

**PAIRWISE SEQUENCE ALIGNMENT
USING DIALIGN ALGORITHM**



I N A Y A H

0303010222



**UNIVERSITY OF INDONESIA
FACULTY OF MATHEMATICS AND SCIENCE
DEPARTMENT OF MATHEMATICS
DEPOK
2008**

**PAIRWISE SEQUENCE ALIGNMENT
USING DIALIGN ALGORITHM**

**This *skripsi* is submitted in partial fulfillment
of the requirements for the degree of
*Sarjana Sains***

By:

I N A Y A H

0303010222



DEPOK

2008

SKRIPSI : PAIRWISE SEQUENCE ALIGNMENT USING
DIALIGN ALGORITHM

NAME : INAYAH

NPM : 0303010222

THIS *SKRIPSI* HAS BEEN APPROVED

DEPOK, JULY 14, 2008

Dra. Denny Riama Silaban, M.Kom

ADVISOR I

Dra. Siti Aminah, M.Kom

ADVISOR II

Date of completion of *sidang sarjana* examination: July 14, 2008

Examiner I : Dra. Denny Riama Silaban, M.Kom

Examiner II : DR. Dian Lestari

Examiner III : Drs. Suryadi MT., M.T

Dedicated to

ibu and papah,

thank you so much for your patience, love, and support.

"Ya Allah, sayangilah kedua orang tuaku jauh melebihi kasih sayang mereka kepadaku. Muliaikanlah mereka berdua di dunia dan di akhirat. Kabulkanlah Ya Rabbi."

It's not the end, but a fight has just begun. Bismillah ...

"Ya Tuhanku, berilah aku ilham untuk mensyukuri nikmat Mu yang telah Engkau anugerahkan kepadaku dan kepada dua orang ibu bapakku dan untuk mengerjakan amal saleh yang Engkau ridhai dan masukkanlah aku dengan rahmat Mu kedalam golongan hamba-hamba Mu yang saleh."

(An Naml 19)

Only in a consciousness that Allah creates everything for good, one's heart will find peace. Subhanallah ...

ACKNOWLEDGEMENT

Alhamdulillahirobbil'alamin, thank God for the graces You bring to my life. "Phiuh, finally I made it!". This skripsi is the result of my study whereby I have been accompanied and supported by many people. I would thank to:

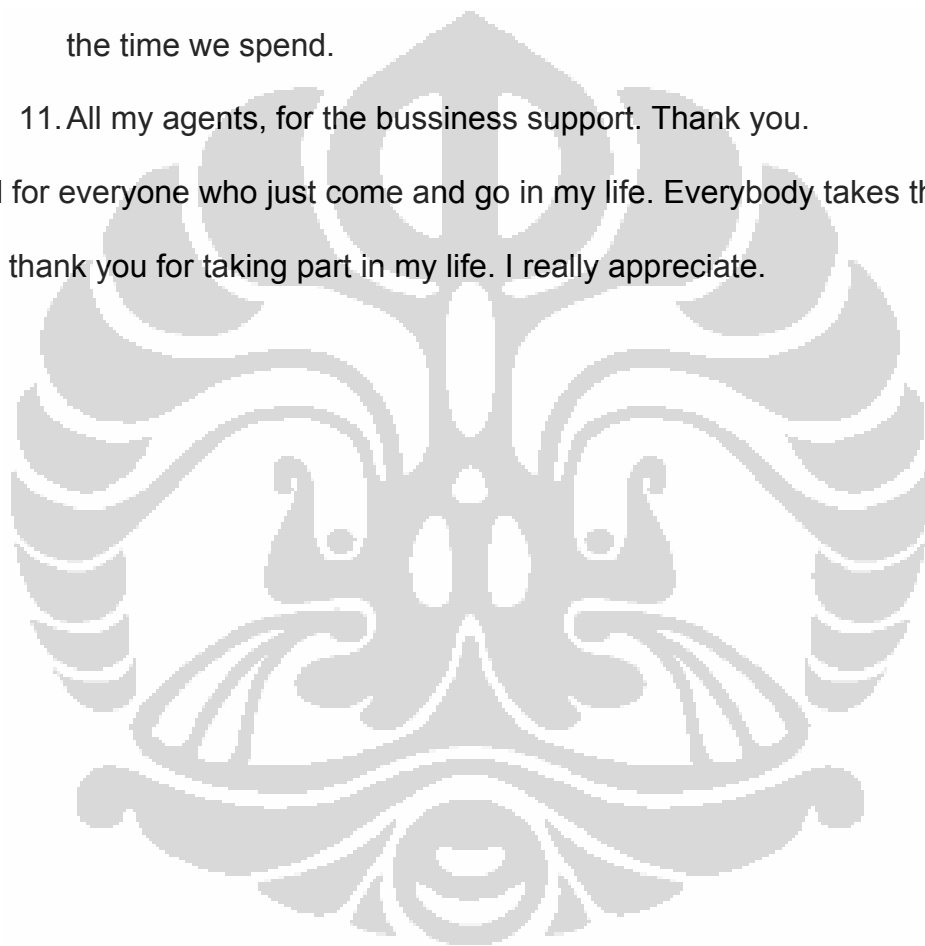
1. Dra. Denny Riama Silaban, M.Kom, my advisor, for the valuable guidance and support through out the skripsi. There is nothing I can say except thank you Ma'am.
2. Dra. Siti Aminah, M.Kom, my advisor, for the patience and support. Thank you Ma'am.
3. The Board of Mathematics Department, University of Indonesia, all the lecturers and the staffs (especially Mba' Santi) that help me through all of the academic stuffs.
4. Bembi, a very big thanks for you dude.
5. Redy Dwijaya, my precious *hohoho...*
6. My beloved sister and the whole family, i'm so proud to be a member of this family. Let me make you proud of me.
7. Bang Udin, for the spiritual support. Thank you so much.
8. A'yuni, *jazakillahu khairan katsir*, also Mel, Bong, and Anton, my *skripsimates*.
9. All members of mathematics year 2003, from A to Z (Adri, Andra, Arief, Asti, Delan, Dewi, Diki, Dody, Gele, Gewe, Gunung, Hadi,

