PROCEEDING

ISBN : 978-979-8969-06-5

International Conference on Biological Science
Faculty of Biology Universitas Gadjah Mada 2011
(ICBS BIO-UGM 2011)

ICBS 2011
Faculty of Biology Universitas Gadjah Mada

ADVANCES IN BIOLOGICAL SCIENCE

Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity

September 23rd-24th 2011
Yogyakarta, INDONESIA

Faculty of Biology Universitas Gadjah Mada

PROCEEDING

ISBN : 978-979-8969-06-5

International Conference on Biological Science
Faculty of Biology Universitas Gadjah Mada 2011
(ICBS BIO-UGM 2011)

ICBS 2011
Faculty of Biology Universitas Gadjah Mada

ADVANCES IN BIOLOGICAL SCIENCE

Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity

September 23rd-24th 2011

Yogyakarta, INDONESIA

FACULTY OF BIOLOGY
UNIVERSITAS GADJAH MADA

I-MHERE
PROJECT

PROCEEDING ICBS BIO-UGM 2011

Published By:
FACULTY OF BIOLOGY
UNIVERSITAS GADJAH MADA
YOGYAKARTA
Jl. Teknika Selatan Sekip Utara
Yogyakarta 55281
Phone: +62-274-580839; +62-274-6492350
Fax : + 62-274-580839
E-mail : biologi-ugm@ugm.ac.id

First Edition, October 2011

ISBN : 978-979-8969-06-5

CONTENT

PREFACE

Proceeding of the **International Conference on Biological Science Faculty of Biology Universitas Gadjah Mada 2011 (ICBS BIO-UGM 2011), Advances in Biological Science: Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity,** organized by and held at the Faculty of Biology, Universitas Gadjah Mada, Yogyakarta, Indonesia on September 23-24, 2011. The conference addressed a range of important research from various fields in biological science likely to play role tropical biodiversity management and conservation for supporting human prosperity. Three kinds of session were held at the conference: plenary session featuring keynote and invited papers, oral presentation session, and poster presentation session. This proceeding features a number of papers presented in these sessions, which represent 5 themes covered in the conference, i.e. genetics and molecular biology, ecology and conservation, systematics and evolution, physiology and developmental biology, and biomedics.

Many people have been involved in the production of these Proceedings, which is started in June 2011 with the launching of a call for abstracts. The abstracts were reviewed by both internal and external reviewers . Those selected abstracts were called for either oral or poster presentations and invited to submit full papers.

Lastly, on behalf of the organizing commite we would like to all participants for their kindness to be part of this conference. We would like to acknowledge each partnerships and sponsorship that involve during this event. I believe that this proceeding still has some weaknesses, therefore any constructive comments are welcome. We hope that the papers contain in this proceeding will prove helpful toward improving the scientific atmosphere. See you in the next two year ICBS 2013.

Yekti Asih Purwestri

Chair of the Organizing Commitee

WELCOMING SPEECH FROM CHAIR PERSON OF THE ORGANIZING COMMITTEE

Distinguish guests

- Executive Director of Indonesia-Managing Higher Education for Relevane and Efficiency (I-MHERE) Project
- Keynote speaker, invited speakers, participants, sponsorships, ladies and gentlemen

Good morning and May God shower us with His blessing.

On behalf of the Conference Organizing Committee, I extend a warm welcome to all participants to the second **International Conference on Biological Science Faculty of Biology Universitas Gadjah Mada 2011 (ICBS BIO-UGM 2011), Advances on Biological Science : Education for Sustainable Development-based tropical biodiversity management and conservation for supporting human prosperity.** Bio-conservation becomes a critical issue not only in Indonesia but also in global community. A good understanding on Education for Sustainable Development- based tropical biodiversity management is necessary to have the right policy regarding bio-conservation action.

For this year, the organizing committee has put together an interesting Scientific Program to accommodate the areas of Biology. The Program comprises of 6 plenary sessions of keynote and invited speakers. The parallel session of 82 oral presentations and more than 50 poster presentations. I realize that you are fully dedicated to the sessions but I do hope that you all will also take time to enjoy Yogyakarta, the multicultural city and may enjoy the special Merapi scenery, the most active volcano in the world.

I would like to take the opportunity to thank Prof Hubert Gijzen (Director of UNESCO-Jakarta) as a keynote speakers and also to these following invited speakers, Hao Yu, Ph.D (National University of Singapore), Prof. Christ Austin (Charles Darwin University, Australia), Prof. Yasumasa Bessho, Ph.D (Nara Institute of Science and Technology, Japan), Dr. Yam Tim Wing (Senior Researcher Orchid Breeding and Conservation Singapore Botanic Gardens), Drs. Langkah sembiring, M.Sc. Ph.D (Faculty of Biology, Universitas Gadjah Mada) for delivering their valuable scientific information.

To make this program happen, I would like to gratefully acknowledge to Indonesia-Managing Higher Education for Relevane and Efficiency (I-MHERE) which support this conference. We also thank to the valuable contributions from personal and institutional sponsorship and funding including Ms. Sachiko Iida, PT Diastika Biotekindo, PT Roche, Prima Grafika Yogyakarta., and Drs. Agus Suryanto - Indogama Yogyakarta.

I also gratefully thank to the Dean and Vices Dean of Biology Faculty, Universitas Gadjah Mada for giving us opportunity and support to organize this conference. My deep appreciation to the Steering Committee, the Academic Reviewers (internal and external: Dr. Sentot Santoso from Institut fuer Klinische Immunologie und Transfusionsmedizin, Justus Liebig Universitaet Giessen, Germany and Prof. Yasumasa Bessho, Ph.D from Gene Expression Research, Biological Sciences, Nara Institute of Science and Technology, Japan), members of the Organizing Committee for their strong support, active participation, cooperation and hard works in preparing and organizing this event a success.

It is inevitable that there is a lack in organizing this conference and I profoundly apologize to all invited speakers, oral and poster presenters, attendants, donators and committee members.

I wish you a pleasant and rewarding two days of scientific discussion.

Thank you,

Yekti Asih Purwestri
Chair person of the Organizing Committee

OPENING REMARKS FROM THE DEAN of THE FACULTY OF BIOLOGY

Bismillahirrahmaanirrahiim.

Director of UNESCO Office Jakarta, Prof. Dr. Hubert Gijzen,
Executive Direktor of Indonesian-Managing Higher education for relevance and Efficiency (I-MHERE) Project
Honorable speakers and distinguished guest, dear participants,

Assalamu'alaikum wr.wb., may God give us healthy and happier life

Welcome to Yogyakarta, the city of youth, education, and culture. It's been an honour for me to be here in front of you to open the prestigious **International Seminar with the special theme of "Advances in Biological Science: Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity",** that invited our honorable speaker from the UNESCO as the keynote, Prof. Hubert Gijzen, Ph.D honorable invited speakers Dr. Yam Tim Wing From Singapore Botanic Garden, Singapore; Prof. Yasumasa Bessho, MD, Ph.D from NAIST, Japan; Prof. Christopher M. Austin, Ph.D from Charles Darwin University, Australia; Dr. Yu Hao from National University of Singapore, and Dr. Langkah Sembiring MSc, from the Faculty of Biology, Universitas Gadjah Mada, Indonesia.

My special gratitute to the speakers who have spent your time travelling to Indonesia in your such busy activity. This international seminar atrracts more than 400 scholars and students mostly come from Indonesia, and some participants come from abroad. This occassion is such a good opportunity for us to share our experiences in research and good practices of ESD based research and community service done, that could inspire students and other researchers, furthermore our keynote speaker today is the Director of UNESCO Jakarta Office, who will talk about Science, Technology and Innovation-an Engine for Sustainable Development.

Honorable and distinguished participants,

The seminar theme taken today is in line with vission of the Faculty of Biology UGM as the center of excellence for higher education that generates biologists who respect to our tropical biodiversity. Since 2010, Faculty of Biology UGM had obtained an ESD based research grant from the World Bank, through I-MHERE (Indonesian Management of Higher Education for Efficiency and Relevance) project. In this project has been conducted 3 activities, these are: improvement of publication and research quality, improvement of integrated collaboration research in tropical diversity with other Institutions, and community based activities that respect to biodiversity conservation. As stated in UNESCO HE information brief, the challenge for higher education in the context of ESD is to innovate the traditional learning environment and learning processes in such a way that they do not only support learning process in the formal education, but also in informal learning.

Our environment is now facing many dilemmas starting from global financial and economic crises highlights the risks of unsustainable economic development models and practices based on short-term goals. These aspects triger economic disparity between the poor and the rich countries, many complex societal contexts, and finally environmental degradation.

Education for Sustainable Development (EfSD) promotes quality education and its inclusive for all people. It is based on values, principles, and practices necessary to respond effectively to current and future challenges. UGM has shown commitment in Education for

Sustainable Development and will continue to conduct ESD in the future. I hope that this Conference will continue to serve as a sustainable forum to provide opportunities for teachers, lecturers, researchers and professionals to share experience and present research activities and action programs. To everyone present here, I wish you have a productive and significant Conference that will benefit humankind, civilization as well as knowledge.

Lastly, I would like to extend my sincere appreciation and profound gratitude to the Director of UNESCO Jakarta and NAIST Japan for their supports. My special thanks should also go to the steering and organizing committee for their hard work in making this event a success. Thank you very much.

Yogyakarta, September 23rd, 2011

Sincerely yours,

Dr. Retno Peni Sancayaningsih, MSc.

WELCOMING SPEECH FROM EXECUTIVE DIRECTOR I-MHERE UGM

Honorable Dean of Faculty of Biology UGM, Dr. Retno Peni Sncayaningsih, M.Sc.
Distinguish Keynote speaker Prof Hubert Gijzen (Director of Unesco in Indonesia)
Distinguish Dr. Yam Tim Wing (Singapore), Prof. Yasumasa Bessho (Japan), Prof Christ Austin (Australia), Dr. Langkah Sembiring (UGM),
Dr. Yu Hao (Singapore)
Distinguish all of participants

Assalamu'alaikum wr.wb.

Welcome to Yogyakarta and participating in International Conference on Biological Science, by Faculty of Biology UGM.

This seminar was supported by IMHERE UGM (Indonesia Managing Higher Education for Relevancy and Efficiency). As we know, UGM get a competitive grant from World Bank trough Directorate General of Higher Education, from 2009 – 2012, and proposed program entitled "Education for Sustainable Development toward World Class Research University" by establishment of Center of Excellence (CoE) on 3 selected academic units, namely (i) "Tropical Biodiversity", in Faculty of Biology (ii) "Medical Herbal and Supplements" in Faculty of Pharmacy and (iii) "Reduction Emission from Deforestation and Degradation (REDD)" in Faculty of Forestry.

Faculty of Biology has attempted for enhancement of the research quality on tropical biodiversity, development of the integrated research on utilizing biodiversity resources to enhance the EfSD and development of network capacity for national and international collaboration on research and community services through Regional Centre of Expertise (RCE) Yogyakarta.

This prestigious international seminar is one of our strategic activities to achieve better key performance indicator, especially in international publication and international research collaboration. As a new paradigm of competitive grant that developed by World Bank, called "Performance Based Contracts", achievement of our key performance indicator in this year was 190% compare to targeted indicator for three years activities. We would like to continuing our "Research based Learning and Services for sustainable reputation as World Class Research University.

Please be enjoy to discuss and active participating in this seminar.

Wassalamu'alaikum wr.wb.

Sincerely yours,

Executive Director I-MHERE UGM

Dr. Cahyono Agus Dwikoranto, M.Agr.Sc.

