SSR MARKER DEVELOPMENT IN SORGHUM &

PHYLOGENETIC STUDIES IN CEREALS.

Dissertation Submitted In Partial Fulfillment Of Requirement For The Award Of Degree Of

MASTER OF TECHNOLOGY in BIOTECHNOLOGY

By

HEMABINDU CHINTALA



CENTRE FOR BIOTECHNOLOGY Institute of postgraduate studies & research Jawaharlal Nehru Technology University Hyderabad-500028

20002

TO LORD SHRI SHIRIDHI SAI BABA





International Crops Research Institute for the Semi-Arid Tropics

atancheru 502 324 Andhra Pradesh India

Tei +91 40 3296161 (19 lines) Fax +91 40 241239 +91 40 3296182 Email ICRISAT@CGIAR.ORG

CERTIFICATE

This is to certified that the work reported in the dissertation entitled "SSR MARKER DEVELOPMENT IN SORGHUM & PHYLOGENETIC STUDIES IN CEREALS" Submitted by CH.Hemabindu have been carried out under my supervision. This work is towards the partial fulfillment of her M.Tech Degree from Jawaharlal Nehru Technological University, Hyderabad. This work is original and has not been submitted in part or full for any other degree or diploma of any university

1. me

IDr. V. MAHALAKSHMIJ Principal Scientist, GT-1, ICRISAT.



JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY CENTRE FOR BIOTECHNOLOGY

Institute of Post Graduate Studies and Research Mahaveer Marg, Hyderabad - 500 028, A.P., India Phone No. : 040-3373020

Dr. M. LAKSHMI NARASU Associate Professor & Head

CERTIFICATE

This is certified that the work reported in the dissertation entitled "SSR MARKER

DEVELOPMENT IN SORGHUM & PHYLOGENETIC STUDIES IN CEREALS"

Submitted in partial fulfillment for the award of M.Tech in biotechnology from

Jawaharlal Nehru Technological University, Hyderabad, is a bonafied work carried out

by Ms CH. HEMEBINDU under the guidance of Dr.V.Mahalakshmi,

Senior scientist, International Crops Research Institute For Semi-Arid Tropic.(ICRISAT)

[Dr.M.Lakshmi Narasu]

DECLARATION

I Hemabindu.Ch, a bonafied student of IPGSR, JNTU. Hyderabad here by declare that the dissertation entitled "SSR Marker Development In Sorghum And Phylogenetic Studies In Cereals" is solely done by me under the expertise guidance of Dr.V.Mahalakshmi at International Crops Research Institute For Semi-Arid Tropics (ICRISAT), Hyderabad.

The facts and figures enumerated in this project work are in accordance with the results of the modeling done in computer. This project work has not been submitted to any university or institution for the award of any degree or diploma.

ACKNOWLEDGEMETS

This project work is carried out with valuable suggestions and guidance under the supervision of Dr.V.Mahalakshmi, seniorscientist, GREP, ICRISAT, patancheru, Hyderabad. Her unstilted encouragement and deep concern helped me to complete my project work in time. I am highly grateful and indebted to her.

It gives me immense pleasure in expressing my deep sense of gratitude to Dr. Lakshmi Narasu, Head, Centre for biotechnology, Jawaharlal Nchru University, Hyderabad.I take this opportunity to thank her for suggesting me to join in the ICRISAT as an Apprentice And helped me in finishing my project work.

My sincere thanks to Dr.Prameela Devi, Dr.Archana giri associate professors and Mr.kiran, Ms.Anuradha and Ms. Radhika academic associates, Centre for biotechnology, JNTU, for their valuable guidance.

I would like to thanks to my friends Ms.Manjula, Ms.Leela, Mr.Kumar, Mr.PVNS Prasad, Ms Rekha and Mr. A.Balakrishna and other colleagues in ICRISAT for their encouragement and cooperation during my course of project work.

I show my gratitude to my beloved mother, father, brother's and sister& my Husband for their constant encouragement and good support throughout this course of work.

HEMABINDU.CH

ABSTRACT

Bioinformatics, the application of computational techniques to analyze the information associated with bio-molecules on a large scale and encompasses a wide range of subject areas from structural biology, genomics to gene expression studies.

Chapter 1 gives an introduction and overview of the SSR marker development in sorghum . For this purpose bioinformatics tools like Tandem Repeats Finder, Primer3, Windows software and MS Access were used. More than 50,000 records were collected and placed in the INTRANET of ICRISAT, which could be accessed by the scholars and scientists for their requirements.

A tandem repeat in DNA is two or more contiguous approximate copies of a pattern of nucleotides. Extensive knowledge about pattern size, copy number, mutational history, etc. for tandem repeats has been limited by the inability to easily detect them in genomic sequence data. These sequences range in size from 3kb up to 700kb. A World Wide Web server interface at <u>http://c3.biomath.mssm.edu_trl_advanced.submit.html</u> has been established for automated use of the program.

Primer design is crucial for the success of PCR. Inappropriate primers cause low yield, equivocal results and misinterpretation. An ideal primer should only bind with its 3' end to a unique sequence.

Chapter 2 is dealt with an introduction about Phylogenetic studies and describes completely about the phylogeny of selected cereals for conserved enzymes. For this purpose we utilized the software tools like CLUSTAL W for multiple alignment, JALVIEW for alignment analysis and phylogenetic tree construction, AND PRIMER 3 for primer designing which is crucial for PCR.Phylogeny is about evolution and is used to reconstruct evolutionary events. It is now possible to construct phylogenetic evolution at a molecular level through analysis of molecular sequences. namely proteins & nucleic acids.To construct phylogenetic tree among grass family, the sequences of conserved enzymes from mitochondria, chloroplast and nucleus are probed using bio-informatics tools.

CONTENTS

1. SSR Marker Development For Sorghum Data Base1
1.1 Introduction1
1.2 Different Tools For Repeat Finder2
1.2.1 Censor_xxx_Humrep [Censor]3
1.2.2 Repeatmasker_Xxx Primate [Repmask]3
1.2.3 SST_Xxx_HumRep [SST]3
1.2.4 Xnun_Repeat_Default [XNUN] 3
1.2.5 TANDEM_Xxx_DEFAULT [TANDEM3
1.2.6 Inverted_Xxx_Default [INVERTED]3
1.3 Tandem Repeats Finder4
1.4.1 Basic
1.4.2 Intermediate4
1.4.3 Advanced4
1.5 Advanced Tandem Repeat Finder Program Parameters6
1.5.1 Alignment Parameters
1.5.2 Minimum Alignment Score6
1.5.3 Maximum Period Size6
1.5.4 Detection parameters
1.5.5 Options
1.5.5.1 Flanking sequence6
1.5.4.2 Masked sequence File6
1.5.4.3 Data File6
1.6 Procedure For Finding Tandem Repeats8
1.7 Entering the sequence for Finding Tandem Repeats8
1.7.1 Sequence
1.7.2 Fasta Format

1.7.3 Submit sequence
1.8 Table Explanation10
1.9 Alignment Explanation11
1.10 Primer Design12
1.10.1 Introduction
1.11 Primer Design Programs 12
1.12 Primer Design Considerations13
1.13 Features Of Primer Design14
1.14 Limits Of Primer Design14
1.15 Primer Design Parameters15
1.15.1 Primer Length15
1.15.2 Primer Sequence15
1.15.3 GC content
1.15.4 Melting temperature16
1.15.5 Secondary structure formation17
1.15.6 Specificity
1.15.7 Primer ends17
1.16 Primer3
1.17 Procedure For Primer Design By Using Primer318
1.18 Primer3 Input Parameters19
1.18.1 Source Sequence
1.18.2 Sequence Id
1.18.3 Targets
1.18.4 Excluded Regions19
1.18.5 Product Size
1.18.6 Number To Return20
1.18.7 Max 3' Stability20
1.18.8 Max Mispriming20
1.18.9 Pair Max Mispriming20
1.18.10 Primer Size20

