

Implementasi Algoritma Needleman-Wunsch dalam Pengujian Tingkat Kemiripan DNA Babi Sulawesi Utara

Rivchi H. Ngovangari¹, Gladly C. Rorimpandey², Olivia Kembuan³, Hendro M. Sumual⁴

^{1,2,3,4}Program Studi Teknik Informatika UNIMA

Email: rivchingovanqari@gmail.com¹

Abstract— Data is a collection generated from various media used by humans. DNA data is a form of data, which is data in the form of genetic code text to store biological identities and characteristics of living things. This study used DNA to test the similarity level of North Sulawesi pigs. The method used is the Pairwise Sequence Alignment method. With sequences of different DNA lengths, the researcher implemented the Needleman-Wunsch algorithm for alignment. Matching is used using the BioPython module, which is in pairwise2. And this study produces a model and proportion of the similarity level of North Sulawesi Pigs that can be used as future references for biological research on molecular biology or for the field of informatics in implementing Needleman-Wunsch algorithm in Python, or references and additional knowledge for research in Bioinformatics.

Keywords— DNA, Needleman-Wunsch Algorithm, Bioinformatics, Pairwise Sequence Alignment, Similarity Level, Sequence Alignment, Python

Abstrak— Data merupakan suatu kumpulan yang dihasilkan dari berbagai media yang digunakan oleh manusia. Data DNA merupakan salah satu bentuk dari data, yang merupakan data dalam bentuk teks kode genetik untuk menyimpan identitas biologis dan ciri – ciri dari makhluk hidup. Penelitian ini menggunakan DNA untuk diuji tingkat kemiripan dari Babi Sulawesi Utara.

Metode yang digunakan adalah metode *Pairwise Sequence Alignment*. Dengan panjang urutan DNA yang berbeda, peneliti mengimplementasikan Algoritma Needleman-Wunsch untuk penyejajaran. Pencocokan digunakan dengan menggunakan modul *BioPython* yaitu *pairwise2*. Dan penelitian ini menghasilkan model dan persentase dari tingkat kemiripan Babi Sulawesi Utara yang dapat digunakan sebagai referensi kedepan untuk penelitian – penelitian biologis mengenai biologi molekuler atau bagi bidang informatika dalam penerapan algoritma Needleman-Wunsch dengan Bahasa Python, ataupun sebagai referensi dan pengetahuan tambahan untuk penelitian dalam bidang Bioinformatika.

Kata Kunci— DNA, Algoritma Needleman-Wunsch, Bioinformatika, *Pairwise Sequence Alignment*, tingkat kemiripan, *Sequence Alignment*, Python

I. PENDAHULUAN

Dalam kehidupan sehari – hari, kita pasti sering berhubungan atau menghasilkan data melalui berbagai macam aktivitas, apalagi dalam masa sekarang ini yang sudah banyak pengguna media teknologi dan media sosial yang tentunya setiap hari kita menggunakan teknologi dan melakukan *browsing* di *internet*, saling mengirim pesan, menggunakan *smartphone*, *notebook* atau laptop dan perangkat elektronik lainnya yang berarti setiap hari juga kita menghasilkan data. Data yang sebelumnya berupa teks, karakter, simbol, angka, gambar, suara dan lain sebagainya, diolah dengan berbagai cara agar lebih mudah untuk diterjemahkan ke dalam bentuk kesimpulan yang merupakan informasi atau pengetahuan yang baru yang menjadi solusi suatu permasalahan atau tambahan ilmu yang sebelumnya belum ditemukan. Dari berbagai macam bentuk data, salah satu data yang berbentuk teks yang menarik untuk diteliti adalah data DNA. DNA adalah kepanjangan dari *Deoxyribonucleic Acid*. yang diketahui dalam bidang biologi molekuler sebagai data

yang menyimpan identitas biologis dari makhluk yang telah diolah terlebih dahulu sehingga disimpan sebagai kode genetik dalam bentuk karakter – karakter atau huruf yang mengandung ciri – ciri dari makhluk hidup yang sulit dipahami jika hanya data mentah DNA-nya saja [1]. Karena kesulitan dalam memahami inilah, sehingga pemanfaatan teknologi informasi dalam mempelajari biologi molekuler sangat dibutuhkan dalam mempermudah komputasi serta menerjemahkan konsep abstrak yang ada pada biologi molekuler menjadi sebuah hasil yang lebih mudah untuk dicermati untuk tujuan penelitian. Dalam penelitian ini, peneliti akan menerapkan teknologi informasi dalam bidang studi biologi yang dikenal dengan penelitian bioinformatika. Penelitian menggunakan data DNA sudah banyak diterapkan untuk menerjemahkan data tersebut ke dalam suatu informasi yang belum diketahui sebelumnya. Seperti contoh, penelitian mengenai data DNA ini didasarkan oleh penelitian yang telah dilakukan sebelumnya oleh Revolson Mege, Yeremia Mokosuli dan tim yang meneliti tentang hubungan filogenik Babi Liar dan babi lokal dari Sulawesi utara berdasarkan hormon pertumbuhan (GH

Gene) yang menunjukkan terdapat variasi gen CO1 babi lokal di Sulawesi Utara dan variasi juga ditemukan pada asam amino yang menjadi dasar seleksi babi lokal unggul Sulawesi Utara [2]. Sulawesi Utara merupakan salah satu daerah di Indonesia yang memiliki jumlah konsumsi serta ternak babi lokal yang cukup tinggi dan sudah banyak peternak – peternak babi lokal di berbagai daerah di Sulawesi Utara. Melalui penelitian tersebut telah dihasilkan kode genetik DNA yang telah berbentuk *file SEQ* yang telah diolah sebelumnya yang akan digunakan dalam penelitian ini. Pengurutan DNA dilakukan dengan menggunakan metode, algoritma dan *coding* serta metode untuk mendapatkan hasil berupa pengetahuan mengenai kedekatan, kemiripan atau kekerabatan dari suatu organisme melalui kode genetik dalam DNA. Sehingga dalam penelitian ini, penulis akan melakukan implementasi algoritma Needleman–Wunsch dalam pengujian tingkat kemiripan DNA Babi Sulawesi Utara.