CONFERENCE COMMITTEE
ICBS 2011 FACULTY OF BIOLOGY UGM

1.	Patron	:	Dean of Faculty of Biology
2.	Steering Committee	:	Dr. Retno Peni Sancayaningsih, M.Sc. Drs. Langkah Sembiring, M.Sc., Ph.D. Dra. Mulyati, M.Si. Dr. Endang Semiarti, M.S., M.Sc. Prof. Dra. Endang S. Soetarto, M.Sc., Ph.D. Prof. Chris Austin (Charles Darwin University, Australia) Prof. Yasumasa Bessho, Ph.D (NAIST, Japan)
3.	Academic Reviewer	:	
	Internal Reviewers	:	Prof. (ret). Dr. Jusup Subagja, M.Sc Prof. (ret). Dr. Jesmant Situmorang, M.Sc Prof. (ret). Sukarti Moeljoprawiro, M.App.Sc., PhD Prof. (ret). Dr. Nyoman Puniawati Soesilo, SU. Prof. (ret). Dr. Istiyati Prof. Dr. Endang Sutariningsih S., M.Sc Langkah Sembiring, M.Sc. Ph.D. Dr. Suwarno Hadisusanto Dr. Endang Semiarti, M.S., M.Sc. Dra. Rarastoeti Pratiwi, M.Sc., Ph.D Dr. Budi Setiadi Daryono, M.Agr.Sc. Dr. rer-nat. Ari Indrianto, SU. Dr. Niken Satuti Nur Handayani, M.Sc. Dr. Kumala Dewi, M.Sc.St. Dr. Rina Sri Kasiamdari Dr. L.Hartanto, M.Agr. Dr. Yekti Asih Purwestri, Dr. Woro Anindito, M.Sc. Dr.biol.hom. Nastiti Wijayanti
	External Reviewers	:	Dr. Sentot Santoso. (Institut fuer Klinische Immunologie und Transfusionsmedizin Justus Liebig Universität Giessen, Germany Prof. Yasumasa Bessho, Ph.D. (Graduate School of Biological Science, Nara Institute of Science and Technology (NAIST), Japan)
4.	Chief of Organizing Committee	:	Dr. Yekti Asih Purwestri, M.Si.
5.	Vice of Chief of Organizing Committee	:	Dr. L. Hartanto Nugroho, M.Agr.
6.	Secretary/Secretariate	:	Ardaning Nuriliani, S.Si., M.Kes. Dra. Ratna Susandarini, M.Sc. Widiastuti, S.Pd. Siti Nurhaida, S.E Dimas Willy, SIP Kukuh Madyaningrana
7.	Treasurer	:	Dr. Diah Rachmawati Yuni Hartati Samiyati, S.E, M.Acc Pardiso

8.	Plenary and Scientific Session	: Dr. Rina Sri Kasiamdari Dr. biol.hom. Nastiti Wijayanti, M.Si. Abdul Rahman Siregar, S.Si., M.Biotech. Dr. Woro Anindito Sri Tunjung, M.Sc Sari Darmasiwi, S.Si, M.Biotech Aries Bagus Sasongko, S.Si, M.Biotech
9.	Publication	: Zuliyati Rohmah, S.Si., M.Si. Donan Satria Yudha, S.Si., M.Sc. Slamet Riyadi, S.Si Aris Setiawan R. Nur Wigunadi
10.	Funding and Sponsorship	: Dr. Suwarno Hadisusanto Donan Satria Yudha, S.Si, M.Sc
11.	Documentation	: Drs. Abdul Rachman, M.Si. Sudarsono
12.	Logistics	: Drs. Sutikno, S.U. Drs. H. Wiyono Yatin, S.Pd Harjana Dodo Priyatno Nahrowi Giyarto Bekti Lamo
13.	Refreshment	: Dra. Siti Susanti, S.U Kodrat Wartini Rusna Nuraini Prapti
14.	Hospitality	: Dr. Niken Satuti Nur Handayani Dr. Rarastoeti Pratiwi, M.Sc Drs. Heri Sujadmiko, M.Si Dra. Upiek Ngesti Wibawaning Astuti, M.Kes Dr. Maryani
15.	Accommodation	: Slamet Widiyanto, S.Si., M.Sc. Donan Satria Yudha, S.Si., M.Sc. Haryanto Suharjito Harsono

ACKNOWLEDGMENT

The following personal and organization are gratefully acknowledged for Suporting this
International Conference on Biological Science
(ICBS 2011 BIO-UGM)
ADVANCES IN BIOLOGICAL SCIENCE:
Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity

Ms. Sachiko Iida, Japan

PT Diastika Biotekindo, Indonesia

PT Roche, Indonesia

Prima Grafika Yogyakarta, Indonesia

Drs. Agus Suryanto, Indogama - Yogyakarta, Indonesia

PLENARY SESSIONS

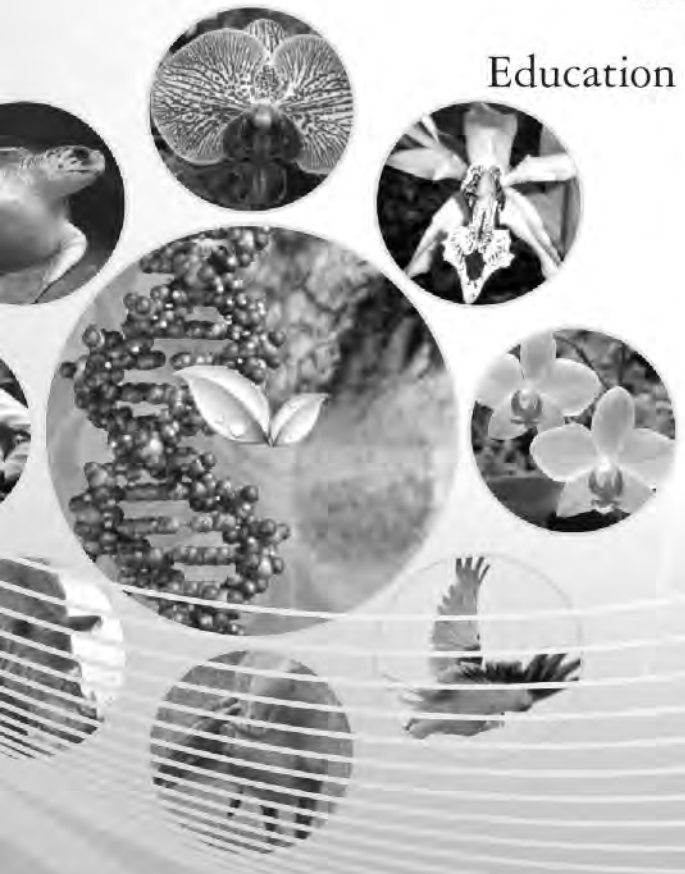
ISBN : 978-979-8969-06-5

International Conference on Biological Science
Faculty of Biology Universitas Gadjah Mada 2011
(ICBS BIO-UGM 2011)

ICBS 2011
Faculty of Biology Universitas Gadjah Mada

ADVANCES IN BIOLOGICAL SCIENCE

Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity



September 23rd-24th 2011

Yogyakarta, INDONESIA

FACULTY OF BIOLOGY
UNIVERSITAS GADJAH MADA

I-MHERE
PROJECT

O-MB12

Computer Aided Simulation of DNA Fingerprint Amplified Fragment Length Polymophism (AFLP) Using Suffix Tree Indexing and Data Mining

Kestrilia Rega P[1], Sulistyo Emantoko[1], Bhinawan Whendy[2], Agung Budiman[1]

[1]Biotechnology Faculty, University of Surabaya
[2]Departement of Informatics, Engineering Faculty , University of Surabaya
kestrilia@ubaya.ac.id, emantoko@ubaya.ac.id

Abstract

AFLP is one of the DNA Fingerprinting techniques which have broad application as genetic marker in various fields. Begin with the DNA sequence digestion using one or more particular restriction enzyme, ligation of the adapters to the overhanging sticky ends followed by DNA fragments amplification using PCR. The PCR reaction uses primers that match the adapter sequence and have some (1 to 3) additional "selective" bases which could be any bases, this reduces the number of bands that will be amplified. Such technique intended to increase the amplified fragments peculiarity so the polymorphism of the organism being studied could be well visualized by gel electrophoresis. The computer aided of AFLP simulation developed in this research was aimed to predict this electrophoresis result by simulate the digestion, ligation and PCR process using some pattern recognition algorithm applied to the DNA sequence from online databases. Through this simulation the researcher could determine the best combination of restriction enzyme and selective bases for their laboratory experiment. Suffix tree indexing was conducted during the exploration process of the genome sequence (in FASTA format) to find the restriction sites rapidly and create fragments of it. Data modeling enable the system draws the fragments into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best restriction enzyme and selective bases combination by calculating certain quantitative criteria.

Keywords : DNA Fingerprint, AFLP, PCR, Suffix Tree Indexing, Data Mining

I. INTRODUCTION

Since its first development in the mid-1980's, technique for DNA fingerprinting has rapidly evolved. In the field of agriculture, this technology assisted seed selection in order to acquire high quality plant such as cereals [1] and tea [2]. Many researcher suggested that Amplified Fragment Length Polymorphism (AFLP) is the best genetic marker nowadays in term of it's information quantity, reproducibility and resolution of genetic polymorphism. With this technique, DNA treated with restriction enzymes is amplified with PCR. It also allows selective amplification of restriction fragments, giving rise to large numbers of useful markers which can be located on the genome relatively quickly and reliably. Users can determine the specificity level of genetic marker by altering the restriction enzyme and sequence of bases in primer's selective bases. Unfortunately, due to the operation cost, it is

not an easy task to conduct trial and error attempt to find the best combination of restriction enzyme and selective bases. Therefore AFLP simulation program (in silico experiment) was developed in this research to help researchers simulate combinations of restriction enzymes and selective bases on virtual AFLP procedure by computational method so that they can determine the combinations that can be used to produce the desired genetic marker through in vitro experiment.

II. MATERIALS AND METHODS

Input of this computational method is DNA sequence from the online database. *Vitis Vinifera* genome sequence was taken from GenBank NCBI as an example and as much as 145 type II restriction enzymes were downloaded from the online Restriction Enzyme Database (rebase.neb.com). In order to make the simulation operational in the wet laboratory these 145 restriction enzymes were selected based on following criteria : (1) palindromic; (2) sticky end; (3) cut the DNA precisely on the restriction site; (4) no ambiguous and methylated bases on the restriction site; (5) at least one supplier available. Virtual restriction digestion then conducted by applying suffix tree algorithm as string pattern matching technique on the genome sequence. This algorithm will rapidly seek the string pattern which is match the restriction site of the enzymes being studied and then separate the genome sequence into subsequences. Hence, virtual PCR is done by exploring the compatibility between sub sequences and the primer-selective bases being studied. At the end of the simulation, exponential regression data modeling would enable the system draws the subsequences into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best possible restriction enzyme and selective bases combination by calculating certain quantitative criteria and conduct cluster analysis.

III. SYSTEM'S DESIGN

III.1 Input

DNA's genome sequence in FASTA format is required as system's raw material as well as the information of enzyme's restriction site pattern. The sequence could be store in several files (one file for each chromosome) in txt format. This FASTA sequence then considered as a text. Hence, all algorithm used in the consecutive processes should be string based algorithms.

III.2 Suffix Tree Algorithm

The first process is tracing the whole text (whole genome sequence) to find the short text (sub sequences) which is match the restriction sites of the restriction enzymes being

studied. The major computational problem when dealing with genome scale sequence is execution time due to computer's processor and memory performance limitation. It can take time up to one hour to find one short sequence along the whole genome [3]. Therefore, an effective string matching technique should be implemented to speed up the process. Hence, more restriction enzyme combination could be simulated. One popular technique to run fast string matching is suffix tree algorithm. Suffix tree are versatile data structures that can help execute short subsequences (queries) very efficiently. In fact, suffix trees are useful for solving a wide variety of string based problems [4]. For instance, the exact substring matching problem can be solved in time proportional to the length of the query, once the suffix tree is built on the database string. The example of suffix tree construction is shown in Figure 1 [5].

.

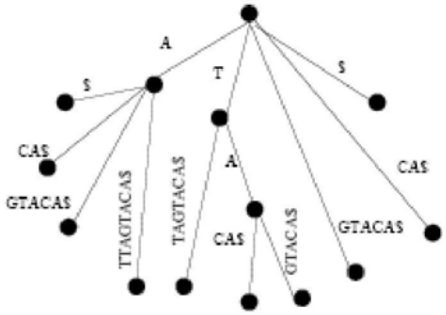


Figure 1. Suffix Tree Representation

The tree will inform every possible subsequence from a sequence as a pattern. One pattern is considered as particular path from the top node (root) to the most bottom node (leaf), for the example on the figure there are 10 possible sub sequences for the ATTAGTACA$ sequence. The $ character is added to inform the end of the sequence. There are three main function in this exploration process :

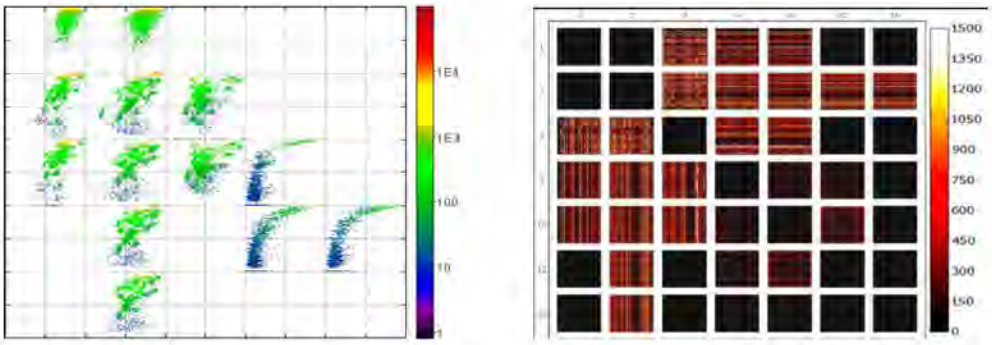
1. Build tree , construct suffix tree on the database. Every sequence (in FASTA format) subjected to the exploration should be transformed to the tree structure. Once it build, the FASTA format no longer needed so that it can be deleted and provide more space on the computer's memory.
2. Node searching, explore the tree for the queries, begin from the root (the top node) and end up at the leaf which is the most bottom node. If the query doesn't exist the system will report as "nothing". Each subsequence being found is indexed by number, represent its location on the sequence and its length (the number of the string).
3. Dispose, automatically erase the tree from the memory after it is stored on the database.