1.18.11 Primer T _m	21
1.18.12 Maximum T _m Difference 1.18.13 Product T _m	
1.10.15 Floudet 1 m	
1.18.14 Primer GC%	21
1.18.15 Max Complementarity	21
1.18.16 Max 3' Complementarity	22
1.18.17 Max Poly-X	23
1.18.18 Included Region	23
1.18.19 Start Codon Position	23
1.18.20 Mispriming Library	23
1.18.21 CG Clamp	23
1.18.22 Salt Concentration	23
1.18.23 Annealing Oligo Concentration	24
1.18.24 Max Ns Accepted	2:4
1.18.25 Liberal Base	24
1.18.26 First Base Index	24
1.18.27 Inside Target Penalty	24

1.18.28 Outside Target Penalty24
1.18.29 Sequence Quality25
1.18.30 Min Sequence Quality25
1.18.31 Min 3' Sequence Quality25
1.18.32 Sequence Quality Range Min25
1.18.33 Sequence Quality Range Max25
1.18.34 Penalty Weights25-26
1.18.35 Hyb Oligos (Internal Oligos)27
1.19 Methods For Finding Tandem Repeats30
1.19.1 Steps Involved After Downloading The Sequence
Of Sorghum From NCB130
1.19.1.1 Fasta Format
1.19.1.2 Enter The Sequence For Finding Tandem Repeats
1.19.1.3 Submit Sequence
1.19.1.4 Tandem Repeats Report
1.19.1.5 Summary Table
1.19.1.6 Alignment Explanation Table34
1.19.1.7 Primer335-37
1.19.1.7 Out Put Of Primer3
1.20 Results41-48
1.21 Discussion49
2. Phylogenetic studies in cereals50
2.1 Introduction
2.2 Phylogenetic terms51-56
2.3 Phylogenetic classifications
2.4 Methods of phylogenic analysis

2.4.1 Cladistic Method	57
2.4.2 Phenetic Method	57
2.4.3 Multiple Alignment Method	58
2.5 Clustalw	58
2.5.1 upload a file	59
2.5.2 Sequences	60
2.5.3 search title	61
2.5.4 CPU mode	61
2.5.5 alignment	61
2.5.6 Output	61
2.5.7 JalView	62
2.5.8 out order	62
2.5.9 color	62
2.5.10 Fast paiwise alignment options	62
2.5.11 multiple sequence alignment options	62
2.5.12 Gap open	63
2.5.13 End gap	63
2.5.14 Gapext	63
2.5.15 Gapdist	63
2.5.16 Phylogenetic Tree	63
2.5.17 kimura correction of distances	63
2.5.18 Ignore Gaps In Alignment	64
2.6 Phylogenetic Tree	
2.7 Methods For The Phylogenetic Studies Of Cereals	66
2.8 Reasons For Taking Enzymes In Phylogenetic Studies	
2.8.1 Nuclear Enzyme	66
2.8.2 Mitochondrial Enzyme	
2.8.3 Chloroplast Enzyme	
2.9 Multiple Alignment Method	
2.10 Steps Involved In Multiple Alignment Method	
2.10.1Select The Most Conserved Enzymes	
••••••••••••••••••••••••••••••••••••	

2.10.2 Search The Sequences From NCB1	68-69
2.10.3 Accession Number	72-80
2.10.4 Exon Regions	81
2.10.5 Multiple Alignment	81
2.10.6 Run Clustalw	82
2.10.7 JalView	83-87
2.10.8 Alignment Graph	
2.10.9 Phylogenetic Tree	93
2.10.10 Tree Analysis	94
2.10.11 Design Primers For The Sequences Of Maize	94
2.10.12 Selection Of First Set Of Primers	95
2.10.12.1 Primer3 For Left Primer	95-97
2.10.12.2 Primer3 Output For Left Primer	98
2.10.12.3 Primer3 For Right Primer	
2.10.12.4 Primer3 Output For Right Primer	102
2.10.12.5 Calculation Of Product Size	103
2.10.13 Selection Of Second Set Of Primers	103
2.10.13 .1 Primer3 For Left Primer	103
2.10.13 .2 Primer3 Output For Left Primer	106
2.10.13 .3 Primer3 For Right Primer	107-109
2.10.13 .4 Primer3 Output For Right Primer	110
2.10.13 .5 Calculation Of The Product Size	111
2.10.14 Mitochondrial And Nuclear Enzyme	111-112
2.11 Results	113-114
2.12 Discussion	115

WEBSITES USED IN THE PROJECT WORK

- 1) http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi
- 2) http://www.ncbi.nlm.nih.gov/entrez/query.fcgi
- 3) http://www.ebi.ac.uk/clustalw/
- 4) http://www.ncbi.nlm.nih.gov/
- 5) http://c3.biomath.mssm.edu/trf.html
- 6) http://c3.biomath.mssm.edu/example.html
- 7) http://c3.biomath.mssm.edu/trf.definitions.html#fasta
- 8) http://c3.biomath.mssm.edu/trf.submit.options.html
- 9) http://c3.biomath.mssm.edu/trf.advanced.submit.html
- 10) http://c3.biomath.mssm.edu/trf.upload.form.html
- 11) http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide
- 12) http://www2.ebi.ac.uk/~michele/jalview/contents.html
- 13) http://www.expasy.ch/enzyme/
- 14) http://www.genome.ad.jp/
- 15) http://www.mssm.edu/school.html

LIST OF FIGURES

S.No	Subject	page No.
1	Tandem Repeat Finder Option Page.	5
2	Tandem Repeat Finder Advanced Submit Page.	7
3	Submission Of Sequence To Advance Tandem Repeat Finder.	9
4	Primer3.	28- 29
5	Alignment Explanation Table.	34
6	NCBI Home Page.	70
7	NCBI Nucleotide Page.	71
8	ClustalW Submission Form.	82

BIOINFORMATICS

Introduction:

Bioinformatics is conceptualizing biology in terms of molecules (in the sense of physicalchemistry) and then applying "informatics" techniques (derived from disciplines such as applied math, Computer Science, and statistics) to understand and organize the associated with these molecules, on a large scale.

In the last few decades, advances in molecular biology and the equipment available for research in this field have allowed the increasingly rapid sequencing of large portions of the genomes of several species. In fact, to date, several bacterial genomes, as well as those of some simple eukaryotes (e.g., *Saccharomyces cerevisiae*, or baker's yeast) have been sequenced in full. The Human Genome Project, designed to sequence all 24 of the human chromosomes, is also progressing. Popular sequence databases, such as Genbank and EMBL, have been growing at exponential rates. This deluge of information has necessitated the careful storage, organization and indexing of sequence information. Information science has been applied to biology to produce the field called **Bioinformatics**.

The most pressing tasks in bioinformatics involve the analysis of sequence information. **Computational Biology** is the name given to this process, and it involves the following:

- 1. Finding the genes in the DNA sequences of various organisms
- Developing methods to predict the structure and/or function of newly discovered proteins and structural RNA sequences.
- 3. Clustering protein sequences into families of related sequences and the development of protein models.
- 4. Aligning similar proteins and generating phylogenetic trees to examine evolutionary relationships.