II. TINJAUAN PUSTAKA

A. Bioinformatika

Bioinformatika merupakan cabang ilmu komputasi dari biologi molekuler yang dapat didefinisikan sebagai teknologi penambangan, penyimpanan, analisis, interpretasi, penyebaran dan aplikasi dari informasi biologis. Menurut Fredj Tekaia dari Institut Pasteur, bioinformatika merupakan metode matematika, statistik dan komputasi yang bertujuan untuk menyelesaikan masalah – masalah biologi dengan menggunakan DNA Sequence, asam amino dan informasi – informasi yang terkait [3]. Dari pendapat ahli dapat dilihat bahwa bioinformatika menjadi solusi yang tepat untuk mengolah data biologis seperti urutan DNA, dengan menggunakan tools komputasi dalam informatika untuk melakukan analisis sehingga data biologis tersebut dapat dipahami dengan mudah. Untuk tools yang dapat digunakan salah satunya adalah bahasa pemrograman python dan perhitungan dengan algoritma seperti Algoritma Needleman-Wunsch untuk mempermudah dalam pengolahan data dengan berbagai keragamannya. Analisis dalam bioinformatika dapat difokuskan dalam tiga dataset yaitu urutan genom, uji fungsional genomik serta struktur dari makromolekul. Namun penerapan analisis bioinformatika dapat juga dilakukan seperti analisis pohon taksonomi, teks artikel ilmiah dan statistik juga hubungan metabolic. Dan macam – macam teknik yang dapat dipakai untuk penerapan adalah pencocokan sequence, 3D Protein Structure, prediksi dan klasifikasi struktur dari protein, struktur RNA, fungsi protein, ekspresi kluster data serta konstruksi pohon filogenetik [4].

B. DNA

DNA atau Asam Deoksiribonukleat (deoxyrinucleic acid) merupakan biomolekul yang berfungsi menyandi informasi genetika dari organisme dan beberapa jenis makhluk mikroba. Molekul dari DNA memiliki peran

dalam menyimpan informasi berjangka Panjang. [5]. Yang pada dasarnya semua makhluk memiliki informasi genetik yang berbeda-beda disimpan dalam DNA, meskipun masih termasuk dalam satu ras yang bentuknya berupa karakter-karakter yang merupakan kode genetik. Hal ini menjadi salah satu bagian dari tujuan penelitian ini dalam melihat langsung perbedaan dalam bentuk persentase angka.

C. Biologi Molekuler

Biologi molekuler adalah cabang ilmu pengetahuan yang merupakan multi disiplin ilmu dari biologisel, biokimia dan genetika yang mempelajari aktivitas biologi pada tingkat molekuler, termasuk interaksi antara perbedaan tipe DNA, RNA, protein dan biosintesisnya. Aktivitas atau mekanisme apa yang terjadi pada tingkat molekuler sangat penting untuk dipelajari karena melalui tingkat terkecil dipelajari maka pengaruh ke tingkat yang lebih tinggi akan terlihat sehingga dapat menunjukkan gen apa yang mempengaruhi suatu organisme atau suatu penyakit genetik, dalam melakukan identifikasi gen, identifikasi DNA, identifikasi DNA forensik untuk pengenalan organisme serta terapi gen untuk mengobati dan mencegah penyakit dan lain sebagainya [6].

D. Babi Lokal

Babi lokal adalah babi hutan yang didomestikkan. Dengan kata lain, babi hutan yang ditenak oleh warga lokal dengan pemeliharaan khusus yang membuat babi hutan tersebut bermutasi sehingga memiliki bentuk serta untuk dikonsumsi sudah memiliki rasa yang berbeda karena faktor makanan dari peternak babi lokal [2].

E. Sequence Alignment

Sequence Alignment atau penjajaran barisan merupakan metode dalam mengatur urutan dari DNA atau protein dengan tujuan untuk mengidentifikasi daerah kesamaan yang boleh saja menjadi konsekuensi hubungan secara fungsional, struktural atau evolusi dari antara *sequence DNA* [7]. Terlihat dari keberagaman data, DNA memiliki perbedaan dari segi kode genetiknya maupun ukuran dari urutan tersebut yang memiliki panjang yang berbeda-beda. Dengan *sequence alignment* dapat melakukan penjajaran untuk setiap urutan sehingga terdapat kesamaan yaitu urutan DNA yang lebih optimal, dapat diidentifikasi daerah kesamaan atau hubungan secara menyeluruh.