There will be 53.248 search on the *Vitis Vinifera* sequence's tree, the detail is explained in the following paragraph.

According to its restriction site, the restriction enzymes were classified into 44 groups and the simulation was conducted on 13 combinations among it. The combinations were determined as follow : (1) Three group having 4 bases of restriction site were paired with 3 group having 6 bases of restriction site, all with the most frequent match on the genome sequence; (2) The *Eco*RI and *Mse*I pair also included in the combinations although *Eco*RI do not fulfil the criteria because of there are facts that thus pair was used frequently for AFLP experiment [6,7,8,9]; (3) The restriction site of each pair do not overlap because such condition could lead bad and unpredicted restriction result. Three nucleotide selective bases were used for each subsequence's right and left hand end. Because there are 4 possible base (A,T,C,G), the total combination for selective bases should be 4^6 = 4.096. Therefore the total run for searching process on the tree is 13 x 4.096 = 53.248.

III.3 Cluster Analysis

The exploration result from the suffix tree then analyse by regarding on some criteria, which are : (1) Fragment (subsequence) length; (2) Percent of "in range" fragment, the number of fragment with the length does not exceed the polyacrylamide gel range criteria divided by the total fragment; (3) Percent of redundancy, the number of fragment with same length but different sequence divided by the total fragment. The analysis was done using multi dimension cluster analysis. The example of cluster representation is shown in Figure 2.



(a) (b)

Figure 2. Cluster Respresentation, (a) restriction enzyme and selective bases combination with their percent of "in range" fragment and percent of redundancy; (b) restriction enzyme and selective bases combination with their fragment length

III.4 The Selection Criteria

In order to find the best ten combinations of restriction enzyme and selective bases, the selection criteria should be well define. The combination will be considered good if : (1) Percent of redundancy less than 25%, too many different subsequence which have same length will reduce the polymorphism information; (2) Percent of fragment "in range" more

than 75%, too many fragment "out range" cause electrophoresis failure due to most of the fragment can not well visualize; (3) The average difference of the fragment length should be large enough so that it could be nicely separate on electrophoresis process. The selection is done by applying IF THEN rules.

III.5 Exponential Regression Model

To simulate the electrophoresis process, the system provide 1 Kbp DNA ladder from which the exponential model was developed. The exponential model between fragment size (bp) and its distance (cm) from the well is as follow:

$$\ln(\text{size}) = 10.81 - 0.736 * \text{distance}$$

IV. RESULT AND DISCUSSION

IV.1 Genome Description

The FASTA format of *Vitis Vinifera* genome sequence was separated in 19 different txt file, one file for one chromosome. Table 1 contains the description of each chromosome sequence component :

Table 1. The *Vitis Vinifera* Chromosome Sequence Description

Chromosome	Ambiguous bases per 1000 bases	GC Content (%)	Size of FASTA file (kb)
1	28	34,45	15.701
2	83	34,48	17.682
3	48	34,42	10.233
4	54	34,40	19.380
5	47	34,85	23.533
6	55	34,45	24.257
7	70	34,46	15.302
8	36	34,47	21.654
9	31	33,69	16.607
10	55	34,53	9.691
11	21	34,46	13.999
12	49	34,51	18.624
13	38	34,19	15.260
14	25	34,57	19.568
15	25	33,72	7.728
16	32	34,14	8.196
17	33	34,89	13.118
18	23	34,70	18.780
19	33	34,05	14.135
Total	**42**	**34,43**	**300.211**

IV.2 Exploration Process Performance

The main problem when facing with simulation of genome scale sequence is the operation time, but it is proven that by conducting suffix tree algorithm the operation time could be reduced significantly. Table 2 describes the time needed for suffix tree construction based on the size of the genome sequence. It is shown that the time needed increase in

linear form with the size of genome sequence, however the system could still operate in reasonable time (less than 3 minute) to handle genome sequence up to 12.1 Mbp long.. Once the suffix tree is constructed, all short pattern searching could be done in no time.

Table 2. Time for Suffix Tree Construction Based on Genome Sequence Size

Size of Genome Sequence (Mbp)	Time (second)
2,43	17
4,87	32
7,3	52
9,74	66
12,1	80

IV.2 Restriction Result Description

By conducting the data mining technique, there are several information that could be infer about the restriction result. It is known that a lot of small fragments were formed using a pair of restriction enzyme with 4 nucleotide restriction site, in the other hand just few bigger fragments were formed using a pair of restriction enzyme with 6 nucleotide restriction site. This facts were inline with the restriction digestion theory, restriction site with many nucleotide will have less probability to match the genome sequence. Therefore, the combination of restriction enzyme with 4 and 6 nucleotide of restriction site seems to be the better choice. These combinations will produce moderate number of fragments with moderate length as well.

IV.3 The Best Ten Combinations

Regarding to the selection criteria, the best ten combinations of restriction enzyme and selective bases were found. Table 3 describes thus combinations.

Table 3. The Best Ten Combinations Description

Rank	Restriction Enzyme		Selective Base		Range	Fragment in range		Redundancy	% of Restriction	% of Amplified Fragment
	1	2	1	2		Total	%			
1	AATT	ATGCAT	GCA	TAA	25-150 (126)	48	80,00%	22,92%	1,24%	0,04%
2	AATT	ATGCAT	GCA	CCA	25-150 (126)	44	80,00%	25,00%	1,24%	0,03%
3	AATT	AAGCTT	GCA	CTC	25-150 (126)	43	82,69%	23,26%	1,23%	0,03%
4	AATT	ATGCAT	GCA	TCT	25-150 (126)	42	76,36%	19,05%	1,24%	0,03%
5	AATT	ATGCAT	GCC	AAA	25-150 (126)	42	80,77%	23,81%	1,24%	0,04%
6	AATT	AAGCTT	GCA	ACA	25-150 (126)	41	77,36%	24,39%	1,23%	0,04%
7	AATT	ATGCAT	GCA	GAA	25-150 (126)	40	75,47%	20,00%	1,24%	0,03%
8	AATT	ATGCAT	GCC	TAA	25-150 (126)	38	77,55%	15,79%	1,24%	0,03%
9	AATT	ATGCAT	GCA	CGA	60-400 (341)	38	77,55%	21,05%	1,24%	0,03%
10	AATT	AAGCTT	GCA	TTT	25-150 (126)	38	77,55%	21,05%	1,23%	0,04%

Figure 3 depicts the visualization of virtual electrophoresis pattern based on the exponential regression model using 1 Kbp DNA Ladder. The blue line indicate that there is only one kind of subsequence with particular size, the green line indicate that there are two kind of subsequences with the same size, the red line indicate that there are three kind of subsequences with the same size and finally the black line indicate that there are more than three kind of subsequences with the same size. The black line should appears as the most thick and bright band in real gel electrophoresis result.

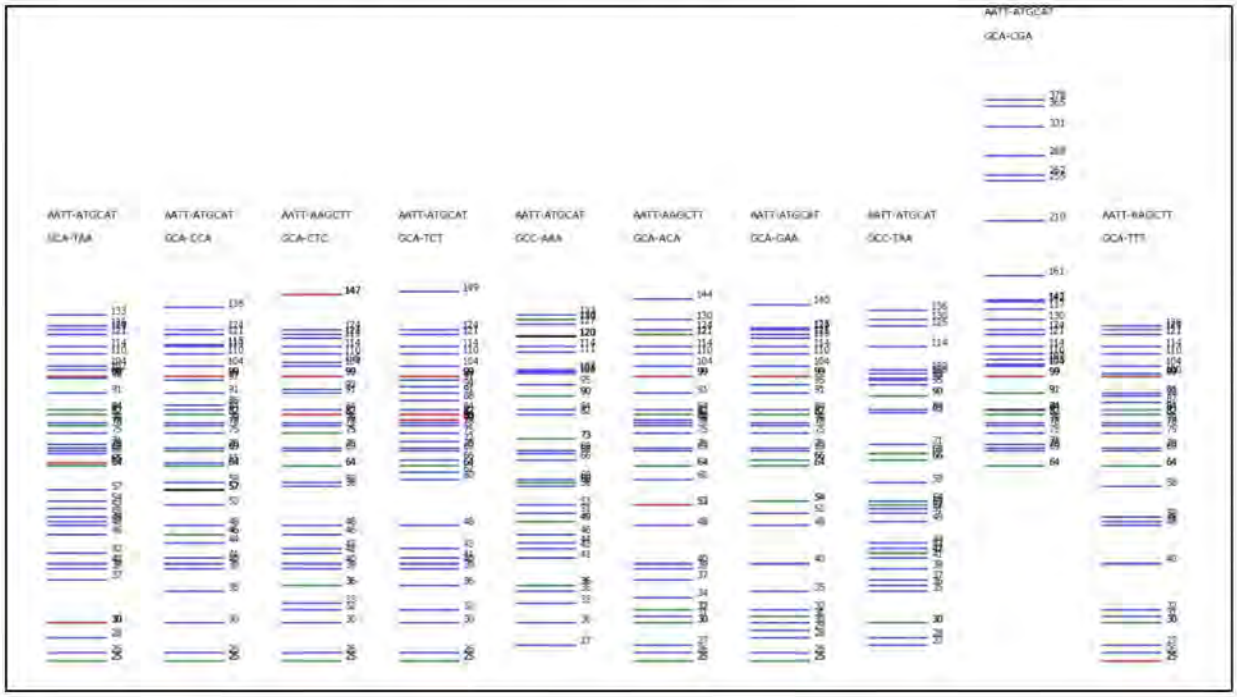


Figure 3. Visualization of The Virtual Electrophoresis Pattern

V. CONCLUSION

Like other simulation software, many factors embedded in laboratory experiment could not completely cover in this system, so that the result should be considered as recommendation (certainly with its probability of failure). However, so far the simulation result of AFLP with suffix tree indexing and data mining shows quite promising guidance for the laboratory experiment. The system developed in this research is a prototype from which more automatic and integrated system could be easily constructed. Machine learning technique such as genetic algorithm could be implemented to automate the optimization of selection criteria. At the end, laboratory conformation for this research result still could not leave behind. Therefore in the short incoming time such laboratory experiment should be conducted.

REFERENCES

[1] Korzun, V. 2003. *Molecular Markers and Their Applications in Cereals Breeding.A paper presented during the FAO international workshop on "Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding?"*. Turin, Italy.

[2] Sui, N., Yao, M., Chen, L., Zhao, L., & Wang, X. 2008. *Germplasm and Breeding Research of Tea Plant Based on DNA Marker Approaches*. Front. Agric. China 2 (2): 200-207.

[3] Budiman, A. 2010. *Pembuatan Program Simulasi AFLP In Silico (Studi Kasus pada Vitis Vinivera)*.Thesis. Surabaya.

[4] D. Gusfield. 1997. *Algorithms on Strings, Trees and Sequences:Computer Science and Computational Biology*. Cambridge University Press.

[5] Tata, Sandeep., Hankins, Richard A., Patel Jignesh M. 2004. *Practical Suffix Tree Construction*. Proceedings of the 30th VLDB Conference, Toronto, Canada.

[6] Cervera, M. T., Gusmao, J., Steenackers, M., Van Gysel, A., Van Montagu, M., & Boerjan, W. 1996. *Application of AFLPTM-Based Molecular Markers to Breeding of Populus spp*. Plant Growth Regulation 20: 47-52.

[7] Koopman, W. J. M. & Gort, G. 2004. *Significance Tests and Weighted Values for AFLP Similarities, Based on Arabidopsis in Silico AFLP Fragment Length Distributions*. Genetics 167: 1915-1928.

[8] Mueller, U. G. & Wolfenbarger, L. L. 1999. *AFLP Genotyping and Fingerprinting*. TREE 14: 389-394.

[9] Stefenon, V. M., Gailing, O., & Finkeldey, R. 2006 *Phylogenetic Relationship Within Genus Araucaria (Araucariaceae) Assessed by Means of AFLPFingerprints*. Silvae Genetica 55 (2): 45-52.

