar Istiadi. *Pengembangan Sistem Identifikasi dan Analisis Single Nucleotide Polymorphism untuk Pemuliaan Tanaman Kedelai*. 2013. Laporan Hasil Penelitian KKP3N. Institut Pertanian Bogor.
-