Hetty, Igun, Ilham, Josua, Laras, Nana, Nita, Pinta, Puput, Putu, Rendie, Rima, Rina, Rini, Sonny, Tebe, Theja, Tony, Tyas, Utie, Yanthie, Yessa), thank you for your help during the class.

10. All seniors (Math '01 and '02) and juniors (Math '04, '05, and '06), for the time we spend.

11. All my agents, for the bussiness support. Thank you.

And for everyone who just come and go in my life. Everybody takes their part, and thank you for taking part in my life. I really appreciate.



Author
2008

ABSTRACT

This skripsi discusses a method known DIALIGN to find the best alignment of two DNA sequences. This algorithm is based on segment-to-segment comparison instead of the commonly used residue-to-residue comparison. Also, this algorithm avoids the wellknown difficulties concerning the choice of appropriate gap penalties. In DIALIGN, all possible diagonals of the input sequences will be weighted and compared to find the diagonals which compose optimal alignment. Diagonal weight is based on match probability of residues in the diagonal. Having the maximum score, the alignment will be constructed by tracing back the components which produce the maximum score. The resulted alignment can be considered as consistent collections of diagonals. In the final, the algorithm is implemented in a program. According to the simulation of the program, DIALIGN algorithm is able to produce optimal sequence alignment from a pair of sequence. And the program performs well on short sequences.

Keyword : Sequence alignment, DIALIGN algorithm, diagonal, scoring scheme, DNA

ix + 45 p.; app.

Bibliography : 7(1991-2007)

ABSTRAK

Skripsi ini membahas suatu metode yang biasa dikenal dengan nama DIALIGN untuk mencari penyejajaran terbaik dari dua barisan DNA. Algoritma ini berdasarkan pada perbandingan segmen dengan segmen bukan seperti yang biasa dilakukan yaitu perbandingan residu dengan residu. Selain itu, algoritma ini juga menghindari kesulitan dalam menentukan pemilihan untuk memberikan penalti yang tepat bagi *gap*. Pada DIALIGN, seluruh diagonal-diagonal yang mungkin dari input barisan yang diberikan akan diberi bobot dan dibandingkan dengan diagonal yang lain untuk mendapatkan diagonal-diagonal yang akan membentuk penyejajaran optimal. Penghitungan bobot dari diagonal berdasarkan pada probabilitas kesamaan residu pada diagonal. Setelah diperoleh skor maksimum, penyejajaran akan dibangun dengan cara menelusuri kembali komponen-komponen yang telah memproduksi skor maksimum. Penyejajaran yang telah dihasilkan merupakan sehimpunan diagonal-diagonal yang konsisten. Di akhir, algoritma DIALIGN diimplementasikan pada suatu program. Berdasarkan simulasi program, algoritma DIALIGN mampu memproduksi penyejajaran optimal dari sepasang barisan. Dan kinerja program sangat baik untuk barisan-barisan pendek.

Kata kunci : penyejajaran barisan, algoritma DIALIGN, diagonal, aturan pengskoran, DNA

ix + 45 hlm.; lamp.