ATTACHMENT

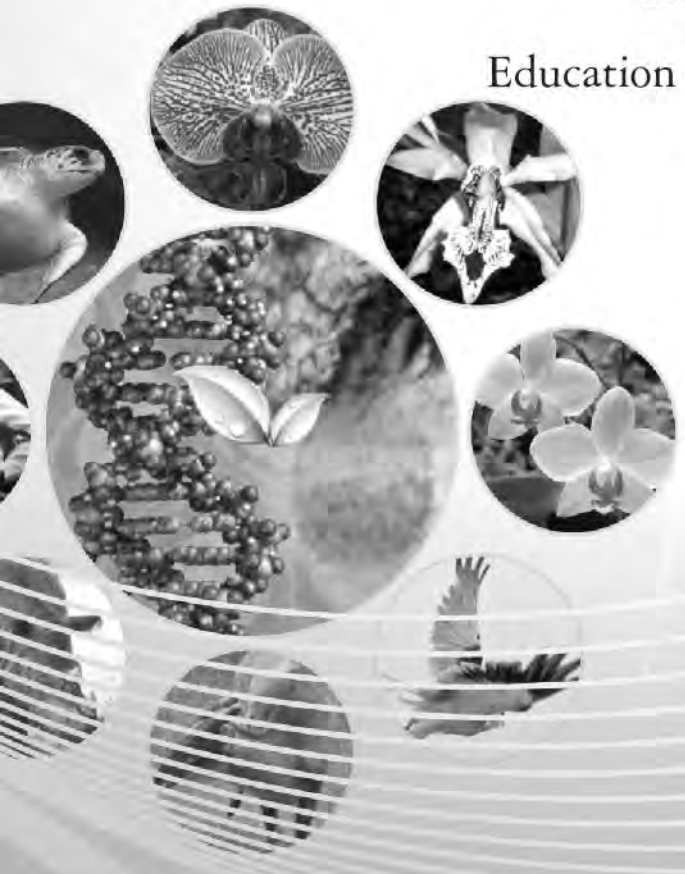
ISBN : 978-979-8969-06-5

International Conference on Biological Science
Faculty of Biology Universitas Gadjah Mada 2011
(ICBS BIO-UGM 2011)

ICBS 2011
Faculty of Biology Universitas Gadjah Mada

ADVANCES IN BIOLOGICAL SCIENCE

Education for Sustainable Development-based Tropical Biodiversity Management and Conservation for Supporting Human Prosperity



September 23rd-24th 2011

Yogyakarta, INDONESIA

FACULTY OF BIOLOGY
UNIVERSITAS GADJAH MADA

I-MHERE
PROJECT

LIST OF STUDENT COMMITTEE

NO	NAME
1.	Kukuh Madyaningrana
2.	Ika Maya Sari
3.	Maria Asih Ristianti
4.	Dyah Ayu Widyastuti
5.	Fariha Wilisiani
6.	Kartika Suryani Putri
7.	Rinatu Siswi
8.	Kurnia Widiyasari
9.	Aninda Retno Utami Wibowo
10.	Ikhya' Udin Ghozali
11.	Anida Metha Anggriasari
12.	Arien Dwitrie
13.	Inda Khoirizzaad
14.	Shinta Dewi Hayuningtias
15.	Yulaika Romadhani
16.	Rachmat Riyadi
17.	Yunita Dwi Setyorini
18.	Muhammad Bima Atmaja
19.	Mohammad Iqbal
20.	Apriliana Dyah Prawestri
21.	Arista Handayani
22.	Junita Christina Sembiring
23.	Repsi Erdiana Nindyasari
24.	Tania Tresna Asih
25.	Hendry Susila
26.	Misbakhul Bait
27.	Akbar Reza
28.	Teo Sukoco
29.	Aiffa Nurindah
30.	Lisa Novita Anggraini
31.	Laksmi Dewanti
32.	R. Aditya Haryandi
33.	Ihlas
34.	Angga Putra Kusumastianto
35.	Rusdi Hartono
36.	Nur Rohmah
37.	Yudha Rahina Putra

LIST OF ORAL PARTICIPANTS

Abdul Rahman Siregar Lab. Microbiology, Faculty of Biology, Universitas Gadjah Mada, Yogyakarta. Email: abdulrahman_siregar@yahoo.co.id
Ade Sumiardi Email: adesumiardi@yahoo.com atau fmipaunmabanten@gmail.com
Agnita Nunung Nugroho W. Faculty of Biology, UGM Email: tricahyantoh@yahoo.com
Aisa Pinaringan Sujalu Jurusan Manajemen Hutan. Fakultas Pertanian Universitas 17 Agustus 1945 Samarinda. Jl. Ir. H. Juanda 80-Samarinda 75123, Indonesia. Email: pinaringan_b@yahoo.co.id
Ali Ikhwan Biotechnology Development Centre, Universitas Muhammadiyah Malang. Jl. Raya Tlogomas no. 246, Malang. Email: ikhwan_umm@yahoo.com
Andi Agussalim Email: andiagussalim_aas@yahoo.com
Andri Frediansyah Indonesian Institute of Science (LIPI). Email: andri.frediansyah@lipi.go.id
Annisa Ridhowati Email: laksmindra.fitria@yahoo.com
Anugrah TP Juni Undergraduate Student of the Biology Departement, State University of Malang, Indonesia. Email: pramuktiajuni@rocketmail.com
Ardaning Nuriliani Faculty of Biology, Universitas Gadjah Mada. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281 Email: d_ningciit@yahoo.com
Aris Haryanto Department of Biochemistry, Faculty of Veterinary Medicine, Gadjah Mada University, Yogyakarta Email: arisharyanto@yahoo.com
Aswi A. Rofiqoh Email: aswi_ar@yahoo.com
Cahya Kurnia Fusianto Marine Study Club, Faculty of Biology Gadjah mada University. Email: cahya.fusianto@gmail.com
Cut Madinna Tiraya Email: laksmindra.fitria@yahoo.com
D. Listyorini Department of Biology, Faculty of Mathematics and Natural Sciences, State University of Malang. Jl. Semarang No. 5 Malang 65145 Indonesia. Email: toni4wan@gmail.com atau listyorini.alj@bio.um.ac.id
Dewi Syahidah Research Institute for Mariculture (RIM) Email: dewi_rimg@yahoo.com

Diah Rachmawati Laboratory of Plant Physiology, Faculty of Biology, Universitas Gadjah Mada Yogyakarta 55281-Indonesia. Email: rbudiharjo@yahoo.com
Dian Aruni Kumalawati Master students, Graduated Program, Faculty of Biology, Gadjah Mada University. Jl, Teknika Selatan, Sekip Utara, Yogyakarta 55281. Phone: +62817272085. Email: dnaruni@yahoo.com.
Dwi Anggorowati Rahayu Undergraduate Student of the Biology Departement, State University of Malang, Indonesia. Email: doewira_89@yahoo.com atau listyorini.alj@bio.um.ac.id
Dwi Listyorini Department of Biology, Faculty of Mathematics and Natural Sciences, State University of Malang. Jl. Semarang 5 Malang 65145, Indonesia. Email: listyorini.alj@bio.um.ac.id atau saraaljabarii@gmail.com
Dwi Umi Siswanti Faculty of Biology, Gadjah Mada University Email: dwi_umi@mail.ugm.ac.id
Eka Martha Della Rahayu Pusat Konservasi Tumbuhan Kebun Raya Bogor – LIPI. Jl. Ir. H. Juanda No. 13 Bogor 16003. Email: eka_mdr@yahoo.com
Endang Linirin Widiastuti Biology Dept – Faculty of Math & Sciences University of Lampung. Email: elwidi@yahoo.com
Endang Semiarti Faculty of Biology, Universitas Gadjah Mada. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281, Indonesia. Email: endsemi@ugm.ac.id ; esemiart@yahoo.com
Fadilah Nurlaili Lutfia Jurusan Biologi, FMIPA, Universitas Negeri Malang, Indonesia. Email: dhilabio@yahoo.com
Fatchiyah Email: fatchiya@gmail.com
Ferdian Andy Yus Setiantono Undergraduate Student of the Biology Department, State University of Malang, Indonesia. Email: ferdy_pe2nk@yahoo.com
Gatot Nugroho Susanto Student of Magister of Biology, Faculty of Biology, Gadjah Mada University, Indonesia. Email: angel_gatot@yahoo.com
Hastin Ambar Asti Kelompok Studi Herpetologi Faculty of Biology Universitas Gadjah Mada. Email: gnux_ok@yahoo.co.uk
Hendra Susanto Animal Physiology Laboratory, Mathematic and Science Faculty State University of Malang. Email: hendrafaal@yahoo.com atau hendrasusantofaal@gmail.com
Herawati Soekardi Email: gparamita@hotmail.com

I GP Suryadarma
Yogyakarta State University,
Faculty of Science and Mathematics, Yogyakarta, Indonesia.
Email: samodhaya@yahoo.com

I Nengah SUWASTIKA
Biological Department Faculty of Sceience Tadulako University, Palu,
Central Sulawesi 94118 Indonesia.
Email: isuwastika@yahoo.com.au

Indrawati Y. Asmara
Research Institute for Environment and Livelihood, Charles Darwin University, Australia
Ellengowen Drive, Darwin, NT 0909.
Email: indra.asmara@cdu.edu.au

Irkham Widiyono
Department of Internal Medicine, Faculty of Veterinary Medicine,
Gadjah Mada University, Yogyakarta, Indonesia.
Jl. Fauna 2 Karangmalang, Yogyakarta, Indonesia 55281
Telp.+62274560862,Fax +62274560861
Email: iwidiyono@yahoo.com atau sarminkh@yahoo.com

Juliyatin Putri Utami
Student of Biology Department, Faculty of Mathematic and Science
State University of Malang, Semarang street 5, Malang Indonesia
Email: juliyatin_putriu@yahoo.com atau hendrafaal@yahoo.com

Kestrilia Rega P.
Email: kestrilia@ubaya.ac.id atau emantoko@ubaya.ac.id

Made Pharmawati
Email: pharmawati@hotmail.com

Martina Kurnia Rohmah
Graduated of Malang State University
Email: martina.kurniarohmah@gmail.com

Miftahul Jannah
College Student of the Faculty of Natural Science,
Malang State University, Malang, Indonesia.
Email: mifta_frozi@yahoo.com

Miftahul Mushlih
The State University of Malang.
Email: mif.mushlih@gmail.com

Muhammad Rifqi Hariri
Department Biology, Faculty of Mathematic and Science,
State University of Malang.

Nesti F. Sianipar
Department of Biology, Faculty of Science and Mathematics, Universitas Pelita Harapan.
JL. M. H. Thamrin Boulevard 1100 Lippo Village, Tangerang.
Email: nestipro@yahoo.com

Nia Lukita Ariani
Email: widyaningsih.anggraeni@gmail.com

Ni Luh Watiniasih
Email: watiniasih@gmail.com

Nita Rukminasari
Department of Fisheries, Marine Science and Fisheries Faculty, Hasanuddin University,
Makassar – Indonesia.
Jl. Perintis Kemerdekaan Km. 10, Makassar - 90245, South Sulawesi – Indonesia.
Tel. : +62 411 586025; Fax: +62 411 586025
Email: nita.r@unhas.ac.id

Nur Christian Email: fatchiya@gmail.com
Penny Wurm Charles Darwin University, Darwin NT. 0909. Australia. Email: Penny.wurm@cdu.edu.au
Rahadyan Aulia Faculty of Biology ,Universitas Gadjah Mada, Yogyakarta, Indonesia. Email: aulia.rahadyan@yahoo.co.id
RCH. Soesilohadi Laboratory of Entomology, Faculty of Biology Gadjah Mada University. Email: hidayat@ugm.ac.id
Reni Ambarwati Department of Biology, Faculty of Mathematics and Natural Science, the State University of Surabaya. Email: renibio95@yahoo.co.id
Rifa Nurulfatia Faculty of Biology, Gadjah Mada University. Jl. Teknika Selatan, Yogyakarta 55281. Email: k.dewi00000@lycos.com
Rochmy Istikharah Email: rochmy@uii.ac.id
Slamet Widiyanto Animal Physiology Laboratory, Faculty of Biology, Gadjah Mada University Yogyakarta. Email: slametbiougm@yahoo.com ; slametbio@ugm.ac.id
Sulistyo Emantoko Fakultas Teknobiologi, Universitas Surabaya E-mail : emantoko@ubaya.ac.id
Takeshi Fujimuro Grad. Sch. Biol. Sci., NAIST, Japan Email: t-fujimuro@bs.naist.jp
Tjut Sugandawaty Djohan Laboratory of Ecology Faculty of Biology. Universitas Gadjah Mada, Yogyakarta 55281. Email: tdjohan95@yahoo.com
Tri Rini Nuringtyas Division Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, Einsteinweg 55, 2300 RA Leiden, The Netherlands. Email: nuningugm@gmail.com
Tuty Arisuryanti School of Environmental and Life Sciences, Charles Darwin University, Ellengowan Drive, Darwin, NT 0909, Australia. Email: tuty.arisuryanti@cdu.edu.au or tuty-arisuryanti@ugm.ac.id
Utaminingsih Faculty of Biology, Gadjah Mada University (UGM) Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281, Indonesia. Email: utaminingsih87@yahoo.com atau Issirepsumardi@yahoo.co.id
Wahyu Aristyaning P. Email: Annisah-endang@ugm.ac.id atau ardhiani.k.hidayanti@gmail.com
Woro Anindito Sri Tunig Email: wanindito@yahoo.co.jp
Yohana Th. Maria Astuti Faculty of Agriculture, Stiper Agriculture Institute. Jl. Petung no.2 , Papringan, Yogyakarta. Email: astuti_maria2000@yahoo.com