F. BioPython

Proyek Biopython adalah asosiasi internasional dari pengembang *tools* Python yang tersedia secara gratis untuk komputasi biologi molekuler [8]. Yang pada dasarnya, tujuan dari Biopython adalah untuk mempermudah dalam menggunakan Bahasa pemrograman Python untuk penerapannya dalam bioinformatika dengan membuat modul dan kelas berkualitas tinggi yang dapat dimanfaatkan [9]. Karena penelitian ini merupakan penelitian bioinformatika dan Bahasa pemrograman Python, maka cocok untuk

penggunaan modul biopython karena dapat mempermudah penelitian dalam pengolahan data DNA.

G. Pairwise Sequence Alignments

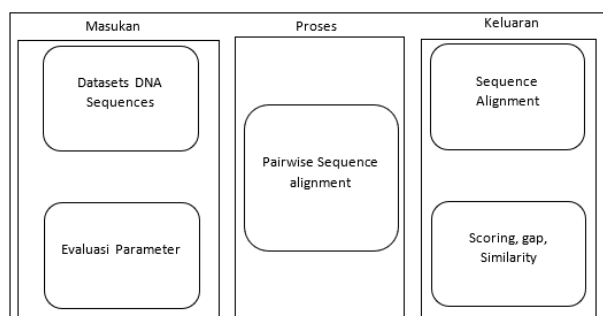
Pairwise Sequence Alignment adalah proses pencocokan antara dua sekuen DNA berdasarkan kesamaan dan untuk mencari kesamaan di database jika dimodelkan untuk *multiple data*. Kesamaan ini dapat dinyatakan dalam persentase [10]. Penelitian ini memiliki hasil dalam bentuk persentase untuk melihat kecocokan dengan menggunakan beberapa data DNA yang diolah. *Pairwise sequence alignment* metode yang cocok untuk dimodelkan karena pencocokan yang sesuai dengan tujuan dari penelitian ini. *Pairwise sequence alignment* dapat digunakan walaupun data yang digunakan tidak dalam database melainkan dapat dilakukan untuk data yang disesuaikan sebagai data utama untuk dicocokkan dengan data – data lainnya.

H. Algoritma Needleman-Wunsch

Merupakan algoritma yang ditemukan pada tahun 1970 oleh Saul B. Needleman Bersama Christian D. Wunsch. Algoritma ini populer dalam penelitian *global sequence alignment* tahapan dalam melakukan pensejajaran dimulai dari *scoring matrix*, *traceback* dan *alignment* [11]. Penggunaan algoritma Needleman-Wunsch ini akan menghasilkan pensejajaran urutan DNA. Selain perbedaan kode genetik dalam data DNA, data DNA yang memiliki ukuran yang berbeda – beda. Sehingga jika dicocokkan dengan ukuran yang berbeda, jadi persentase yang didapatkan kurang optimal karena hanya mencakup kemiripan dari keseluruhan data DNA yang punya ukuran yang lebih pendek, sedangkan untuk ukuran yang lebih panjang akan tersesa kode genetik yang tidak dicocokkan. Algoritma Needleman-Wunsch dapat melakukan penjajaran dari urutan DNA sehingga didapatkan urutan DNA yang sejajar dan optimal untuk dilakukan pencocokan dan hasil yang optimal pula karena hasil pencocokan berdasarkan keseluruhan data DNA.

III. METODE

Dalam penelitian ini, peneliti menggunakan metode pairwise sequence alignment untuk melakukan sequence alignment terhadap urutan DNA babi lokal [12].



Gbr. 1 Metode Pairwise Sequence Alignment

Sequence alignment merupakan proses dimana urutan DNA dibandingkan dengan tujuan untuk mencari pola dari karakter pada bagian yang paling umum dengan

urutan yang saling terkait. Pairwise Sequence Alignment adalah proses perataan antara dua urutan DNA berdasarkan kesamaan dari urutan DNA yang dinyatakan dalam persentase. Yang artinya urutan DNA itu memiliki kemiripan dengan sifat fisika-kimia. Kesamaan juga dapat menunjukkan hubungan fungsional atau hubungan struktural dan evolusi [13].

A. Datasets DNA Sequence

Dataset yang digunakan adalah data yang merupakan data dalam bentuk FASTA urutan DNA Babi lokal.

B. Evaluasi Parameter

Untuk melakukan *Sequence Alignment* perlu dilakukan pergeseran urutan pada posisi tertentu dari urutan agar mendapatkan kesamaan yang maksimal. Ada dua acara dalam melakukan perataan urutan DNA yaitu *Global Sequence Alignment* dan *Local Sequence Alignment*. Dalam Penelitian ini akan digunakan *Global Sequence Alignment* yaitu membandingkan dengan keseluruhan urutan. Cara ini cocok untuk melakukan penyelarasan urutan DNA yang mempunyai derajat kemiripan yang tinggi dari keseluruhan urutan.

C. Pairwise Sequence Alignment

Untuk melakukan *Pairwise Sequence Alignment*, peneliti menggunakan Algoritma Needleman-Wunsch untuk menemukan kesejajaran urutan optimal secara global dan memungkinkan adanya suatu cela. Dan peneliti menggunakan Algoritma Needleman-Wunsch untuk menguji tingkat kemiripan DNA drngan cara *global sequence alignment*. Dengan tahapan sebagai berikut [5].

1) Inisialisasi Tabel

Untuk melakukan *sequence of 2 alignment* diberikan S1 untuk urutan DNA pertama dan S2 untuk urutan DNA kedua. Dengan panjang dari urutan adalah m dan n , langkah awal ini dimulai dengan membuat tabel dengan $m + 1$ sebagai kolom dan $n + 1$ untuk baris.