The Need For Bioinformatics

- 1. Whole Genome Analyses and Sequences
- 2. Experimental Analyses involving Thousands of Genes simultaneously
- 3. DNA Chips and Array Analyses
- 4. Expression Arrays
- 5. Comparative Analyses between Species and Strains
- 6. Proteomics: 'Proteome' of an Organism ... 2D gels, Mass Spec
- 7. Medical applications: Genetic Disease ... SNPs
- 8. Pharmaceutical and Biotech Industry
- 9. Forensic applications
- 10. Agricultural applications

Evolution Of Bioinformatics

After years of research in structure-function relationships of genes and proteins, the last decade proved to be extremely important and immensely satisfying due to its technical advances in genome sequences of several species and protein structure and identification. To handle this ever-increasing voluminous data, computer processing power and disk storage has been instrumental. Besides gathering all these data, it is necessary to compare these nucleotide and amino acid sequences to find similarities and differences. Since it is not convenient to compare the sequences, that are several hundreds of nucleotides by hand, several computational techniques were developed to approach this problem. In addition these are less error-prone than the manual approach. So bioinformatics has taken its place to cater the needs of biological community.

Divisions Of Bioinformatics

Bioinformatics is a multi disciplinary subject. Though only about a decade old, it has become very important for the growth of biosciences, biotechnology, and the economic prosperity of nations. Three well-identified subdivisions of Bioinformatics are:

- a) Molecular Bioinformatics
- b) Cellular and sub-cellular Bioinformatics and
- c) Orgasmic and community Bioinformatics.

Out of these three, most Bioinformatics scientists and workers practice molecular Bioinformatics. The other two areas are more recent and are at different stages of development. In the next 5-10 years, cellular and sub-cellular Bioinformatics that will include metabolic pathways, epigenetic, and neuro Bioinformatics on one hand and Bioinformatics of Species Diversity, behavior, evolution and the effect of pollutants on higher as well lower species, on the other will occupy the main stage.

Global Importance Of Bioinformatics

Bioinformatics has acquired great importance due to its recent application in vast amount of data generated in the Genome sequence projects. The target of decoding the three billion base pairs of the human DNA has become achievable only through the use of various innovative techniques and methods evolved by the Bioinformatics scientists. Bioinformatics has become an essential component of biotechnology based product and process development. The process of drug design and development is expensive and time consuming. The application of the tools and techniques of Bioinformatics has resulted in the reduction in cost and the development cycle of the drugs. This aspect has a tremendous impact on the society. If a newly discovered drug is a life-saving one, then the resulting gains are not only in terms of financial savings but also in saving the lives of several million people. Major pharmaceutical and biotechnology companies have set-up large R&D groups in Bioinformatics. Current research has identified biotechnology the fastest growing sector of production technology. Further advances in this sector will depend quite a lot upon the progress of Bioinformatics and hence there is a great emphasis on Bioinformatics world over. The following are a few important Websites for international networks/ institutions/ groups on Bioinformatics.

The Role Of Bioinformatics In Research

As traditional way of research is very time consuming and laborious, in these days there is introduction of computers and even biological sciences are no exception to this. The long sequences and enormous flow data in molecular biology it becomes difficult to manage data by persons or by organizations. Further with the help of computers through Internet and Intranet, sharing of data with others in the organization or organizations through out the world is becoming routine. This aids in the research and makes it easier for scientists to download, analyze, compare and exchange information. This also saves a lot of time and finances.

Uses Of Bioinformatics

- The main role of bioinformatics for today's world is to increase productivity and quality of all plants and animals.
- For understanding evolution, comparative genomics will add hither to unimaginable comprehensiveness to the study of the relationships between species.
- In cell biology. the components of cellular activity, how these components interact and how they are influenced by environmental states can be identified comprehensively.
- 4. In medicine, sequences will provide a basis for the study of susceptibility to disease, it's pathogenesis and the development of new preventive and therapeutic approaches.
- 5. Bioinformatics is extremely useful to mankind as it can help in increasing crop yields and animal produces, to cater the needs of ever increasing human

population. It also helps to develop more effective drugs to protect public health as well as to control pests and diseases on the crop.

1. SSR Marker Development In Sorghum

1.1 Introduction

DNA molecules are subject to a variety of mutational events. One of the less well understood is **TANDEM DUPLICATION** in which a stretch of **DNA**, which we call the pattern, is converted into two or more copies, each following the preceding one in a contiguous fashion.for example we could have

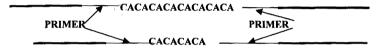
....TCGGA.... 🛛 TCGGCGGCGGA....

In which the single occurrence of triplet CGG has been transformed into three identical, adjacent copies. The result of a tandem duplication event is termed as **TANDEM REPEAT.** Over time, individual copies within a tandem repeat may undergo additional, uncoordinated mutations so that so typically, only approximate tandem copies are present. Tandem repeats are ubiquitous sequence features in both prokaryotic and eukaryotic genomes. These are highly useful as genetic markers. They are codominant, occur in high frequency, and appear to be distributed through out the genomes of most, not all the higher plants and animals. They also display a high level of polymorphism, even among closely related accessions, and are amenable to simple and inexpensive Polymerase Chain Reaction (**PCR**) assays (Brown et. al. 1996).

Tandem repeats are usually classified among satellites (spanning megabases of DNA, associated with heterochromatin), minisatellites (repeat units in the range 6-100 bp, spanning hundreds of base pairs) and microsatellites (repeat units in the range 1-5 bp, spanning a few tens of nucleotides). The minisatellites are also called "various number tandem repeats" or VNTRs. The microsatellites are also called "short tandem repeats" or STRs. (Short tandem repeats, STRs) contain 2-5 bp repeats VNTR 'S are scattered at various locations in the Genome are regions that are highly variable. These regions contain a type of DNA sequence called variable number Tandem Repeat. Tandem repeats are multiple copies of sequence of base pairs arranged in head to tail fashion. For example, a frequently found Tandem Repeat is CA, and one strand containing this type of repeat reads CACACA...

GTGTGT.... In this example, the number of repeating base pairs is two, but it can be more. When the repeating unit is less than four, the VNTR is called a MICROSATELLITE and when the repeating unit is longer it is a MINISATELLITE.

MICROSATELLITE are DNA regions with variable numbers of short tandem repeats flanked by a unique sequence. Microsatellites make good genetic Markers because they each have many different 'alleles'-i. e. There can be many different lengths of the repeat region. The number of repeats there are at the same location defines an allele. With many alleles, most individuals are heterozygous, giving power to note association between marker allele and performance in progeny inheriting a favorable linked QTL allele.



Tandemly repeated sequences are especially liable to undergo misalignments during chromosome pairing, and the size of tandem clusters tends to be highly polymorphic. The smaller clusters of this simple sequence can be used to characterize individual genomes in the technique of "DNA finger printing". Comparisons of corresponding regions of simple sequence DNA with in and between species are informative about the mechanisms involved in manipulating sequences

1.2 Different Tools For Repeat Finder

1.2.1 censor_xxx_humrep [censor]

This tool searches for genome wide repeats and classifies its findings. Each incarnation of this tool produces two tables containing ALU (xxx=ALU) and non ALU (xxx=NONALU) repeats. In this incarnation the ``Humrep" repeat database [humrep] is used and the search sensitivity is set to ``moderate". If the input sequence was derived from other mammalians than human the ``MamRep" repeat database - which comes along with CENSOR too - is used.