LIST OF POSTER PARTICIPANTS

Agung Budiharjo Department of Biology, Faculty of Math and Science, Sebelas Maret University. Jalan Ir Sutami 36A Kentingan Surakarta. Email: budiharjo_ag@yahoo.com
Agustin Fatimah Marine Study Club, Faculty of Biology, Universitas Gadjah Mada. Jalan Teknika Selatan, Sekip Utara, Bulaksumur Yogyakarta. E-mail : agustinfatimah@gmail.com.
Angga Putra Kusumastianto Undergraduate Student of Faculty of Biology, UGM. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281. Tel: +62-274-580839, 6492350 Fax: +62-274-580839. Email: kusumastianto@gmail.com
Anita Pratimi
Ardhiani.K.Hidayanti Faculty of Biology, Gadjah Mada University. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281, Indonesia. Email: ardhiani.k.hidayanti@gmail.com
Ariana Faculty of Veterinary Medicine Gadjah Mada University. Jl. Fauna No. 2, Yogyakarta, Indonesia. Email: arianasutejo@yahoo.com
Arif Habibal Umam Faculty of Biology, Universitas Gadjah Mada. Jl. Teknika Selatan, Sekip Utara, 55281, Yogyakarta, Indonesia. Email : arifhabibalumam@yahoo.com
Arnis Damayanti Molekular Genetics Laboratorium of Puslitbang Perum Perhutani
Chatarina Wariyah Faculty of Agroindustry. Mercu Buana University of Yogyakarta. Jl. Wates Km 10 Yogyakarta 55753. Email: chatarina_wariyah@yahoo.co.id
Chomsun Hadi Kurniawan Biology Faculty- Gadjah Mada University
Dewi Kusumawati Faculty of Biology, Gadjah Mada University, Yogyakarta. Email: little_dhe16@yahoo.com
Dian Ayuning Tyas Research Center for Biotechnology, Gadjah Mada University, Yogyakarta 55281, Indonesia. Email: dn.tyas@gmail.com
Dini Astika Sari Faculty of Biology, Universitas Gadjah Mada, Yogyakarta. Email : mohammadiqbl@gmail.com
Dodin Koswanudin Biology Post-graduate Programme, Faculty of Mathematic and Natural Science, University of Indonesia. Mobile: 08129918579. Email: dodin.koswanudin@yahoo.com
Dwi Murti Puspitaningtyas Center for Plant Conservation-Bogor Botanical Gardens-Indonesian Institute of Sciences (LIPI). Email: puspitakrb@yahoo.com

Dwi Umi Siswanti Faculty of Biology, Gadjah Mada University. Email : dwi_umi@mail.ugm.ac.id
Elizabeth B. E. Kristiani Satya Wacana Christian University. Jl. Diponegoro 42-60 Salatiga. Email: betty_kristiani@yahoo.com
Farid Kuswantoro Undergradute Student of Faculty of Biology, UGM. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281. Tel: +62-274-580839, 6492350. Fax: +62-274-580839 . Email: farid_kus@yahoo.co.id
Helena T. Tuririday Jurusan Biologi Universitas Negeri Papua- Manokwari. Email: vona2003@yahoo.com
Imam Bagus N. Faculty of Biology, Gadjah Mada University. Jalan Teknika Selatan, Sekip Utara, Bulaksumur Yogyakarta. Email: ibnugroho_biogama@yahoo.com
Krisnawati Dept of Physics, Gadjah Mada University Yogyakarta 55281 Indonesia. Email: chryso_02@yahoo.co.id or gbsuparta@ugm.ac.id
Lianah Kuswanto Doctorate Program of Environmental Studies, Diponegoro University, Semarang. Email: lianahkuswanto@yahoo.co.id
Lina Winarti Faculty of Pharmacy Jember University. Email: lhinna_w@yahoo.com
Lisna Hidayati Biotechnology, Postgraduate Programe, GMU. Email: lizna_21@yahoo.com
Luthfi Nurhidayat Posgraduate Program , Faculty of Biology Universitas Gadjah Mada, Yogyakarta. Email: Luthfiturtle@gmail.com
Made Ria Defiani Biology Department, Math and Natural Sciences, Udayana University, Indonesia. Email: defiani_ria@yahoo.com
Marlina Ummas Genisa Genetic Laboratory, Faculty of Biology, Gadjah Mada University. Email: linagenisa@yahoo.com
Mifta Pratiwi Rachman Research Center for Biotechnology Gadjah Mada University, Yogyakarta, Indonesia. Email: mifta_rachman@yahoo.com
Murni Saptasari Department of Biology, State University of Malang. Jl. Surabaya no. 6 Malang. Email: saptasarimurni@yahoo.com
Nastiti Wijayanti Faculty of Biology, Universitas Gadjah Mada. Email: nastitiw@yahoo.com

Nunung Sulistyani AAK Manggala Yogyakarta. Email: fi2_yk@yahoo.com
Nur Khikmah AAK Manggala Yogyakarta. Email: nurkhikmah@yahoo.com
Nur Qomariyah Department of Biology, Faculty of Mathematics and Science, Universitas Negeri Surabaya. Email: nurq.biounesa@yahoo.com
Repsi Erdiana N. Lab. Microbiology, Faculty of Biology, Gadjah Mada Univesity, Indonesia. Jl. Teknika Selatan, Sekip Utara, Yogyakarta 55281. Tel: +62-274-580839, 6492350 Fax: +62-274-580839. Email: repsi.erdiana@gmail.com
Ria Cahyaningsih Center for Plant Conservation, Bogor Botanical Garden. Email: ria.cahya@gmail.com
Riana Murti Handayani Departemen Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor. Email : gayuhrahayu@yahoo.com
Risang Pandegan Faculty of Biology UGM. Email: risangpandegan@yahoo.com
Rudi Nirwantono Kelompok Studi Kelautan, Faculty of Biology, Universitas Gadjah Mada. Jalan Teknika Selatan, Sekip Utara, Bulaksumur Yogyakarta. E-mail : rudi.nirwantono@gmail.com
Rungkat J. Jurusan Biologi FMIPA Universitas Negeri Manado (UNIMA). Jln.UNIMA di Tondano 95618. Email: jovialine.unima@yahoo.co.id
Senlie Octaviana Biotechnology, Gadjah Mada University. Jl. Teknika Utara Barek Yogyakarta 55281, Indonesia
Siti Mukhlishoh Setyawati IAIN Walisongo, Semarang
Seodjono Purwodadi Botanic Garden-Indonesian Institute of Sciences. Jl. Surabaya-Malang Km.65, Purwodadi, Pasuruan. Email: soejono@lipi.go.id
Sri Rahayu Bogor Botanical Gardens, Indonesian Institute of Sciences (LIPI). Jl Ir. H. Juanda 13 Bogor. Telp/fax 0251-8322187. Email: srirahayukrb@yahoo.com
Sri Rahayu Bogor Botanical Gardens, Indonesian Institute of Sciences (LIPI). Jl Ir H. Juanda 13 Bogor, Indonesia. Telp/Fax. 0251-8322187. Email: srirahayukrb@yahoo.com
Sudaryanto Department of Biology Faculty of Mathematic and Natural Sciences Udayana University. Email: Sudaryanto2000@yahoo.com

Sumarmi
Faculty of Agriculture, Slamet Riyadi Surakarta University.
Email: sumarmi_mp@yahoo.com

Tety Hartatik
Laboratory of Animal Breeding, Faculty of Animal Science, UGM, Indonesia.
Email: tety.hartatik@gmail.com

Titi Lasmini
Universitas Gadjah Mada.
Email: Grie_sala@yahoo.co.id

V. Irene Meitiniarti
Biology Faculty of Satya Wacana Christian University.
Jl. Diponegoro 52-60 Salatiga.
Telp: 0816 66 3564. Email: irene_meiti@yahoo.com.

Winastuti Dwi Atmanto

Yekti Asih Purwestri
Faculty of Biology, Universitas Gadjah Mada.
Jl. Teknika Selatan, Sekip Utara, Yogyakarta 5 5281, Indonesia.
Email: yektiugm@yahoo.com

Zuliyati Rohmah
Posgraduate Program, Faculty of Biology Universitas Gadjah Mada, Yogyakarta.
Jl. Teknika Selatan, Skip Utara Yogyakarta.
Email: zuliyati.rohmah@ugm.ac.id or luthfiturtle@gmail.com

computer_aided_simulation_of_dna_fingerprint.pdf

by Sulistyo Rega

Submission date: 01-Feb-2021 04:39PM (UTC+0700)
Submission ID: 1498957337
File name: computer_aided_simulation_of_dna_fingerprint.pdf (522.47K)
Word count: 2930
Character count: 15540

O-MB12

Computer Aided Simulation of DNA Fingerprint Amplified Fragment Length Polymophism (AFLP) Using Suffix Tree Indexing and Data Mining

Kestrilia Rega P[1], Sulistyo Emantoko[1], Bhinawan Whendy[2], Agung Budiman[1]

[1]Biotechnology Faculty, University of Surabaya
[2]Departement of Informatics, Engineering Faculty , University of Surabaya
kestrilia@ubaya.ac.id, emantoko@ubaya.ac.id

Abstract

AFLP is one of the DNA Fingerprinting techniques which have broad application as genetic marker in various fields. Begin with the DNA sequence digestion using one or more particular restriction enzyme, ligation of the adapters to the overhanging sticky ends followed by DNA fragments amplification using PCR. The PCR reaction uses primers that match the adapter sequence and have some (1 to 3) additional "selective" bases which could be any bases, this reduces the number of bands that will be amplified. Such technique intended to increase the amplified fragments peculiarity so the polymorphism of the organism being studied could be well visualized by gel electrophoresis. The computer aided of AFLP simulation developed in this research was aimed to predict this electrophoresis result by simulate the digestion, ligation and PCR process using some pattern recognition algorithm applied to the DNA sequence from online databases. Through this simulation the researcher could determine the best combination of restriction enzyme and selective bases for their laboratory experiment. Suffix tree indexing was conducted during the exploration process of the genome sequence (in FASTA format) to find the restriction sites rapidly and create fragments of it. Data modeling enable the system draws the fragments into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best restriction enzyme and selective bases combination by calculating certain quantitative criteria.

Keywords : DNA Fingerprint, AFLP, PCR, Suffix Tree Indexing, Data Mining

I. INTRODUCTION

Since its first development in the mid-1980's, technique for DNA fingerprinting has rapidly evolved. In the field of agriculture, this technology assisted seed selection in order to acquire high quality plant such as cereals [1] and tea [2]. Many researcher suggested that Amplified Fragment Length Polymorphism (AFLP) is the best genetic marker nowadays in term of it's information quantity, reproducibility and resolution of genetic polymorphism. With this technique, DNA treated with restriction enzymes is amplified with PCR. It also allows selective amplification of restriction fragments, giving rise to large numbers of useful markers which can be located on the genome relatively quickly and reliably. Users can determine the specificity level of genetic marker by altering the restriction enzyme and sequence of bases in primer's selective bases. Unfortunately, due to the operation cost, it is

not an easy task to conduct trial and error attempt to find the best combination of restriction enzyme and selective bases. Therefore AFLP simulation program (in silico experiment) was developed in this research to help researchers simulate combinations of restriction enzymes and selective bases on virtual AFLP procedure by computational method so that they can determine the combinations that can be used to produce the desired genetic marker through in vitro experiment.