2) Pengisian Tabel

Tabel diisi dengan melakukan hubungan rekursif. Dengan rumus,

$$F(i, j) = \max \begin{cases} F(i - 1, j - 1) + S(x_i, y_i) \\ F(i - 1, j) + S(indel) \\ F(i, j - 1) + S(indel) \end{cases}$$

3) Pelacakan Kembali (Traceback)

Traceback adalah langkah akhir dalam algoritma Needleman-Wunsch. Pada tahapan ini, digunakan tabel yang telah terisi yang membentuk matriks untuk memperoleh *sequence* yang terbaik yaitu pensejajaran yang opetimal dari ke dua *sequence*. Dengan rumus,

$$F(i, j) = \max \begin{cases} F(\nearrow) + S(Match/Mismatch) \\ F(\uparrow) + S(indel) \\ F(\leftarrow) + S(indel) \end{cases}$$

D. Sequence Alignment

Pada tahapan ini ditampilkan Urutan DNA yang telah dilakukan pensejajaran atau DNA yang telah optimal ditampilkan. DNA yang telah diurutkan dapat ditampilkan dalam bentuk matriks dengan menggunakan *dotplot method*.

E. Similarity

Pada tahapan ini, dilihat tingkat kemiripan dari urutan DNA babi lokal Sulawesi Utara yang dianalisis. Pada tahapan ini peneliti menggunakan modul *BioPython* yang ada pada Bahasa pemrograman Python yaitu modul *pairwise2*.

IV. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini mengacu pada metode *Sequence Alignment* dan pensejajaran dari urutan DNA dengan penerapan algoritma Needleman-Wunsch dan menggunakan modul *BioPython* yaitu *pairwise2* dalam menguji tingkat kemiripan dari urutan DNA yang telah dilakukan. Berikut merupakan tahapan yang telah dilakukan:

A. Datasets DNA Sequence

Data DNA yang diolah dalam penelitian ini ada 7 data DNA yaitu 6 data DNA babi lokal dan 1 data DNA Babi Hutan.

TABEL I. Data Urutan DNA

No	Urutan DNA	Kode
1.	> NNNNNNNNCTATATGTATTGTTACAGCTCATGCCT TTGTAATAATCTCTTTATAGTAATACCCATTATGATT GGGGGTTTTGGTAACTGACTCGTACCGCTAATAATCG GAGCTCCCGATATGGCCTTTCCACGTATAAACAACAT AAGTTTCTGACTACTTCCACCATCTTCTATTACTAC TGGCATCTCAATAGTAGAAGCCGGGCGGCTACTG GATGAACCGTATACCCACCTTTAGCTGGAACCTTAGC CCATGCAGGAGCTTCAAGTTGATCTAACAATTTTTC CTACACCTTGCAAGGTGATCATCAATCTAGGGGCTA TTAATTTTCAATACCAAAATTTAACAATAAACCTCC CGCAATGTCTCAATACCAAAACCCCTGTTGTCTGA TCAGTACTAATACAGCGGTACTACTTACTATCCC TGCCAGTTCTAGCAGCTGGCATTACTACTACTGAC AGACCCGCAACCTAAA	
2.	> NNNNNNNNCTGCTNACTGGCAGGNNAGTAGAAGTA GTACGGCTGTGATTAGTACTGATCAGACAAACAGGG GTGTTTGGTATTGAGACATTGGCGGAGGTTTATGTT AATAATTGTTGTAATGAAATTAATAGCCCTAGGATT GATGATACACCTGCAAGGTGTAGGGAAAAAATTGTT AGATCAACTGAAGCTCTGCAATGGGCTAAGTTTCCAG CTAAAGGTGGGTATACGGTTTATCCAGTACCCGCCCC GGCTTCTACTATTGAGGATGCCAGTAGTAATAGGAAG GATGGTGAAGTAGTCAGAACTTATGTTGTTTATAC GTGGAAGGCCATATCGGGAGCTCCGATTATTAGCG GTACGAGTCAGTTACCAAAACCCCAATCATAATGG GTATTACTATAAGAAGATTATTACAAAGGCATGAG CTGTAACAATTACATTATAGATTGATCATCGCCAAG TAGAGTTCCGGGCTGACCTAAA	
3.	> NNNNNNNNCTATATGTNNTGTTACAGCTCATGCCTT TGTAATAATCTCTTTATAGTAATACCCATTATGATT GGGGTTTTGGTAACTGACTCGTACCGCTAATAATCGG AGCTCCCGATATGGCCTTTCCACGTATAAACAACATA AGTTTCTGACTACTTCCACCATCTTCTATTACTACT GGCATCTCAATAGTAGAAGCCGGAGCGGCTACTGG ATGAACCGTATACCCACCTTTAGCTGGAACCTTAGCC CATGCAGGAGCTTCAGTTGATCTAACAATTTTCTCC TACACCTTGCAAGGTGATCATCAATCTAGGGGCTAT TAATTTTCAATACCAAAATTTAACAATAAACCTCCC GCAATGTCTCAATACCAAAACCCCTGTTGTCTGAT CAGTACTAATACAGCGGTACTACTTCTACTATCCCT GCCAGTTCTAGCAGCTGGCATTACTACTACTGACA GACCCGCAACCTA	