1.2.2 repeatmasker_xxx primate [repmask]

This tool searches for both genome wide and simple repeat and classifies its findings. Each incarnation of this tool produces three tables containing ALU (xxx=ALU), non ALU (xxx=NONALU) and simple (xxx=REPEAT) repeats. In this incarnation the "Primate" repeat database is used. The search sensitivity is set to "high" and tagging of simple repeats is turned on. If the input sequence was derived from other mammalians than human the "Rodent" or "Vertebrate" repeat database - which comes along with RepeatMasker too - is used.

1.2.3 sst_xxx_humrep [sst]

This tool searches for both genome wide and simple repeats and classifies its findings. Each incarnation of this tool produces three tables containing ALU (xxx=ALU), non ALU (xxx=NONALU) and simple (xxx=REPEAT) repeats. The ``HumRep" repeat database [humrep] is used.

1.2.4 xnun_repeat_default [xnun]

This tool searches for short-period repeats (micro-satellites) and masks its findings. In order to reduce the rate of false positives the default probability cut is lowered from 1% to 0.1%.

1.2.5 tandem_xxx_default [tandem]

This tool searches for short-period repeats (micro-satellites) and reports the consensus repeat unit. Each incarnation of this tool produces two tables. The first table (xxx=STRAND) contains all repeats found together with the consensus repeat unit and the strand information. The second table (xxx=REGION) contains just the repeat regions and is cross-linked with the first table.

1.2.6 inVerted_xxx_default [inverted]

This tool searches for inverted repeats within a window of 2 kbp. The window size is an internally fixed parameter of the analysis tool itself. Each incarnation of this tool produces two tables. The first table (xxx=STRAND) contains two entries for each

inverted repeats. One for the forward and one for the reverse strand repeat unit. The second table (xxx=REGION) contains just the repeat regions and is cross-linked with the first table.

1.3 Tandem Repeats Finder

A Tandem repeat in DNA is two or more adjacent, approximate copies of a pattern of nucleotides. Tandem Repeats Finder is a program to locate and display Tandem Repeats in DNA sequences. In order to use this program, we have to submit the sequence in **FASTA** format. There is no need to specify the pattern, the size of the pattern or any other parameter. The program's analysis is sent back as two files, a summary table file and an alignment file. The summary table contains information about each repeat, including its location, size, number of copies and nucleotide content. Clicking on the location indices for one of the table entries opens a second web browser that shows an alignment of the copies against a consensus pattern. The program is very fast, analyzing sequences on the order of .5Mb in just a few seconds. Submitted sequences may be up to 5Mb in length. Repeats with pattern size in the range from 1 to 500 bases are detected.

1.4 Levels Of Tandem Repeat Finder:

There are 3 LEVELS of tandem repeat finders.

1.4.1 Basic

It uses default parameters (recommended for beginners.)

1.4.2 Intermediate

It provides the parameter Maximum period size, and options Flanking sequence and Masked Sequence File.

1.4.3 Advanced

It provides the parameters like Alignment parameters (match, mismatch and indels), Minimum Alignment Score To Report Repeat, Maximum Period Size and options like Flanking sequence, Masked sequence file, Data file. File Edit View Favorites Tools Help

TANDEM REPEATS FINDER

Submitting Your Sequence

Please Select one of the following options

Use default parameters (recommended for beginners.)

anticontenties and Sel some parameters and selent options

CONCRETE Set all parameters and options

1997 (Stomathematical) Sed as cusine a same so 1997 Sciences

Last revised July 22, 1999

TANDEM REPEAT FINDER OPTION PAGE

ŝ

1.5 Advanced Tandem Repeat Findeer Program Parameters:

Input to the program consists of a sequence file and the following parameters:

1.5.1 Alignment Parameters

Weights for match, mismatch and indels. Lower weights allow alignments with more mismatches and indels. Match weight is +2 in all options here. Mismatch and indels weights [interpreted as negative numbers] are 3, 5, or 7. A 3 is more permissive and a 7 is less permissive of these types of alignment choices.

1.5.2 Minimum Alignment Score

The alignment score must meet or exceed this value for the repeat to be reported.

1.5.3 Maximum Period Size

The period size must be no larger than this value for the repeat to be reported. The program will find all repeats with period size between 1 and 500, but the output table can be limited to some other range.

1.5.4. Detection parameters

Matching probability P_m and indeal probability P_i . $P_m = .80$ and $P_i = .10$ by default and it cannot be modified in this version of the program.

1.5.5 Options

1.5.5.1 Flanking sequence

Flanking sequence consists of the 200 nucleotides on each side of a repeat. Flanking sequence is recorded in the alignment file. This may be useful for PCR primer determination.

1.5.4.2 Masked sequence File

The masked sequence file is a FASTA format file containing a copy of the sequence with every character that occurred in a tandem repeat changed to the letter 'N'. The word "masked" is added to the sequence description line just after the '>' character.

1.5.4.3 Data File: The data file is a text file, which contains the same information, in the same order, as the summary table file, plus consensus sequences. This file contains no labeling and is suitable for additional processing, for example with a perl script, outside of the program.

nterangenerangen in den service son den state en	
IANDEM REPEATS FINDER	
Parameters:	
Aligeneter Paraseters (maska menetik, möste) [(2,7,7)]	
Minimum Abgament Stores To Report Report 55 🔄	$= \frac{1}{2} \sum_{i=1}^{n} \left(\left(-\frac{1}{2} + \frac{1}{2} \right) + \left(-\frac{1}{2} + \frac{1}{2} $
Marinet Report Size 100 -	and the second ready
Opticas: Flashing Sequence F Mesked Sequence File F Dels File F	
Your data must be a DNA sequence in FASTA frames. The second second	
Your data must be a DNA sequence in FASTA frames. The second second	
Your deste must be a E Né, se quante in FASTA format. Choose one of the following weys to early your site	
You dala nuset be a DNA sequence in PASTA themes. So Theose and of the following weys to send your dita) <u>A Uplace e file from your theorory</u>	
Econce.	
You date must be a DNA sequence in PASTA format. Choose mit of the following weys to send you date 1 I Johnson Effe from your disectory Econory Econory	
You date must be a DNA sequence in PASTA format. Choose mit of the following weys to send you date 1 I Johnson Effe from your disectory Econory Econory	
You date must be a DNA sequence in PASTA format. Choose mit of the following weys to send you date 1 I Johnson Effe from your disectory Econory Econory	

TANDEM REPEAT FINDER ADVANCED SUBMIT PAGE.

1.6 Procedure For Finding Tandem Repeats:

Web sites like <u>http://c3.biomath.mssm.edu/trf.advanced.submit.html</u> are used to find out the tandem repeats of particular sequence (source sequence). This web site is provided by The Department of Biomathematical sciences, Mount sinai School of Medicine.

- Down load the source sequence from the www.ncbi.nlm.nih/gov/entrez
- Open the page http://c3. Biomath.mssm.edu/trf.advanced.submit.html
- Enter the source sequence in cut and paste sequence blank.
- Click the submit sequence button.

1.7 Entering The Sequence For Finding Tandem Repeats:

1.7.1 Sequence:

Data must be a DNA sequence in FASTA format.