Bibliografi : 7(1991-2007)

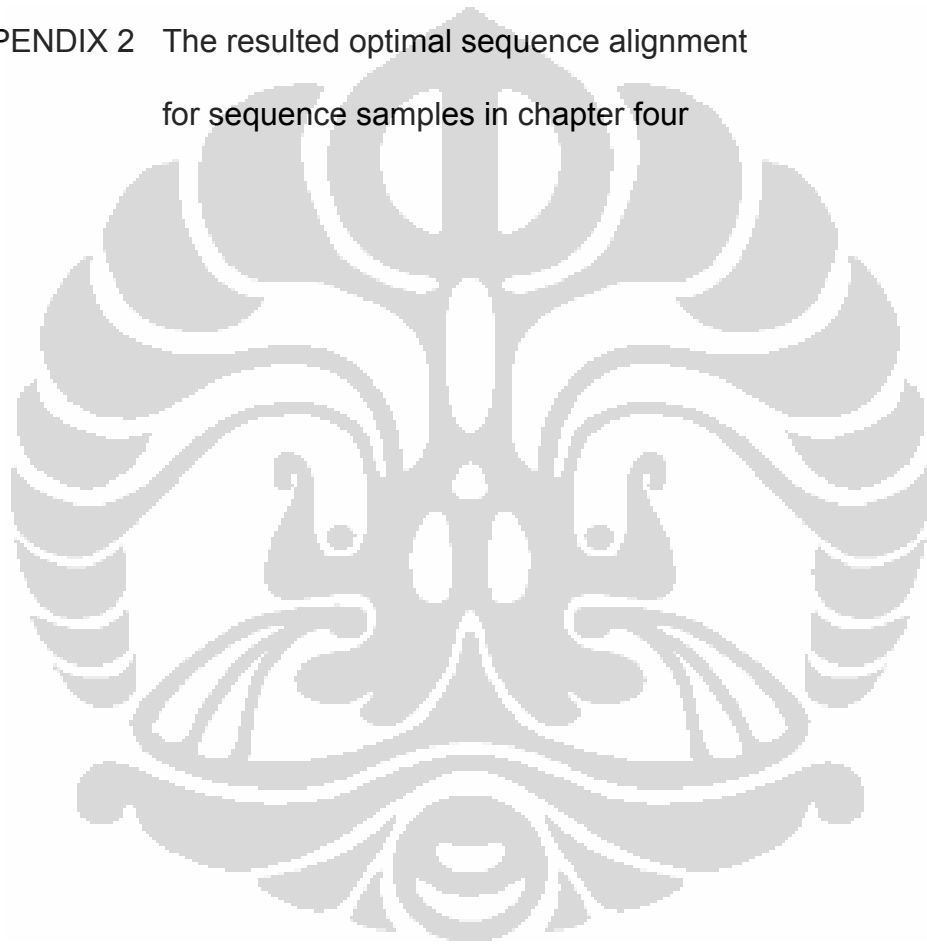
CONTENTS

	Page
ACKNOWLEDGEMENT	i
ABSTRACT	iii
CONTENTS	v
LIST OF APPENDIX	vii
LIST OF FIGURE	viii
LIST OF TABLE	ix
CHAPTER I PRELIMINARIES	1
1.1 Background	1
1.2 Problem Statement	3
1.3 Objective of Writing	4
1.4 Scope of Discussion	4
1.5 Structure	4
CHAPTER II CONCEPTS AND THEORIES OF SEQUENCE ALIGNMENT	5
2.1 Sequence Alignment	5
2.2 Sequence	8
2.3 Pairwise Sequence Alignment	10
2.4 Scoring Scheme	15
2.5 Local Alignment	17

CHAPTER III	DIALIGN ALGORITHM	18
3.1	Diagonal Construction	19
3.2	Diagonal Weight	21
3.3	Alignment Score	24
3.4	Backtracking Process	29
CHAPTER IV	IMPLEMENTATION OF DIALIGN ALGORITHM	34
4.1	Implementation	34
4.2	Simulation	37
CHAPTER V	CONCLUSION	43
REFERENCES		44
APPENDIX		45

LIST OF APPENDIX

	Page
APPENDIX 1 Program Listing	45
APPENDIX 2 The resulted optimal sequence alignment for sequence samples in chapter four	50



LIST OF FIGURE

Figure		Page
2.1	The new sequences which are formed from sequence Z	10
2.2	Some possible equal length segment pairs (diagonals)	12
3.1	Diagonals at position (3,2)	20
4.1	Content of sequen.txt	36
4.2	Input display for example 3.1	36
4.3	Output display for example 3.1	37
4.4	Output display for sequen1	37
4.5	Output display for sequen2	39
4.6	Output display for sequen3	40

LIST OF TABLE

Table		Page
3.1	Possible diagonals for every position	21
3.2	Diagonal weights of each position	23
3.3	Scores at each position	28
3.4	Alignment path	32
4.1	The number of optimal alignment and running time of the seven samples taken at random	41