No	Urutan DNA	Kode
4.	> NNNNNNNNNNCTGCTAGACTGGCAGGNNAGTAGAAG TAGTACGGCTGTGATTAGTACTGATCAGACAAACAG GGGTGTTTGGTATTGAGACATTGCGGGAGGTTTTATG TTAATAATTGTTGTAATGAAATTAATAGCCCTTAGGA TTGATGATACACCTGCAAGGTGTAGGGAGAAAAATTG TTAGATCAACTGAAGCTCCTGCAATGGGCTAAGTTTCC AGCTAAAGGTGGGTATACGGTTCATCCAGTACCCGCT CCGGCTTCTACTATTGAGGATGCCAGTAGTAATAGGA AGGATGGTGAAGTAGTCAGAACTTATGTTGTTTAT ACGTGGAAGGCCATATCGGGAGCTCCGATTATTAG CGGTACGAGTCAGTTACCAAAACCCCAATCATAATG GGTATTACTATAAGAAGATTATTACAAAGGCATGA GCTGTAACAATTACATTATAGATTGATCATCGCCAA GTAGAGTTCCGGGCTGACCTAAA	
5.	> NNNNNNNNNNCTATATGTAATTGTTACAGCTCATGCC TTTGTATAATCTCTTTATAGTAATACCCATTATGAT TGGGGGTTTTGGTAACTGACTCGTACCGCTAATAATC GGAGCTCCCGATATGGCCTTTCCACGTATAAACAACA TAAGTTTCTGACTACTTCCACCATCTTCTATTACTA CTGGCATCTCAATAGTAGAAGCCGGGCGGCTAGT GGATGAACCGTATACCCACCTTTAGCTGGAACCTTAG CCCATGCAGGAGCTTCAAGTTGATCTAACAATTTTCT CCTACACCTTGCAAGGTGATCATCAATCTAGGGGCT ATTAATTTCAATACCAAAATTTAACAATAAACCTCC CCGCAATGTCTCAATACCAAAACCCCTGTTGTCTG ATCAGTACTAATACAGCGGTACTACTTCTACTATCC CTGCCAGTTCTAGCAGCTGGCATTACTACTACTGTA CAGACCCGCAACCTAAA	
6.	> NNNNNNNNCTGCTAGACTGGCAGGNNAGTAGAAGTA GTACGGCTGTGATTAGTACTGATCAGACAAACAGGG GTGTTTGGTATTGAGACATTGGCGGAGGTTTATGTT AATAATTGTTGTAATGAAATTAATAGCCCTAGGATT GATGATACACCTGCAAGGTGTAGGGAAAAAATTGTT AGATCAACTGAAGCTCTGCAATGGGCTAAGTTTCCAG CTAAAGGTGGGTATACGGTTTATCCAGTACCCGCCCC GGCTTCTACTATTGAGGATGCCAGTAGTAATAGGAAG GATGGTGAAGTAGTCAGAACTTATGTTGTTTATAC GTGGAAGGCCATATCGGGAGCTCCGATTATTAGCG GTACGAGTCAGTTACCAAAACCCCAATCATAATGG GTATTACTATAAGAAGATTATTACAAAGGCATGAG CTGTAACAATTACATTATAGATTGATCATCGCCAAG TAGAGTTCCGGGCTGACCTAAA	
7.	>EF113597.1 Sus scrofa selenoprotein X mRNA, complete cds AGGCTCAAGACCCCTGCGGTGGAGCTGTGCTCTCGGTT CCGCCGGAGCTTCCCGCTCGGCCCATGTGCTTCTGCT AGCTTCTTCCGGGGCGAGGTTTCCAGAACCACTTTG AGCCAGGTGTCTACGTGTGTGCCAAGTGTGGCTATGA GCTCTTCTCCAGCCTCAAAAGTACGCGCACTCGTCC CCATGGCGGGCTTACCCGAGACCATCCAGCTGACA GTGTGGCAAGCGTCCAGAGCACAAATCGGCTGGAG CCTTAAAGGTGTCTGTGGCAGATGTGGCAACGGACT GGGCCATGAGTTCTGAACGATGCCGCCAAGCGGGG ACAGTCCCGCTTCTGAATATTCAGCAGTTCCCTGAAG TTCATCCCTAAAGGCCAAGATCATCTCCCTCCAGG GACAGTAGGCGGGCAGCCACGCGCTTCCACAGTGA ACACTGCTTGGTGGCCATGCTTGGCGGGCCATCAT AGAAGTGAACCCAGCCATTACAGACTGAATGTCA GGCAGGCTTCCGAGGCTTGGCTGTGAACAGGCTCCT TGGGAGAAGGAGAAGGCCCATCTCGGCTGCCAGGG CGGCCCCGTGGGCTGGCTGTGGCTGGCACCAGG GTCTCAGGAGTGGGAGCCAGGCTGGCTCGTCACC CCCACCTTCTCAGCCATGGCCCTCCATGCCCTGTCT GGACTTCAACCCGGCGGGCCAGTTCTCAGATGGGC CTGGCTCAGCAAGCTCTACCCACCCCGCGCAGG CTGTGCTGGGACAGAAGGTGAGCAGAGGCGGCCAG CCGGCTGTCCAGGCTCAGTATGATACACTTGGC TAAACCTTTCAGGCGGCCAGAGGGGTGATGGTCTG TGACTCCGAGGGGAGCAGGCGGAGGGGGGGTCT CTGGGCTTTGAGTTACGAAGAGAAATCTCGGACCTT CTGGCACAGCTGTATGGGTGTGCTCTCCAGCTCA GGCTTAGGAAAGCAGCTTCTGCTCCACAGCCTGATC GGGCGCGGCCAGGGCCGAGCTCCAGCCTCAGCT CGCGCTTTTGGAGAGAAGGTGGCCGTGCTTACTGC GCATGCTTTGAGGCTCGTGTGGCGGGGAGGGTGGG TGAAGCCCTCCTGCTGAATGCTCTGTAATCGGGAG GCGATCAATACACAAATGGAACCGGAAAAA AAAAA	