1.7.2 FASTA format:

The FASTA format looks like this,

gi|18063613|gb|BM324719.1|BM324719 PIC1 34 G05.b1 A002 Pathogen-infected (PIC1) compatible 1 Sorghum bicolor cDNA. mRNA sequence GCACGAGGCACACACACACACTGACGTACGATATCGAACACACTGACACAA CACAACCGCCGGGCGCCGCCGGCTTTGTCGTCAACGTCACCATTCCTTAACA CACTCTCTACATACACGTCGCAGCCGGCCGGCAGCTCATCTTCTGTATTTAA ACGCGCTCCATCCATTCCGGTCAGGGCACGGAGTCGTCACTCCTAGCTCCGC **GCGCACCGAGCTCCCACACACTCTCGCAGCAATGGCACGCGGCGGCGGCTCTACC** GCGGCTCCTCCACCCCTCCCGCCGCTGCTGCTGCTGCTGGTGGTAGTAGTGG CACTACGCCGGGTACGTGACGGTGGACGAGGCGCACGGCAGGAGGCTCTTC TACTACCTGGTGGAGCCGAGCGTGACCCCGCCAAGGACCCCGTCGTGCTGTG GCTCAACGGCGGGCCTGGCTGCTCCAGCTTCGACGGC

The first line starts with a greater than sign "..." and contains a name or other identifier for the sequence. The remaining lines contain the sequence data. The sequence can be in upper or lower case letters. Anything other than letters mambers for example its ignored

1.7.3 Submit sequence:

By clicking this submit sequence the program analysis will exists on screen

	e Sile ikun Panderan Sunda yak 🚺 🚺 👘
TANDEM REPEATS FI	NDER
Parameters: Alegeneerd For meters (noteds, mercarity, so bein 1997)	
Ministration Alegonitional Bactors Tai Forgatet Regiones 💱 🔡	$= \int \sum_{i=1}^{n} \frac{1}{i} \sum_{j=1}^{n} \frac{1}{i} $
M contrast 7 contra la finte (
<mark>Opideman</mark> Handwarg Bengaennes F., Mandard Vanguennes Van F., Dietwieller F.	
Sequences show for second to a LNA sequence on APTA former (2007) and the Choices one of the following map and exile pole data.	
 ²⁷⁷ ปัฐปัตสะดิษที่ไขที่อย่อง รูปตาสะเสียงการอาก 	
j 🛛 🕅 Tradiensk engelsk	
4-11 (a) - 42. For a power power particular of the antimic relation of a state of the land a post power particular of the state of the state of the land a power particular of the state of the state of the state of the the state of the state of the state of the state of the state of the the state of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the state of the state of the state of the power particular of the state of the power particular of the state of the power particular of the state o	nina Tana Tana
Sectors operates : Claence - Normalien	
TABLE RACES FOR A CONTRACTOR STREET, ST	

SUBMISSION OF SEQUENCE TO ADVANCED I ANDEM REPEAT FINDER.

The first line starts with a measured an sign and contains a name or other identifier for the sequence. The remaining lines contain the sequence d to The source can be in upper or lower case letters. Anothing other than letters number or example is a mored.

173 Submit sequence

By cheking this submit sequence the program inclusion will axists on screen

	şá	-39	ç.	(CACA)
TANDEM REPEATS FI	NDER			
Parameters: an all trans at them				
المنت "دور مراجع العند" العند العن العند الع المنابع				
Ogdano a_ I [™] I [™] c ^a				
Sequence Land the Line Levi the No. 2010				
<u> </u>				
	-			
10.00				

SUBMISSION OF SEQUENCE TO ADVANCED FANDEM REPEAT FINDER

1.8 Table Explanation:

The summary table includes the following information:

- 1. Indices of the repeat relative to the start of the sequence.
- 2. Period size of the repeat.
- 3. Number of copies aligned with the consensus pattern.
- 4. Size of consensus pattern (may differ slightly from the period size).
- 5. Percent of matches between adjacent copies overall.
- 6. Percent of indels between adjacent copies overall.
- 7. Alignment score.
- 8. Percent composition for each of the four nucleotides.
- 9. Entropy measure based on percent composition.
- 10. If the output contains more than 140 repeats, multiple linked tables are produced.

يات. Back) Stop I	글 Refresh Home		Al J vonites Histor		Print	BY Edu	•	Uints "	Eile	<u>E</u> di "	T
ddress 🚑	http://c3.bi	oniath missin e	da/ht/4/zE2/Oos	yde2.2.3.5.00.1	n au 10,1 ktol							- e	×6
				athens of B Hount Sinai		nical Scie Medicino							
heleic A	u Seperator f Neid Pepe	inder: a arch(1999 p. 573-58		alγse ⊨₩k s	ooptones"								
		- 50 10 30 3	10										
ables:	1												
h13 13 1	able 1	of 1											
lick on	indices	to view à	lignment										
able Expl	nation												
In	lices	Peno- Ciza	i Copy Number	Consensus Size	Ferent Matcheo	Percent Indels	Scote	٨	с	0	т	Entropy (0.2)	
ų	32	1	23.0	2	20	1.5	30	43	ŝo	10	10	174	
	<u>.</u>	- 6	37	6	07	0	34	6	31	40	27	1.56	
. 23	<u></u>	9	2.4	0	54	0	34	0	51	-40	27	1.56	
The End													

SUMMARY TABLE

1.9 Alignment Explanation:

The alignment is presented as follows:

- In each pair of lines, the actual sequence on top and a consensus sequence for all the copies is on the bottom.
- 2. Each pair of lines is one period except very small patterns.
- 3. The 10 sequence characters before and after a repeat is shown.
- 4. Symbol * indicates a mismatch.
- 5. Symbol indicates an insertion or deletion.
- Statistics refers to the matches, mismatches and indels overall between adjacent copies in the sequence, not between the sequence and consensus pattern.

- Distances between matching characters at corresponding positions are listed as distance, number at that distance, percentage of all matches.
- 8. A, C, T, G count is percentage of each nucleotide in the repeat sequence.

1.10 Primer Design

1.10.1 Introduction:

Designing **PCR** and sequencing primers are essential activities for molecular biologists around the world. Primer design was developed to find suitable primers for PCR or oligo nucleotides for probes and **DNA** sequencing. Primer design is crucial for the success of PCR. Inappropriate primers cause low yield and misinterpretation. Primers that bind to multiple DNA loci can synthesize side products and render sequencing illegible, especially with high amplification of small amounts of DNA and with impure DNA. They are generally the result of short DNA sequence repeats. An ideal primer should only bind to a unique sequence. To ensure this the given sequence must be compared with itself to identify repeats.

Primer Design is a DOS-program to choose primer for PCR or oligonucleotide probes. Napiwotzki, J. and Becker, A. wrote this program in 1995. It is tailored to check known sequences for repeats and unique sequences and subsequently to create proper primers according to this data.

1.11 Primer Design Programs:

- PrimerGen searches strings of amino acid residues in order to reverse-translate oligonucletide primers of a desired range of lengths and maximum number of degeneracies.
- Primer (Stanford) Sun Sparcstations only
- Primer (Whitehead) Unix, Vms (and DOS and Mac if you can compile it)

- Amplify, Bill Engels (Macintosh only), is for use in designing, analyzing, and simulating experiments involving the polymerase chainreaction (PCR). You can obtain your copy of Amplify here
- **PrimerDesign 1.04**, a DOS-program to choose primer for PCR or oligonucleotide probes. See also the PrimerDesign Welcome Page.
- PC-Rare, a new software by R. Griffais, which uses a rare octamer at the 3' termini of the primer. This powerful (but user friendly) software is available for Macintosh and Windows environment.
- Primer Design, a Java applet by Luca Ida Giovanni TOLDO.
- CODEHOP, PCR primers designed from protein multiple sequence alignments (Local mirror site at WIS).
- **Primer3**, an online service to pick PCR primers from nucleotide sequence. (Local mirror site at WIS).
- NetPrimer (PREMIER Biosoft International). NetPrimer combines the latest
 primer design algorithms with a web-based interface allowing the user to analyze
 primers over the internet

1.12 Primer Design Considerations:

One of the single most important factors in successful automated **DNA** sequencing is proper primer design. It is important that a primer has the following characteristics:

- Primers should be at least 18-20 nucleotides in length to minimize the chance of encountering problems with a secondary hybridization site on vector or insert.
- 2. Primers with long runs of a single base should generally be avoided. It is especially to avoid **3** or more G's or C's in a row.
- For cycle sequencing, primers with melting temperatures above 55°C are generally produce better results than primers with lower melting temperatures.