II. MATERIALS AND METHODS

Input of this computational method is DNA sequence from the online database. *Vitis Vinifera* genome sequence was taken from GenBank NCBI as an example and as much as 145 type II restriction enzymes were downloaded from the online Restriction Enzyme Database (rebase.neb.com). In order to make the simulation operational in the wet laboratory these 145 restriction enzymes were selected based on following criteria : (1) palindromic; (2) sticky end; (3) cut the DNA precisely on the restriction site; (4) no ambiguous and methylated bases on the restriction site; (5) at least one supplier available. Virtual restriction digestion then conducted by applying suffix tree algorithm as string pattern matching technique on the genome sequence. This algorithm will rapidly seek the string pattern which is match the restriction site of the enzymes being studied and then separate the genome sequence into subsequences. Hence, virtual PCR is done by exploring the compatibility between sub sequences and the primer-selective bases being studied. At the end of the simulation, exponential regression data modeling would enable the system draws the subsequences into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best possible restriction enzyme and selective bases combination by calculating certain quantitative criteria and conduct cluster analysis.

III. SYSTEM'S DESIGN

III.1 Input

DNA's genome sequence in FASTA format is required as system's raw material as well as the information of enzyme's restriction site pattern. The sequence could be store in several files (one file for each chromosome) in txt format. This FASTA sequence then considered as a text. Hence, all algorithm used in the consecutive processes should be string based algorithms.

III.2 Suffix Tree Algorithm

The first process is tracing the whole text (whole genome sequence) to find the short text (sub sequences) which is match the restriction sites of the restriction enzymes being

studied. The major computational problem when dealing with genome scale sequence is execution time due to computer's processor and memory performance limitation. It can take time up to one hour to find one short sequence along the whole genome [3]. Therefore, an effective string matching technique should be implemented to speed up the process. Hence, more restriction enzyme combination could be simulated. One popular technique to run fast string matching is suffix tree algorithm. Suffix tree are versatile data structures that can help execute short subsequences (queries) very efficiently. In fact, suffix trees are useful for solving a wide variety of string based problems [4]. For instance, the exact substring matching problem can be solved in time proportional to the length of the query, once the suffix tree is built on the database string. The example of suffix tree construction is shown in Figure 1 [5].

.

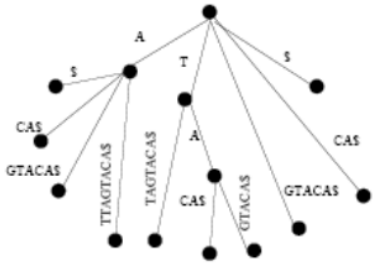


Figure 1. Suffix Tree Representation

The tree will inform every possible subsequence from a sequence as a pattern. One pattern is considered as particular path from the top node (root) to the most bottom node (leaf), for the example on the figure there are 10 possible sub sequences for the ATTAGTACA$ sequence. The $ character is added to inform the end of the sequence. There are three main function in this exploration process :

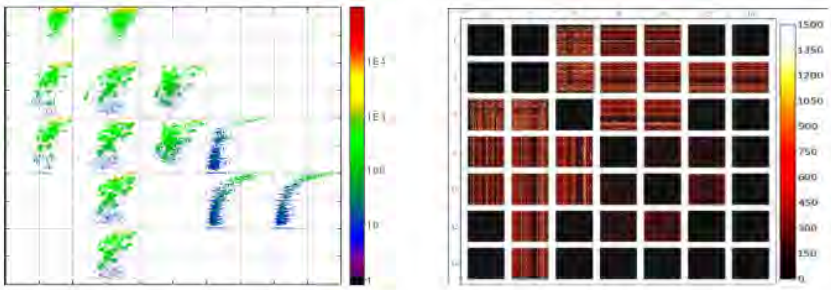
1. Build tree , construct suffix tree on the database. Every sequence (in FASTA format) subjected to the exploration should be transformed to the tree structure. Once it build, the FASTA format no longer needed so that it can be deleted and provide more space on the computer's memory.
2. Node searching, explore the tree for the queries, begin from the root (the top node) and end up at the leaf which is the most bottom node. If the query doesn't exist the system will report as "nothing". Each subsequence being found is indexed by number, represent its location on the sequence and its length (the number of the string).
3. Dispose, automatically erase the tree from the memory after it is stored on the database.

There will be 53.248 search on the *Vitis Vinifera* sequence's tree, the detail is explained in the following paragraph.

According to its restriction site, the restriction enzymes were classified into 44 groups and the simulation was conducted on 13 combinations among it. The combinations were determined as follow : (1) Three group having 4 bases of restriction site were paired with 3 group having 6 bases of restriction site, all with the most frequent match on the genome sequence; (2) The *Eco*RI and *Mse*I pair also included in the combinations although *Eco*RI do not fulfil the criteria because of there are facts that thus pair was used frequently for AFLP experiment [6,7,8,9]; (3) The restriction site of each pair do not overlap because such condition could lead bad and unpredicted restriction result. Three nucleotide selective bases were used for each subsequence's right and left hand end. Because there are 4 possible base (A,T,C,G), the total combination for selective bases should be 4^6 = 4.096. Therefore the total run for searching process on the tree is 13 x 4.096 = 53.248.

III.3 Cluster Analysis

The exploration result from the suffix tree then analyse by regarding on some criteria, which are : (1) Fragment (subsequence) length; (2) Percent of "in range" fragment, the number of fragment with the length does not exceed the polyacrylamide gel range criteria divided by the total fragment; (3) Percent of redundancy, the number of fragment with same length but different sequence divided by the total fragment. The analysis was done using multi dimension cluster analysis. The example of cluster representation is shown in Figure 2.



(a) (b)

Figure 2. Cluster Respresentation, (a) restriction enzyme and selective bases combination with their percent of "in range" fragment and percent of redundancy; (b) restriction enzyme and selective bases combination with their fragment length

III.4 The Selection Criteria

In order to find the best ten combinations of restriction enzyme and selective bases, the selection criteria should be well define. The combination will be considered good if : (1) Percent of redundancy less than 25%, too many different subsequence which have same length will reduce the polymorphism information; (2) Percent of fragment "in range" more

than 75%, too many fragment “out range” cause electrophoresis failure due to most of the fragment can not well visualize; (3) The average difference of the fragment length should be large enough so that it could be nicely separate on electrophoresis process. The selection is done by applying IF THEN rules.

III.5 Exponential Regression Model

To simulate the electrophoresis process, the system provide 1 Kbp DNA ladder from which the exponential model was developed. The exponential model between fragment size (bp) and its distance (cm) from the well is as follow:

$$\ln(\text{size}) = 10.81 - 0.736 * \text{distance}$$

IV. RESULT AND DISCUSSION

IV.1 Genome Description

The FASTA format of *Vitis Vinifera* genome sequence was separated in 19 different txt file, one file for one chromosome. Table 1 contains the description of each chromosome sequence component :

Table 1. The *Vitis Vinifera* Chromosome Sequence Description

Chromosome	Ambiguous bases per 1000 bases	GC Content (%)	Size of FASTA file (kb)
1	28	34,45	15.701
2	83	34,48	17.682
3	48	34,42	10.233
4	54	34,40	19.380
5	47	34,85	23.533
6	55	34,45	24.257
7	70	34,46	15.302
8	36	34,47	21.654
9	31	33,69	16.607
10	55	34,53	9.691
11	21	34,46	13.999
12	49	34,51	18.624
13	38	34,19	15.260
14	25	34,57	19.568
15	25	33,72	7.728
16	32	34,14	8.196
17	33	34,89	13.118
18	23	34,70	18.780
19	33	34,05	14.135
Total	**42**	**34,43**	**300.211**

IV.2 Exploration Process Performance

The main problem when facing with simulation of genome scale sequence is the operation time, but it is proven that by conducting suffix tree algorithm the operation time could be reduced significantly. Table 2 describes the time needed for suffix tree construction based on the size of the genome sequence. It is shown that the time needed increase in

linear form with the size of genome sequence, however the system could still operate in reasonable time (less than 3 minute) to handle genome sequence up to 12.1 Mbp long.. Once the suffix tree is constructed, all short pattern searching could be done in no time.

Table 2. Time for Suffix Tree Construction Based on Genome Sequence Size

Size of Genome Sequence (Mbp)	Time (second)
2,43	17
4,87	32
7,3	52
9,74	66
12,1	80

IV.2 Restriction Result Description

By conducting the data mining technique, there are several information that could be infer about the restriction result. It is known that a lot of small fragments were formed using a pair of restriction enzyme with 4 nucleotide restriction site, in the other hand just few bigger fragments were formed using a pair of restriction enzyme with 6 nucleotide restriction site. This facts were inline with the restriction digestion theory, restriction site with many nucleotide will have less probability to match the genome sequence. Therefore, the combination of restriction enzyme with 4 and 6 nucleotide of restriction site seems to be the better choice. These combinations will produce moderate number of fragments with moderate length as well.

IV.3 The Best Ten Combinations

Regarding to the selection criteria, the best ten combinations of restriction enzyme and selective bases were found. Table 3 describes thus combinations.

Table 3. The Best Ten Combinations Description

Rank	Restriction Enzyme		Selective Base		Range	Fragment in range		Redundancy	% of Restriction	% of Amplified Fragment
	1	2	1	2		Total	%			
1	AATT	ATGCAT	GCA	TAA	25-150 (126)	48	80,00%	22,92%	1,24%	0,04%
2	AATT	ATGCAT	GCA	CCA	25-150 (126)	44	80,00%	25,00%	1,24%	0,03%
3	AATT	AAGCTT	GCA	CTC	25-150 (126)	43	82,69%	23,26%	1,23%	0,03%
4	AATT	ATGCAT	GCA	TCT	25-150 (126)	42	76,36%	19,05%	1,24%	0,03%
5	AATT	ATGCAT	GCC	AAA	25-150 (126)	42	80,77%	23,81%	1,24%	0,04%
6	AATT	AAGCTT	GCA	ACA	25-150 (126)	41	77,36%	24,39%	1,23%	0,04%
7	AATT	ATGCAT	GCA	GAA	25-150 (126)	40	75,47%	20,00%	1,24%	0,03%
8	AATT	ATGCAT	GCC	TAA	25-150 (126)	38	77,55%	15,79%	1,24%	0,03%
9	AATT	ATGCAT	GCA	CGA	60-400 (341)	38	77,55%	21,05%	1,24%	0,03%
10	AATT	AAGCTT	GCA	TTT	25-150 (126)	38	77,55%	21,05%	1,23%	0,04%

Figure 3 depicts the visualization of virtual electrophoresis pattern based on the exponential regression model using 1 Kbp DNA Ladder. The blue line indicate that there is only one kind of subsequence with particular size, the green line indicate that there are two kind of subsequences with the same size, the red line indicate that there are three kind of subsequences with the same size and finally the black line indicate that there are more than three kind of subsequences with the same size. The black line should appears as the most thick and bright band in real gel electrophoresis result.

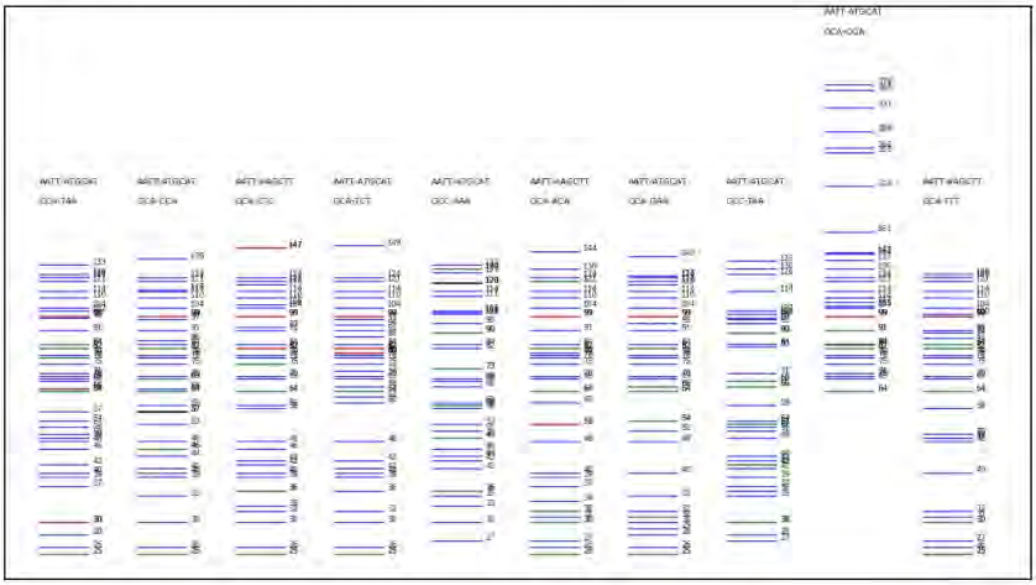


Figure 3. Visualization of The Virtual Electrophoresis Pattern

V. CONCLUSION

Like other simulation software, many factors embedded in laboratory experiment could not completely cover in this system, so that the result should be considered as recommendation (certainly with its probability of failure). However, so far the simulation result of AFLP with suffix tree indexing and data mining shows quite promising guidance for the laboratory experiment. The system developed in this research is a prototype from which more automatic and integrated system could be easily constructed. Machine learning technique such as genetic algorithm could be implemented to automate the optimization of selection criteria. At the end, laboratory conformation for this research result still could not leave behind. Therefore in the short incoming time such laboratory experiment should be conducted.