Dari data DNA yang ada, peneliti menggunakan data DNA Babi Tipe L1F sebagai tolak ukur untuk menguji tingkat kemiripan dengan DNA Babi lainnya yang didefinisikan sebagai data DNA Babi Lokal.

B. Evaluasi Parameter

```
print ("Jumlah Sequence DNA Babi Lokal : ",len(babi_lokal_seq))
print ("Jumlah Sequence DNA Babi Hutan : ",len(babi_hutan_seq))
print ("Jumlah Sequence DNA Babi Sample B1F : ",len(babi_sampleB1F_seq))
print ("Jumlah Sequence DNA Babi Sample B1R : ",len(babi_sampleB1R_seq))
print ("Jumlah Sequence DNA Babi Sample L1R : ",len(babi_sampleL1R_seq))
print ("Jumlah Sequence DNA Babi Sample M1F : ",len(babi_sampleM1F_seq))
print ("Jumlah Sequence DNA Babi Sample M1R : ",len(babi_sampleM1R_seq))
```

Gbr. 2 Code Evaluasi Parameter

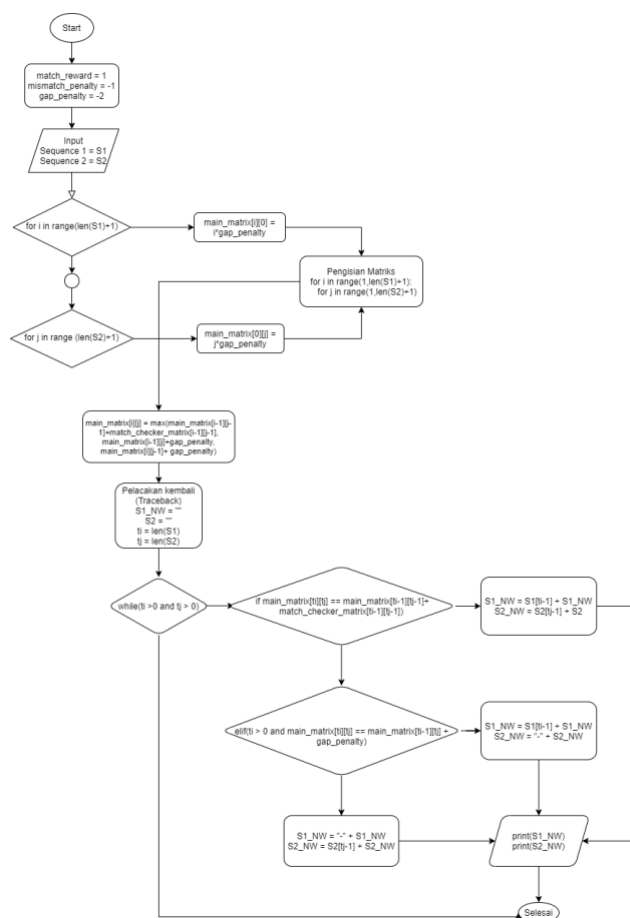
Dari code tersebut didapatkan hasil sebagai berikut.

```
Jumlah Sequence DNA Babi Lokal : 498
Jumlah Sequence DNA Babi Hutan : 1215
Jumlah Sequence DNA Babi Sample B1F : 495
Jumlah Sequence DNA Babi Sample B1R : 500
Jumlah Sequence DNA Babi Sample L1R : 497
Jumlah Sequence DNA Babi Sample M1F : 499
Jumlah Sequence DNA Babi Sample M1R : 498
```

Gbr. 3 Hasil Evaluasi Parameter

Dari hasil tersebut terlihat parameter dari tiap DNA berbeda. Sehingga untuk menguji kemiripan lebih optimal, peneliti memilih untuk melakukan *global sequence alignment*.