- 4. Primers should have a G/C content between 40 and 60 percent. For primers GC content of less than 50%, it may be necessary to extend the primer sequence beyond 18 bases to keep the melting temperature above the recommended lower limit of 55°C.
- Primers should be "stickier" on their 5' ends than on their 3' ends. A "sticky" 3' end as indicated by a high G/C content could potentially anneal at multiple sites on the template.
- 6. "G" or "C" is desirable at the 3' end.
- 7. Primers should not contain complementary (palindromes) within themselves, which is they should not form hairpins. If this state exists a primer will fold back on itself and result in an unproductive priming event, which decreases the overall signal obtained.
- Primers should not contain sequences of nucleotides that would allow one primer molecule to anneal to it self or to the other primer used in a PCR reaction (primer dimer formation).
- If possible, run a computer search against the vector and insert DNA sequences to verify that the primers, and especially the 8-10 bases of its 3' end, are unique.
- 10. Do not design degenerate primers. Do not request inosine in sequencing primers. They either do not work or give poor cycle sequencing results.

1.13 Features Of Primer Design:

- Creating of new primer pairs.
- Creating one suitable primer to a given primer.
- Finding of repeats within a sequence.
- Finding of unique sequences within a sequence.
- Handling of sequences up to 32,000bp.

1.14 Limits Of Primer Design:

• The sequence length, which can be used for primer design, repeat and unique search is limited to 32,000bp.

- Maximal 16000 repeats can be found and sorted.
- Primer combinations can be explorated up to 6000 pairs.

1.15 Primer Design Parameters:

The most critical parameter for successful PCR is the design of primer. The primer sequence determines several things such as the length of the product, it's melting temperature and ultimately the yield. A poorly designed primer can result in little or no product due to non-specific amplification or primer dimer formation, which can be competitive enough to suppress to product formation. The main parameters for primer design are:

1.15.1 primer length

Specificity, the temperature and the time of annealing are at least partly dependent on primer length. Length of primer is critical parameter in primer designing. It is also proportional to annealing efficiency. For general studies primers of typically 17-34 nucleotides in length are the best primers, <16 nucleotides generally anneal non-target DNA sequences. Longer primers are required when a perfect complementary sequence to the entire template is not found.

1.15.2 primer sequence

- a) 3' self/cross complementarity 3 or more bases of complementarity at the 3' end can give a strong primer dimer and reduce PCR efficiency.
- b) Runs of single bases avoid long runs of single bases (i.e.5 or more).
- c) Internal self-complementarity We should check and avoid sequences of internal self-complementarity otherwise these regions can form hairpin loops.
- d) Internal sequence structure Primer stability is affected by nearest neighbour interactions between bases. Consequently it is better to have Gs and Cs together in pairs rather than dispersed within As and Ts.

1.15.3 GC content:

Melting temperature and annealing temperature are strictly dependent on each other. GC is an important characteristic of DNA and provides information about the strength of annealing (GC have hydrogen bonds between them). The base composition of primer should be between 45% and 55% of GC. The primer sequence must be chosen such that there is no poly G or poly C stretches that can promote non-specific annealing. Poly A, poly T stretches also to be avoided as these will breath and open up stretches of primer template complex.

1.15.4 Melting temperature:

a) Matching the primer for melting temperature – the stability of a primertemplate DNA duplex can be measured by its melting temperature(T_m). The T_m of any duplex of DNA is defined as the temperature at which 50% is dissociated as ssDNA. T_m is affected by sequence, length, p¹¹, salt concentration etc. The T_m of any primer pair (or multiplexed set) should be reasonably well matched otherwise amplification efficiency may be compromised.

There are many formulas used for calculating T_m the most accurate account of nearest neighbour interactions and are best left to computer programs.

 $T_m(^0C) = 2(A+T) + 4(G+C)$

This formula becomes inaccurate for longer primers.

Alternatively another more versatile formula is as follows: -

 $T_m(^0C) = 69.3 + 0.41(\%G+C) - (650/L)$

%G + C= %age GC content of the primer

L = length of primer (bp)

After calculating the T_ms of a primer pair they differ by too much it is best to add further bases to the less stable of the pair from the 5' end.

b) Choosing an annealing temperature for a primer set – the first time a pair of primers is used it is best to attempt an annealing temperature 5^0 C below the average T_m for the primer pair. Annealing temperature is then adjusted depending on the preliminary results or alternatively a Mg²⁺ titration can be useful especially where the product size is large or the GC content is high.

1.15.5 Secondary structure formation:

An important factor to consider when designing a primer is the presence of secondary structures. Presence of secondary structures greatly reduces the number of primer molecules available for the reaction. The presence of hairpin loops reduces the efficiency by limiting the ability to bind to the target site. The single stranded nucleic acid sequences may have secondary structure due to the presence of complementary sequences in the primer sequence. Secondary structure formation occurs mainly because of repetitive sequence.

1.15.6 Specificity:

Primer specificity is partly dependent on primer length. Primers must be chosen so that they have a unique sequence with in the template DNA that is to be amplified. A primer designed with a high repetitive sequence will result a smear when amplifying genomic DNA. How ever the same primer may give a single band if clone from a genomic library is amplified. Because Taq DNA Polymerase is active over a broad range of temperature, primer extension will occur at the lower temperatures of annealing. If the temperature is too low, non-specific priming may occur which can be extended by the Polymerase if there is a short homology at the 3' end. In general, a melting temperature of 55° C to 72° C gives the best results.

1.15.7 Primer ends:

The 3' terminal position in PCR primers is essential for the control of mis-priming. Another variable to look at is the inclusion of a G or C residue at the 3' end of primer. This will helps to improve the efficiency of the reaction by minimizing local unwind of DNA double helix. This occurs because of AT rich sequence.

1.16 PRIMER3

To design primers for a region of interest, Genotator i.e. *Primer3* is used. The development of *Primer3* and the Primer3 WWW interface were funded by Howard Hughes Medical institute and by the National Institutes of Health, National Iluman Genome Research Institute, under grants ROI – HG00257 and P50-IIG0098.

Primer 3 started as a reimplementation of *Primer .5* as software component; the design of *Primer 3* draws heavily on the design of *Primer .5* and Primer v2 and WWW interface designed by Richard Resnick for *Primer v2*.

Primer 3 is a computer program that suggests PCR primers for a variety of applications.

- a) To create STS (sequence tagged sites).
- b) To amplify sequences for single nucleotide polymorphism discovery.
- c) To select single primers for sequencing reactions.
- d) Do design oligo nucleotide hybridization probes.

1.17 Procedure For Primer Design By Using Primer3:

- · Select the query sequence and Id, which has the Tandem repeats in i
- Paste source sequence in FASTA format.
- Paste the sequence Id number in the sequence Id blank.
- Then put the tandem repeats in brackets [], which are present in the source sequence.
- Then adjust parameters according to our requirement.
- Then click the Pick Primers option.
- Then it shows the results as output.