REFERENCES

[1] Korzun, V. 2003. *Molecular Markers and Their Applications in Cereals Breeding.A paper presented during the FAO international workshop on "Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding?"*. Turin, Italy.

[2] Sui, N., Yao, M., Chen, L., Zhao, L., & Wang, X. 2008. *Germplasm and Breeding Research of Tea Plant Based on DNA Marker Approaches*. Front. Agric. China 2 (2): 200-207.

[3] Budiman, A. 2010. *Pembuatan Program Simulasi AFLP In Silico (Studi Kasus pada Vitis Vinivera)*.Thesis. Surabaya.

[4] D. Gusfield. 1997. *Algorithms on Strings, Trees and Sequences:Computer Science and Computational Biology*. Cambridge University Press.

[5] Tata, Sandeep., Hankins, Richard A., Patel Jignesh M. 2004. *Practical Suffix Tree Construction*. Proceedings of the 30th VLDB Conference, Toronto, Canada.

[6] Cervera, M. T., Gusmao, J., Steenackers, M., Van Gysel, A., Van Montagu, M., & Boerjan, W. 1996. *Application of AFLPTM-Based Molecular Markers to Breeding of Populus spp*. Plant Growth Regulation 20: 47-52.

[7] Koopman, W. J. M. & Gort, G. 2004. *Significance Tests and Weighted Values for AFLP Similarities, Based on Arabidopsis in Silico AFLP Fragment Length Distributions*. Genetics 167: 1915-1928.

[8] Mueller, U. G. & Wolfenbarger, L. L. 1999. *AFLP Genotyping and Fingerprinting*. TREE 14: 389-394.

[9] Stefenon, V. M., Gailing, O., & Finkeldey, R. 2006 *Phylogenetic Relationship Within Genus Araucaria (Araucariaceae) Assessed by Means of AFLPFingerprints*. Silvae Genetica 55 (2): 45-52.

computer_aided_simulation_of_dna_fingerprint.pdf

ORIGINALITY REPORT

11%	11%	6%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.unusa.ac.id Internet Source	7%
2	docplayer.net Internet Source	2%
3	hdl.handle.net Internet Source	1%
4	genome.cshlp.org Internet Source	<1%
5	www.tandfonline.com Internet Source	<1%
6	"Data Production and Analysis in Population Genomics", Springer Science and Business Media LLC, 2012 Publication	<1%

Exclude quotes	Off	Exclude matches	< 4 words
Exclude bibliography	On		

Fakultas Biologi UGM adakan International Conference on Biological Science (ICBS) 2011

Rilis Berita 23 September 2011, 11.01 Oleh: dedieko 0

Post Views: 827

Berdasarkan data dari Sekertaris Panitia, ibu Ardaning Nuriliani, S.Si., M.Kes., konferensi ini dihadiri lebih dari 400 orang, berasal dari berbagai negara antara lain Indonesia, Jepang, Australia, Belanda dan Afrika. Sebanyak 82 makalah akan disampaikan dalam bentuk presentasi oral dan 59 makalah disampaikan dalam bentuk poster. Semua makalah tersebut meliputi 5 kelompok topik yaitu 1) *Molecular Biology, Genetic and Bio-informatics*; 2) *Ecology and Conservation*; 3) *Systematic and Evolution*; 4) *Physiology and Developmental Biology*; dan 5) *Bio-medics*, imbuh beliau.

Pada konferensi yang dibuka secara resmi oleh Dekan Fak. Biologi UGM (Dr. Retno Peni Sancayaningsih, M.Sc.), Prof. Hubert Gijzen (*Director of UNESCO in Indonesia*) didaulat untuk menjadi pembicara kunci pada acara ini. Beliau menyampaikan bahwa kemajuan teknologi hendaknya sejalan dengan tindakan konservasi alam. Acara ini dihadiri lima pakar sebagai pembicara tamu, yaitu: Dr. Yam Tim Wing (Singapore Botanic Garden, Singapore) yang akan membawakan makalah tentang *Physiology and Developmental Biology,* Prof. Yasumasa Bessho (Graduate School of Biological Sciences, Nara Institute of Science and Technology, Japan) dengan makalah bertema *Biomedics*, Prof. Chris Austin (School of Environmental and Life Sciences, Charles Darwin University, Australia) untuk bidang *Ecology and Conservation*, Drs. Langkah Sembiring, M.Sc., Ph.D (Faculty of Biology, Universitas Gadjah Mada, Indonesia) dengan makalah yang bertema *Systematic and Evolution,* dan Dr. Yu Hao (Department of Biological Science, National University of Singapore, Singapore) akan membawakan makalah untuk bidang *Molecular Biology, Genetic, and Bioinformatics.*

Sebagai rangkaian acara ICBS 2011, sehari sebelumnya (22/09) telah diadakan *Stadium General* tentang *Animal Cell Culture* oleh Prof. Yasumasa Bessho. Acara yang dilaksanakan di Ruang Sidang Pascasarjana Fakultas Biologi ini kemudian dilanjutkan dengan Praktek laboratorium di Fasilitas penelitian Bersama (FALITMA) Fak. Biologi UGM. (**ZR**)

Recent Posts

Leave A Comment

Alamat email Anda tidak akan dipublikasikan. Ruas yang wajib ditandai *

Comment *

Nama * Email *

Simpan nama, email, dan situs web saya pada peramban ini untuk komentar saya berikutnya.

KIRIM KOMENTAR

O-MB12

Computer Aided Simulation of DNA Fingerprint Amplified Fragment Length Polymophism (AFLP) Using Suffix Tree Indexing and Data Mining

Kestrilia Rega P[1], Sulistyo Emantoko[1], Bhinawan Whendy[2], Agung Budiman[1]

[1]Biotechnology Faculty, University of Surabaya
[2]Departement of Informatics, Engineering Faculty , University of Surabaya
kestrilia@ubaya.ac.id, emantoko@ubaya.ac.id

Abstract

AFLP is one of the DNA Fingerprinting techniques which have broad application as genetic marker in various fields. Begin with the DNA sequence digestion using one or more particular restriction enzyme, ligation of the adapters to the overhanging sticky ends followed by DNA fragments amplification using PCR. The PCR reaction uses primers that match the adapter sequence and have some (1 to 3) additional "selective" bases which could be any bases, this reduces the number of bands that will be amplified. Such technique intended to increase the amplified fragments peculiarity so the polymorphism of the organism being studied could be well visualized by gel electrophoresis. The computer aided of AFLP simulation developed in this research was aimed to predict this electrophoresis result by simulate the digestion, ligation and PCR process using some pattern recognition algorithm applied to the DNA sequence from online databases. Through this simulation the researcher could determine the best combination of restriction enzyme and selective bases for their laboratory experiment. Suffix tree indexing was conducted during the exploration process of the genome sequence (in FASTA format) to find the restriction sites rapidly and create fragments of it. Data modeling enable the system draws the fragments into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best restriction enzyme and selective bases combination by calculating certain quantitative criteria.

Keywords : DNA Fingerprint, AFLP, PCR, Suffix Tree Indexing, Data Mining

I. INTRODUCTION

Since its first development in the mid-1980's, technique for DNA fingerprinting has rapidly evolved. In the field of agriculture, this technology assisted seed selection in order to acquire high quality plant such as cereals [1] and tea [2]. Many researcher suggested that Amplified Fragment Length Polymorphism (AFLP) is the best genetic marker nowadays in term of it's information quantity, reproducibility and resolution of genetic polymorphism. With this technique, DNA treated with restriction enzymes is amplified with PCR. It also allows selective amplification of restriction fragments, giving rise to large numbers of useful markers which can be located on the genome relatively quickly and reliably. Users can determine the specificity level of genetic marker by altering the restriction enzyme and sequence of bases in primer's selective bases. Unfortunately, due to the operation cost, it is

not an easy task to conduct trial and error attempt to find the best combination of restriction enzyme and selective bases. Therefore AFLP simulation program (in silico experiment) was developed in this research to help researchers simulate combinations of restriction enzymes and selective bases on virtual AFLP procedure by computational method so that they can determine the combinations that can be used to produce the desired genetic marker through in vitro experiment.

II. MATERIALS AND METHODS

Input of this computational method is DNA sequence from the online database. *Vitis Vinifera* genome sequence was taken from GenBank NCBI as an example and as much as 145 type II restriction enzymes were downloaded from the online Restriction Enzyme Database (rebase.neb.com). In order to make the simulation operational in the wet laboratory these 145 restriction enzymes were selected based on following criteria : (1) palindromic; (2) sticky end; (3) cut the DNA precisely on the restriction site; (4) no ambiguous and methylated bases on the restriction site; (5) at least one supplier available. Virtual restriction digestion then conducted by applying suffix tree algorithm as string pattern matching technique on the genome sequence. This algorithm will rapidly seek the string pattern which is match the restriction site of the enzymes being studied and then separate the genome sequence into subsequences. Hence, virtual PCR is done by exploring the compatibility between sub sequences and the primer-selective bases being studied. At the end of the simulation, exponential regression data modeling would enable the system draws the subsequences into virtual DNA's electrophoresis pattern. Data mining accomplish the simulation by exploring overall possible virtual DNA's electrophoresis pattern and determine the best possible restriction enzyme and selective bases combination by calculating certain quantitative criteria and conduct cluster analysis.

III. SYSTEM'S DESIGN

III.1 Input

DNA's genome sequence in FASTA format is required as system's raw material as well as the information of enzyme's restriction site pattern. The sequence could be store in several files (one file for each chromosome) in txt format. This FASTA sequence then considered as a text. Hence, all algorithm used in the consecutive processes should be string based algorithms.

III.2 Suffix Tree Algorithm

The first process is tracing the whole text (whole genome sequence) to find the short text (sub sequences) which is match the restriction sites of the restriction enzymes being

studied. The major computational problem when dealing with genome scale sequence is execution time due to computer's processor and memory performance limitation. It can take time up to one hour to find one short sequence along the whole genome [3]. Therefore, an effective string matching technique should be implemented to speed up the process. Hence, more restriction enzyme combination could be simulated. One popular technique to run fast string matching is suffix tree algorithm. Suffix tree are versatile data structures that can help execute short subsequences (queries) very efficiently. In fact, suffix trees are useful for solving a wide variety of string based problems [4]. For instance, the exact substring matching problem can be solved in time proportional to the length of the query, once the suffix tree is built on the database string. The example of suffix tree construction is shown in Figure 1 [5].

.

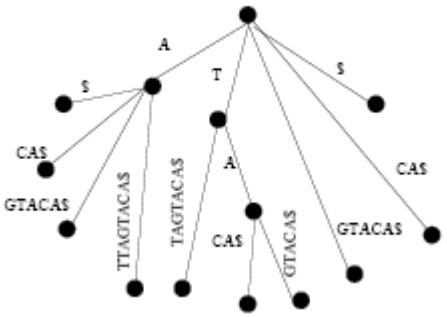


Figure 1. Suffix Tree Representation

The tree will inform every possible subsequence from a sequence as a pattern. One pattern is considered as particular path from the top node (root) to the most bottom node (leaf), for the example on the figure there are 10 possible sub sequences for the ATTAGTACA$ sequence. The $ character is added to inform the end of the sequence. There are three main function in this exploration process :

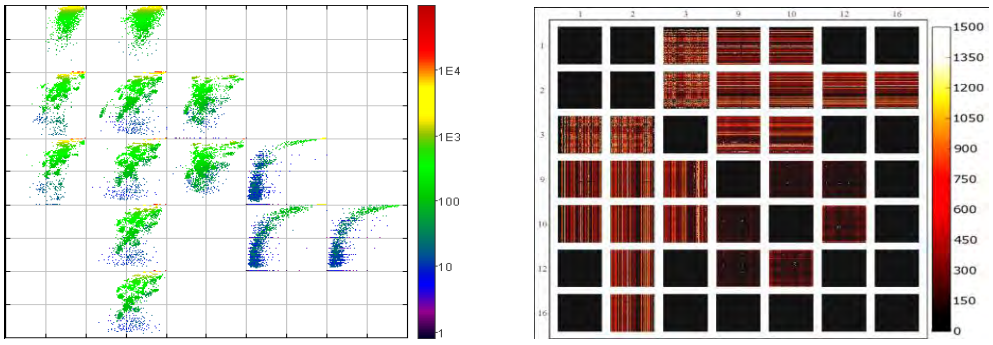
1. Build tree , construct suffix tree on the database. Every sequence (in FASTA format) subjected to the exploration should be transformed to the tree structure. Once it build, the FASTA format no longer needed so that it can be deleted and provide more space on the computer's memory.
2. Node searching, explore the tree for the queries, begin from the root (the top node) and end up at the leaf which is the most bottom node. If the query doesn't exist the system will report as "nothing". Each subsequence being found is indexed by number, represent its location on the sequence and its length (the number of the string).
3. Dispose, automatically erase the tree from the memory after it is stored on the database.