C. Pairwise Sequence Alignment



Gbr. 4 Flowchart code dari algoritma Needleman-Wunsch

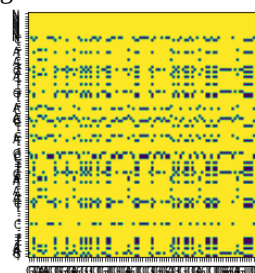
```
NNNNNNNNNNCT-ATATGTA-TTGT-TAC-AGCTCATGC
CTTTG--TAATAATCTTCTTTATAGTAATAC-CCA--TT
-AT-GATTG-GGGGTTTT-GGTAACGAC-TCGTACC-G
CTAATAATCG-GAGCTCCCGA-TATGGC-CTTTCCACGT
AT-AAACAAC-ATAAGT-TTCTGACTACTTCCACCAT--
CCTTCCTATTACTACTGGCATCCTCAATAGTAGAAGCCG
GGGCGGGTACTGGATGAACCGTATACCCACCTTTAGCTG
GAACTTAGCCCAT-GCAGGAGCTTCAGTTGATCT-AAC
AATTTTTTCCCTACAC-CTTG-CAGGTGTATCATCAATC
CTAGGGGCT-ATTAAT-TTCA-TTACCACAA-TTATTAA
CATAAAACCTCCCGCAAT-GTCTCAATACCAACACCCC
TGTTTGTCTGATCA-GTACTAATCAGCCGTACTACTT
CTACTATCCCTGCCA-GTTCTAGCAGCTGGCATTACTAT
ACTACTGACAGACCCGCAACCTAAA
```

```
NNNNNNCTGCTNACTGGCAGGNNAGTAGAAGTAGTACG
GCTGTGATTAGTACTGATCAGAC-AAACAGGGGTGTTT
GTATTGAGACATTGCGGGAGGTTTT-ATGTTAATAATTG
TG-GTAATGAAATT-AATAGCCCCTAGGATTGATGATAC
ACCTGCAAGGTGTAGGGAATAATTG-TTAGA-TCAACT
GAAGCTC-CTGCATGGGCTAAGTTTCCA-G-C-TAA-AG
GTGGGTATACGGTT-C-ATCCAG-TACCCGCCCGGCTT
CTACTATTGAGGATGCCAGTAGTAATAGGAAG-GATGGT
GGAAGTAG-TCAGAACTTATGTTGTTTATACGTGGAAA
G-GCCATATCGGGAGCTCCGATTATTAGCGGTACGAG-T
CAGTTACCA-AAAC-CCCCAATCATAATGGGTATTACTA
TAAAGAAGATTATTACAAAG--GCATG--AGC-TGTA-A
CAATT-ACATTATAGATTTGA-T--CATCGC-CAAGTAG
AGTTCCG-GG-CTGA-CCTAA---A
```

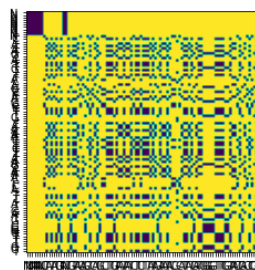
Urutan DNA tersebut adalah urutan DNA yang telah dioptimalkan atau disejajarkan dengan pergeseran beberapa karakter dalam urutan atau biasa disebut dengan *indel* yang disimbolkan dengan “-“ yang membuat dua urutan tersebut menjadi sama panjang untuk nantinya dicocokkan dan diuji tingkat kemiripannya.

D. Sequence Alignment

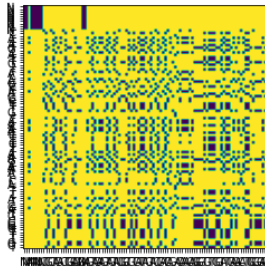
Pencocokan dibuat dalam bentuk matriks yang dapat digambarkan sebagai berikut.



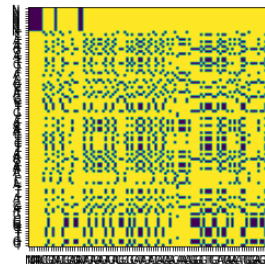
Gbr. 5. Matriks DNA babi lokal dan babi hutan



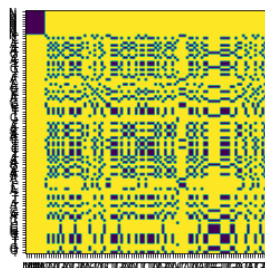
Gbr. 6 Matriks DNA babi lokal dan DNA Babi B1F



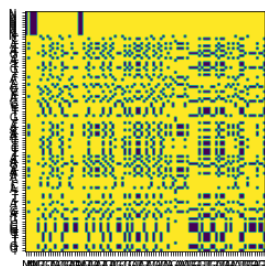
Gbr. 7 Matriks DNA babi lokal dan DNA babi B1R



Gbr. 8 Matriks babi lokal dan DNA babi L1R



Gbr. 9 Matriks babi lokal dan DNA babi M1F



Gbr. 10 Matriks babi lokal dan DNA babi M1R

Matriks yang ditampilkan yaitu pada bagian vertikal atau sumbu y berisikan urutan DNA dari Babi Lokal dan untuk bagian horizontal atau sumbu x berisikan urutan DNA Babi tipe lainnya yang akan dicocokkan dengan DNA Babi Lokal. Gambar tersebut merupakan matriks yang diperkecil agar dapat melihat bentuk atau pola kecocokan dari kedua urutan yang dapat disimpulkan bahwa semakin berwarna gelap matriks tersebut maka akan semakin tinggi tingkat kemiripan dari kedua urutan DNA.

E. Similarity

Hasil dari tahapan ini adalah berupa persentase tingkat kemiripan dari urutan DNA Babi lokal dan urutan DNA Babi tipe lainnya. Dan dalam penelitian ini, peneliti menggunakan modul *pairwise2* dari *BioPython* untuk menghitung persentase tingkat kemiripan.

```
from Bio import pairwise2
from Bio.pairwise2 import format_alignment

kemiripan = pairwise2.align.globalxx(babi_lokal_NW[0:100],data_NW[0:100],
                                     one_alignment_only=True,score_only=True)

print("Tingkat Kemiripan :",kemiripan,"%")
```