1.18 Primer3 Input Parameters:

1.18.1 Source Sequence

The sequence from which to select primers or hybridization oligos.

1.18.2 Sequence Id

An identifier that is reproduced in the output to enable the user to identify the chosen primers.

1.18.3Targets

If one or more Targets are specified then a legal primer pair must flank at least one of them. A Target might be a simple sequence repeat site (for example a CA repeat) or a single-base-pair polymorphism.

1.18.4 Excluded Regions

Primer oligos may not overlap any region specified in this tag. The associated value must be a space-separated list of

Start, length

Pairs where *start* is the index of the first base of the excluded region, and *length* is its length.

E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. e.g. ...ATCT<CCCC>TCAT. Forbids primers in the central CCCC.

1.18.5 Product Size

Minimum, Optimum, and Maximum lengths (in bases) of the PCR product. Primer3 will not generate primers with products shorter than Min or longer than Max, and with default arguments Primer3 will attempt to pick primers producing products close to the Optimum length,

1.18.6 Number To Return

The maximum number of primer pairs to return. Primer pairs returned are sorted by their "quality", in other words by the value of the objective function (where a lower number indicates a better primer pair). Setting this parameter to a large value will increase running time.

1.18.7 Max 3' Stability

The maximum stability for the five 3' bases of a left or right primer. Bigger numbers mean more stable 3' ends.

1.18.8 Max Mispriming

The maximum allowed weighted similarity with any sequence in Mispriming Library. Default is 12.

1.18.9 Pair Max Mispriming

The maximum allowed sum of similarities of a primer pair (one similarity for each primer) with any single sequence in Mispriming Library. Default is 24

1.18.10 Primer Size

Minimum, Optimum, and Maximum lengths (in bases) of a primer oligo. Primer3 will not pick primers shorter than Min or longer than Max, and with default arguments will attempt to pick primers close with size close to Opt. Min cannot be smaller than 1. Max cannot be larger than 36. (This limit is governed by maximum oligo size for which melting-temperature calculations are valid.) Min cannot be greater than Max.

1.18.11 Primer T_m

Minimum, Optimum, and Maximum melting temperatures (Celsius) for a primer oligo. Primer3 will not pick oligos with temperatures smaller than Min or larger than Max, and with default conditions will try to pick primers with melting temperatures close to Opt.

1.18.12 Maximum T_m Difference

Maximum acceptable (unsigned) difference between the melting temperatures of the left and right primers.

1.18.13 Product Tm

The minimum, optimum, and maximum melting temperature of the amplicon. Primer3 will not pick a product with melting temperature less than min or greater than max.

 $T_m = 81.5 + 16.6(\log_{10}([Na+])) + .41*(\%GC) - 600/length,$

where [Na+] is the molar sodium concentration, (%GC) is the percent of Gs and Cs in the sequence, and length is the length of the sequence.

1.18.14 Primer GC%

Minimum, Optimum, and Maximum percentage of Gs and Cs in any primer.

1.18.15 Max Complementarity:

The maximum allowable local alignment score when testing a single primer for (local) self-complementarity and the maximum allowable local alignment score when testing for complementarity between left and right primers. For example, the alignment

```
5' ATCGNA 3'
||||
3' TA-CGT 5'
```

is allowed (and yields a score of 1.75), but the alignment

5' ATCCGNA 3'

```
11
```

3' TA--CGT 5'

is not considered. Scores are non-negative, and a score of 0.00 indicates that there is no reasonable local alignment between two oligos.

1.18.16 Max 3' Complementarity:

The maximum allowable 3'-anchored global alignment score when testing a single primer for self-complementarity, and the maximum allowable 3'-anchored global alignment score when testing for complementarity between left and right primers. The 3'-anchored global alignment score is taken to predict the likelihood of PCR-priming primer-dimers, for example

```
5' ATGCCCTAGCTTCCGGATG 3'
||| |||||
3' AAGTCCTACATTTAGCCTAGT 5'
```

or

```
5° AGGCTATGGGCCTCGCGA 3'
```

11111

3' AGCGCTCCGGGTATCGGA 5'

The scoring system is as for the Max Complementarity argument. In the examples above the scores are 7.00 and 6.00 respectively. Scores are non-negative, and a score of 0.00 indicates that there is no reasonable 3'-anchored global alignment between two oligos. In order to estimate 3'-anchored global alignments for candidate primers and primer pairs, Primer assumes that the sequence from which to choose primers is presented 5'->3'. It is nonsensical to provide a larger value for this parameter than for the Maximum (local) Complementarity parameter because the score of a local alignment will always be at least as great as the score of a global alignment.

1.18.17 Max Poly-X:

The maximum allowable length of a mononucleotide repeat, for example AAAAAA.

1.18.18 Included Region:

A sub-region of the given sequence in which to pick primers. For example, often the first dozen or so bases of a sequence are vector, and should be excluded from consideration. The value for this parameter has the form

start,length

where *start* is the index of the first base to consider, and *length* is the number of subsequent bases in the primer-picking region.

1.18.19 Start Codon Position:

This parameter should be considered EXPERIMENTAL at this point. some erroneous inputs might cause an error in Primer3. Index of the first base of a start codon. This parameter allows Primer3 to select primer pairs to create in-frame amplicons.

1.18.20 Mispriming Library:

This selection indicates what mispriming library (if any) Primer3 should use to screen for interspersed repeats or for other sequence to avoid as a location for primers.

1.18.21 CG Clamp:

Require the specified number of consecutive Gs and Cs at the 3' end of both the left and right primer. (This parameter has no effect on the hybridization oligo if one is requested.)

1.18.22 Salt Concentration:

The millimolar concentration of salt (usually KCl) in the PCR. Primer3 uses this argument to calculate oligo melting temperatures.

23

BR 63107

1.18.23 Annealing Oligo Concentration:

The nanomolar concentration of annealing oligos in the PCR. Primer3 uses this argument to calculate oligo melting temperatures.

1.18.24 Max Ns Accepted:

Maximum number of unknown bases (N) allowable in any primer.

1.18.25 Liberal Base:

This parameter provides a quick way to get Primer3 to accept JUB / JUPAC codes for ambiguous bases (i.e. by changing all unrecognized bases to N.

1.18.26 First Base Index:

The index of the first base in the input sequence. For input and output using 1-based indexing (such as that used in GenBank and to which many users are accustomed) set this parameter to 1. For input and output using 0-based indexing set this parameter to 0. (This parameter also affects the indexes in the contents of the files produced when the primer file flag is set.) In the WWW interface this parameter defaults to 1.

1.18.27 Inside Target Penalty:

Non-default values valid only for sequences with 0 or 1 target regions. If the primer is part of a pair that spans a target and overlaps the target, then multiply this value times the number of nucleotide positions by which the primer overlaps the (unique) target to get the 'position penalty'. The effect of this parameter is to allow Primer3 to include overlap with the target as a term in the objective function.

1.18.28 Outside Target Penalty:

Non-default values valid only for sequences with 0 or 1 target regions. If the primer is part of a pair that spans a target and does not overlap the target, then multiply this value times the number of nucleotide positions from the 3' end to the (unique) target to get the

'position penalty'. The effect of this parameter is to allow Primer3 to include nearness to the target as a term in the objective function.

1.18.29 Sequence Quality

A list of space separated integers. There must be exactly one integer for each base in the Source Sequence if this argument is non-empty. High numbers indicate high confidence in the base call at that position and low numbers indicate low confidence in the base call at that position.