There will be 53.248 search on the *Vitis Vinifera* sequence's tree, the detail is explained in the following paragraph.

According to its restriction site, the restriction enzymes were classified into 44 groups and the simulation was conducted on 13 combinations among it. The combinations were determined as follow : (1) Three group having 4 bases of restriction site were paired with 3 group having 6 bases of restriction site, all with the most frequent match on the genome sequence; (2) The *Eco*RI and *Mse*I pair also included in the combinations although *Eco*RI do not fulfil the criteria because of there are facts that thus pair was used frequently for AFLP experiment [6,7,8,9]; (3) The restriction site of each pair do not overlap because such condition could lead bad and unpredicted restriction result. Three nucleotide selective bases were used for each subsequence's right and left hand end. Because there are 4 possible base (A,T,C,G), the total combination for selective bases should be 4^6 = 4.096. Therefore the total run for searching process on the tree is 13 x 4.096 = 53.248.

III.3 Cluster Analysis

The exploration result from the suffix tree then analyse by regarding on some criteria, which are : (1) Fragment (subsequence) length; (2) Percent of "in range" fragment, the number of fragment with the length does not exceed the polyacrylamide gel range criteria divided by the total fragment; (3) Percent of redundancy, the number of fragment with same length but different sequence divided by the total fragment. The analysis was done using multi dimension cluster analysis. The example of cluster representation is shown in Figure 2.



(a) (b)

Figure 2. Cluster Respresentation, (a) restriction enzyme and selective bases combination with their percent of "in range" fragment and percent of redundancy; (b) restriction enzyme and selective bases combination with their fragment length

III.4 The Selection Criteria

In order to find the best ten combinations of restriction enzyme and selective bases, the selection criteria should be well define. The combination will be considered good if : (1) Percent of redundancy less than 25%, too many different subsequence which have same length will reduce the polymorphism information; (2) Percent of fragment "in range" more

than 75%, too many fragment “out range” cause electrophoresis failure due to most of the fragment can not well visualize; (3) The average difference of the fragment length should be large enough so that it could be nicely separate on electrophoresis process. The selection is done by applying IF THEN rules.

III.5 Exponential Regression Model

To simulate the electrophoresis process, the system provide 1 Kbp DNA ladder from which the exponential model was developed. The exponential model between fragment size (bp) and its distance (cm) from the well is as follow:

$$\ln(\text{size}) = 10.81 - 0.736 * \text{distance}$$

IV. RESULT AND DISCUSSION

IV.1 Genome Description

The FASTA format of *Vitis Vinifera* genome sequence was separated in 19 different txt file, one file for one chromosome. Table 1 contains the description of each chromosome sequence component :

Table 1. The *Vitis Vinifera* Chromosome Sequence Description

Chromosome	Ambiguous bases per 1000 bases	GC Content (%)	Size of FASTA file (kb)
1	28	34,45	15.701
2	83	34,48	17.682
3	48	34,42	10.233
4	54	34,40	19.380
5	47	34,85	23.533
6	55	34,45	24.257
7	70	34,46	15.302
8	36	34,47	21.654
9	31	33,69	16.607
10	55	34,53	9.691
11	21	34,46	13.999
12	49	34,51	18.624
13	38	34,19	15.260
14	25	34,57	19.568
15	25	33,72	7.728
16	32	34,14	8.196
17	33	34,89	13.118
18	23	34,70	18.780
19	33	34,05	14.135
Total	**42**	**34,43**	**300.211**

IV.2 Exploration Process Performance

The main problem when facing with simulation of genome scale sequence is the operation time, but it is proven that by conducting suffix tree algorithm the operation time could be reduced significantly. Table 2 describes the time needed for suffix tree construction based on the size of the genome sequence. It is shown that the time needed increase in

linear form with the size of genome sequence, however the system could still operate in reasonable time (less than 3 minute) to handle genome sequence up to 12.1 Mbp long.. Once the suffix tree is constructed, all short pattern searching could be done in no time.

Table 2. Time for Suffix Tree Construction Based on Genome Sequence Size

Size of Genome Sequence (Mbp)	Time (second)
2,43	17
4,87	32
7,3	52
9,74	66
12,1	80

IV.2 Restriction Result Description

By conducting the data mining technique, there are several information that could be infer about the restriction result. It is known that a lot of small fragments were formed using a pair of restriction enzyme with 4 nucleotide restriction site, in the other hand just few bigger fragments were formed using a pair of restriction enzyme with 6 nucleotide restriction site. This facts were inline with the restriction digestion theory, restriction site with many nucleotide will have less probability to match the genome sequence. Therefore, the combination of restriction enzyme with 4 and 6 nucleotide of restriction site seems to be the better choice. These combinations will produce moderate number of fragments with moderate length as well.

IV.3 The Best Ten Combinations

Regarding to the selection criteria, the best ten combinations of restriction enzyme and selective bases were found. Table 3 describes thus combinations.

Table 3. The Best Ten Combinations Description

Rank	Restriction Enzyme		Selective Base		Range	Fragment in range		Redundancy	% of Restriction	% of Amplified Fragment
	1	2	1	2		Total	%			
1	AATT	ATGCAT	GCA	TAA	25-150 (126)	48	80,00%	22,92%	1,24%	0,04%
2	AATT	ATGCAT	GCA	CCA	25-150 (126)	44	80,00%	25,00%	1,24%	0,03%
3	AATT	AAGCTT	GCA	CTC	25-150 (126)	43	82,69%	23,26%	1,23%	0,03%
4	AATT	ATGCAT	GCA	TCT	25-150 (126)	42	76,36%	19,05%	1,24%	0,03%
5	AATT	ATGCAT	GCC	AAA	25-150 (126)	42	80,77%	23,81%	1,24%	0,04%
6	AATT	AAGCTT	GCA	ACA	25-150 (126)	41	77,36%	24,39%	1,23%	0,04%
7	AATT	ATGCAT	GCA	GAA	25-150 (126)	40	75,47%	20,00%	1,24%	0,03%
8	AATT	ATGCAT	GCC	TAA	25-150 (126)	38	77,55%	15,79%	1,24%	0,03%
9	AATT	ATGCAT	GCA	CGA	60-400 (341)	38	77,55%	21,05%	1,24%	0,03%
10	AATT	AAGCTT	GCA	TTT	25-150 (126)	38	77,55%	21,05%	1,23%	0,04%

Figure 3 depicts the visualization of virtual electrophoresis pattern based on the exponential regression model using 1 Kbp DNA Ladder. The blue line indicate that there is only one kind of subsequence with particular size, the green line indicate that there are two kind of subsequences with the same size, the red line indicate that there are three kind of subsequences with the same size and finally the black line indicate that there are more than three kind of subsequences with the same size. The black line should appears as the most thick and bright band in real gel electrophoresis result.

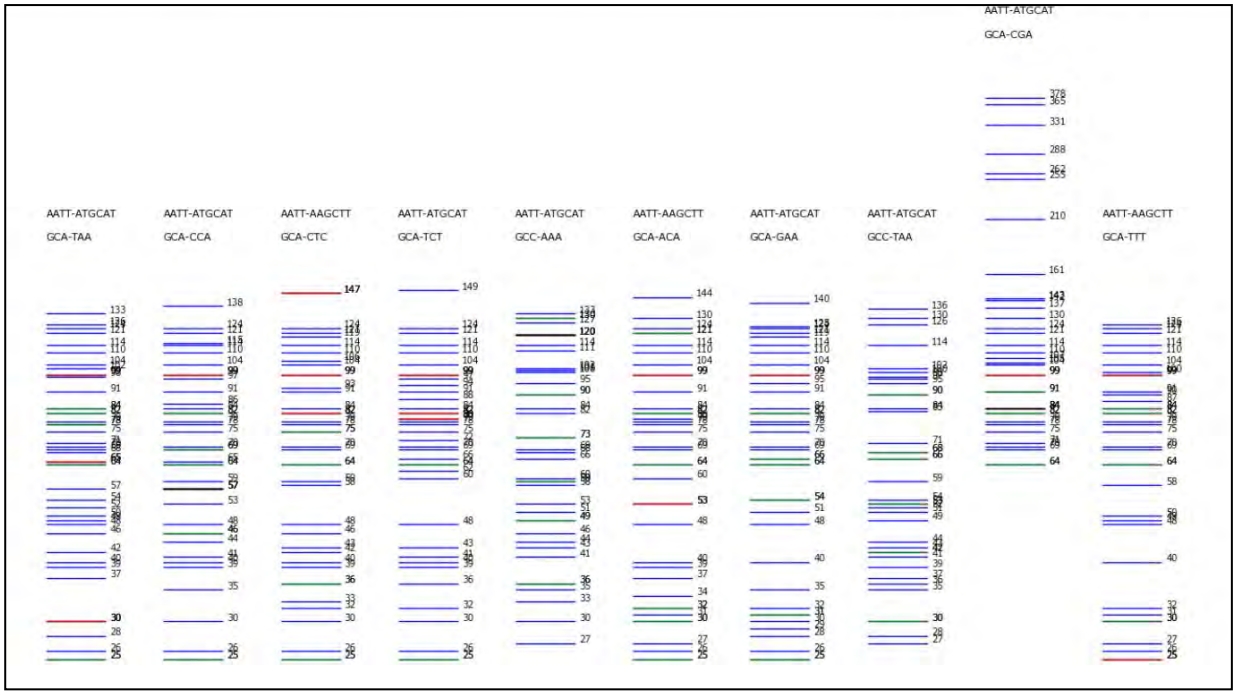


Figure 3. Visualization of The Virtual Electrophoresis Pattern

V. CONCLUSION

Like other simulation software, many factors embedded in laboratory experiment could not completely cover in this system, so that the result should be considered as recommendation (certainly with its probability of failure). However, so far the simulation result of AFLP with suffix tree indexing and data mining shows quite promising guidance for the laboratory experiment. The system developed in this research is a prototype from which more automatic and integrated system could be easily constructed. Machine learning technique such as genetic algorithm could be implemented to automate the optimization of selection criteria. At the end, laboratory conformation for this research result still could not leave behind. Therefore in the short incoming time such laboratory experiment should be conducted.

REFERENCES

[1] Korzun, V. 2003. *Molecular Markers and Their Applications in Cereals Breeding.A paper presented during the FAO international workshop on "Marker assisted selection: A fast track to increase genetic gain in plant and animal breeding?"*. Turin, Italy.

[2] Sui, N., Yao, M., Chen, L., Zhao, L., & Wang, X. 2008. *Germplasm and Breeding Research of Tea Plant Based on DNA Marker Approaches*. Front. Agric. China 2 (2): 200-207.

[3] Budiman, A. 2010. *Pembuatan Program Simulasi AFLP In Silico (Studi Kasus pada Vitis Vinivera)*.Thesis. Surabaya.

[4] D. Gusfield. 1997. *Algorithms on Strings, Trees and Sequences:Computer Science and Computational Biology*. Cambridge University Press.

[5] Tata, Sandeep., Hankins, Richard A., Patel Jignesh M. 2004. *Practical Suffix Tree Construction*. Proceedings of the 30th VLDB Conference, Toronto, Canada.

[6] Cervera, M. T., Gusmao, J., Steenackers, M., Van Gysel, A., Van Montagu, M., & Boerjan, W. 1996. *Application of AFLPTM-Based Molecular Markers to Breeding of Populus spp*. Plant Growth Regulation 20: 47-52.

[7] Koopman, W. J. M. & Gort, G. 2004. *Significance Tests and Weighted Values for AFLP Similarities, Based on Arabidopsis in Silico AFLP Fragment Length Distributions*. Genetics 167: 1915-1928.

[8] Mueller, U. G. & Wolfenbarger, L. L. 1999. *AFLP Genotyping and Fingerprinting*. TREE 14: 389-394.

[9] Stefenon, V. M., Gailing, O., & Finkeldey, R. 2006 *Phylogenetic Relationship Within Genus Araucaria (Araucariaceae) Assessed by Means of AFLPFingerprints*. Silvae Genetica 55 (2): 45-52.