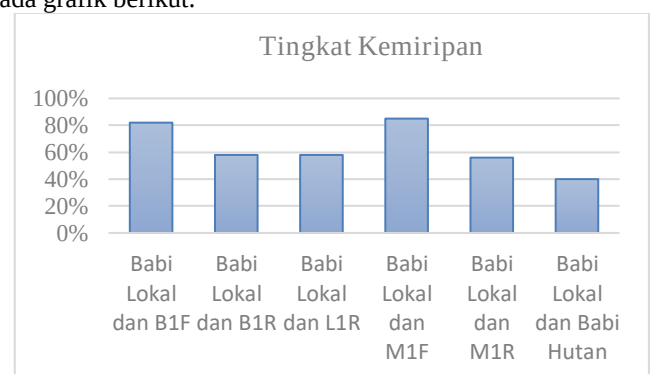
Gbr. 11 Coding modul *pairwise2*

Dari *coding* tersebut, perhitungan *Similarity* dibuat dalam fungsi Python. Dimana, **babi_lokal_NW** menyimpan data DNA Babi lokal dan **data_NW** menyimpan urutan data DNA yang akan dicocokkan yang telah disejajarkan dengan Algoritma Needleman-Wunsch. Dan perhitungan ini menunjukkan persentase tingkat kemiripan dari dua urutan DNA yang dapat dilihat sebagai berikut.

1. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Hutan adalah 40.0 %
2. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Tipe B1F adalah 82.0 %
3. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Tipe B1R adalah 58.0 %
4. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Tipe L1R adalah 58.0 %
5. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Tipe M1F adalah 85.0 %
6. Tingkat kemiripan dari urutan DNA Babi Lokal dan DNA Babi Tipe M1R adalah 56.0 %

V. KESIMPULAN

Dari penelitian ini, dapat disimpulkan bahwa telah dilakukan pengujian tingkat kemiripan DNA babi Sulawesi Utara dengan pengimplementasian algoritma Needleman-Wunsch dan pensejajaran urutan DNA menggunakan global sequence alignment berhasil menghasilkan urutan DNA yang telah optimal disejajarkan. Dan setelah dilakukan pencocokan dengan menggunakan modul *pairwise2* pada *BioPython* ditemukan hasil persentase dari tingkat kemiripan urutan DNA yang dianalisa, yaitu dapat dilihat pada grafik berikut.



Gbr. 12 Grafik Tingkat Kemiripan

Kiranya penelitian ini dapat menjadi referensi tambahan untuk penelitian bioinformatika dan bidang lainnya serta model yang telah dibuat dapat menjadi dasar untuk melakukan penelitian serupa di kemudian hari.

DAFTAR PUSTAKA

- [1] Y. P. P. Chen, *Bioinformatics Technologies*, 2005.
- [2] A. M. Revolson, Y. S. Mokosuli, J. J. R. Debby, H. A. Ellen, R. Christny, M. Nony och M. Merry, "Phylogenetic relationship of wild pigs and local pig from north sulawesi based on the growth hormone gene (Gh gene)," *Materials Science Forum*, vol. 967 MSF, nr August, pp. 71-82, 2019.
- [3] Seprianto, "Modul Mata Kuliah Pengantar Bioinformatika," nr Ibt 431, pp. 1-86, 2017.
- [4] V. Narita, A. L. Arum, S. I. M och N. Y. Fawzya, "Analisis Bioinformatika Berbasis WEB untuk Eksplorasi Enzim Kitosanase Berdasarkan Kemiripan Sekuens," *Al-Azhar Indonesia Seri Sains dan Teknologi*, vol. I, nr 4, pp. 197-203, 2012.
- [5] H. Kholiq, M. U. Romdhini och M. Susanto, "Algoritma Needleman-Wunsch dalam Menentukan Tingkat Kemiripan Urutan DNA Rusa Timor (*Cervus timorensis*) dan Rusa Merah (*Cervus elaphus*)," *EIGEN MATHEMATICS JOURNAL*, vol. 3, nr 2, 2020.
- [6] I. A. Wahyudi, "RESENSI Biologi Molekuler dalam Ilmu yang menyenangkan dan mudah," *TEKNOSAINS*, vol. IV, nr 2, 2015.
- [7] A. A. Sunarto, "Perbandingan Program Sequence Alignment," *Jurnal Rekayasa Nusaputra*, vol. 1, nr 1, pp. 1-5, 2015.
- [8] P. S. Foundation, "Python," 2014. [Online]. Available: <http://www.python.org>. [Använd 30 11 2020].
- [9] J. Chang, B. Chapman, I. Friedberg, T. Hamelryck, M. D. Hoon och P. Cock, "Biopython Tutorial and Cookbook," *Update*, vol. 2010, nr November, pp. 15-19, 2010.
- [10] J. Xiong, "Pairwise Sequence Alignment," *Essential Bioinformatics*, pp. 31-50, 2012.
- [11] M. A. Malendes och H. Bunyamin, "Perbandingan Needleman-Wunsch dan Lempel-Ziv dalam Teknik Global Sequence Alignment: Keunggulan Faktorisasi Sempurna," *Jurnal Teknik Informatika dan Sistem Informasi (JUTISI)*, vol. 3, nr 1, pp. 57-68, 2017.
- [12] L. Muflikhah och E. Santoso, "Pairwise sequence alignment between HBV and HCC using modified Needleman-Wunsch algorithm," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 15, nr 4, pp. 1785-1793, 2017.
- [13] H. C. Genome Campus, "EMBL-EBI," [Online]. Available: <https://www.ebi.ac.uk/Tools/psa/>. [Använd 13 2 2021].