1.18.30 Min Sequence Quality:

The minimum sequence quality (as specified by Sequence Quality) allowed within a primer.

1.18.31 Min 3' Sequence Quality:

The minimum sequence quality (as specified by Sequence Quality) allowed within the 3' pentamer of a primer.

1.18.32 Sequence Quality Range Min:

The minimum legal sequence quality (used for interpreting Min Sequence Quality and Min 3' Sequence Quality).

1.18.33 Sequence Quality Range Max:

The maximum legal sequence quality (used for interpreting Min Sequence Quality and Min 3' Sequence Quality).

1.18.34 Penalty Weights:

This section describes "penalty weights", which allow the user to modify the criteria that Primer3 uses to select the "best" primers. There are two classes of weights: for some parameters there is a 'Lt' (less than) and a 'Gt' (greater than) weight. These are the weights that Primer3 uses when the value is less or greater than (respectively) the specified optimum. The following parameters have both 'Lt' and 'Gt' weights:

- Product Size
- Primer Size
- Primer T_m
- Product T_m
- Primer GC%
- Hyb Oligo Size
- Hyb Oligo T_m
- Hyb Oligo GC%

For the remaining parameters the optimum is understood and the actual value can only vary in one direction from the optimum:

- Primer Self Complementarity
- Primer 3' Self Complementarity
- Primer #N's
- Primer Mispriming Similarity
- Primer Sequence Quality
- Primer 3' Sequence Quality
- Primer 3' Stability
- Hyb Oligo Self Complementarity
- Hyb Oligo 3' Self Complementarity
- Hyb Oligo Mispriming Similarity
- Hyb Oligo Sequence Quality
- Hyb Oligo 3' Sequence Quality

The following are weights are treated specially:

• Position Penalty Weight

- Determines the overall weight of the position penalty in calculating the penalty for a primer.
- Primer Weight
- Determines the weight of the 2 primer penalties in calculating the primer pair penalty.
- Hyb Oligo Weight

Determines the weight of the hyb oligo penalty in calculating the penalty of a primer pair plus hyb oligo.

The following govern the weight given to various parameters of primer pairs (or primer pairs plus hyb oligo).

- T_m difference
- Primer-Primer Complementarity
- Primer-Primer 3' Complementarity
- Primer Pair Mispriming Similarity

1.18.35 Hyb Oligos (Internal Oligos):

Parameters governing choice of internal oligos are analogous to the parameters governing choice of primer pairs.

Prove age antico manage of the Maximum			
Bin Sell Your Constant Sonis Unio	1	a	·
Address (P) http://www.concome.us.mit.edu/cur.tu	kallapati Hiterae	South Faunties Herry	
Addresse (at) http://www.gencome.uk.mit.edu/cgn/ba	nditerenter (transmitter)	P	<u></u>
Primer3	dire.)	- A RAMANE BRANDE E	电子外的分析 计定分进程
(new interface) pick primer: DNA sequence	s tron & courts		6.7.19 2.03.11.1.1.2.2.5.1.
Parte source sequence below trated as N numbers and undesiteble sequence (verte hittarr trapest threathr M	, ALUS, LINES,	g of ACGINSCUID	other letters Please N-out
p r Dich Dick	ization	D Pict	
primer probe		primer	
use (opile	na i	02 184	
left or primer upo		right primer	
below. eligo below.		balow (5'-	
		>3"	
والمنافع والمراجع والمراجع والمحاج		opposite strand).	
[Г	
Pick Primers Recet Form	string to ident	ify your output.	
	50.7 Femilia		nd the 2 bares
Serverer .	positions 50 a	and 31. Or mark the and 3. w.gArcti	OCCLEGAT
	ans that prime: g. 401,7 65,3 f	s must flank the ce orbids selection of	stral cocc.
RASINARS OF	• 7 beses start mark the	a must flank the co orbids selection of ing at 401 and the constant with 4 a by forkads primers	3 knowes at 68. and >: a.a.
Commonda Com	APCT CCCCCMECAL	forhads primers	in the central
PLAZARS SAA Nin: 100	ope : 200)	las: 1000	
HUMEN F. S. REFUTER: B MAN.	ati. Sunkaans	6.0	
MAN. MARGARMANN, 12.00 Dours. Pick Primers Reset Form	.Ман. Мірукатаца	1: 24.00	
General Primer Picking Co	aditions		
Prines Arenin: 13 Ope:	20 Hax: 27	-	
PAARAL IL Min: PP.O Opt:		Malal Tr. C) I'F + CONSEC.	1000
RESEARCH THE MENT Opt.:	MAR:	-	
FRENC SAL Min: DOD Opt:		-	
	注意最优化与这些上式之人。	5.00	
Mark, Attalan D Mark	125. 3 V 1523	F	
Asside Texas Casta Fenalizza	1945 <u>- Tiak 1995</u> Réf. Mil	Set Steads In St. Siles prov Langet	6128555777565577 61285577756
CLEAR BASE LUGARS P	1.4m;	George and the contract	
Rall Contractions South Con-	alland Glilles Mandeleni	50000 002 K3000 20 Chin 040 K300 20 Chin 040 K5 200 20 Chin	1. 1. 19 (1977) 1. 1971 - 1986 18 1971 - 1976 1. 1971 - 1976 1971 - 1976 1. 1971 - 1976
P LABREAL Brite C Show Enter	192213 I.N.S.S.		
Other Per-Sequence Inputs			
Agaluand	egion starting	ly pick primers in at position 20. Or	1050 (0101)
ELECTION CON	nd of the incl	at position 20. Or halfblich to mark the uded region: e.g. i	a Aic
Start Seden	TTCTCT AT L	he included region	is TTCTCT.
POSILION:			
Bequence Quality			

Segmence Quality	¥.			
Massana J Suda	Man Finel Socyae	<u>.</u>	- <u>Bergamer / Diabys Bangy</u> Julie	10 Sugarous Roder Laure por
				2266.
Dejective Panels	an Punaity Weigh	is for Pyles		
Sa La pa Ga				
<u>2055 (4 80 –</u> G) 2017 - Maintenne Anto				
Sel Constraints				
4 <u>27'</u> -:	8.9			
WART TRANS	a a			
Carpence Junio:	10.0			
Sad Seguence Outling				
issian inder	p a			
inst Englishing	R 0			
Ibjective Punctio	n Penalty Weight	s for Prime	a Pairs	
(12 that Date La 005				
bostar: Ter Lt PD	Gt. 0.9			
an Latercome	ae			
us Connectory	· · ·			
Conviencements	0.0			
no Mannana	0.0			
Trains Person Viewers				
fat silves Receive We Pick Primare 1	Reset Fam			
lyb Olige (Intern	ul Oligo) Per-Seq	nence Inpa	its.	
het sides Exclude a R	18004			
iyi Olige (inters	ul Oligo) General	Condition	•	
hit Chan See. Min	10 Opt 20 M	27		
	570 Opt 600 M			
ht Class GOM Min		80.0		
lato Otan Sedi Comp	ernentady 12.00	Hat: Class M	las E. Sell Complementation	12.00
the the	jo	Hyt. Chec M	10: 2 dy - X	5
in Line Madert Lab		HTS Over M	see Martist.	12.00
ht Ohes Mr. Seque				
ht Obeo Salt Course	· · ·	Hyb Chips D	NA Concentration	50.0
	Reset Form			
_		i far Hyb C	Miges (Internal Olige	4)
ht Olico Ten La Fi	•			
<u>ivi: Obser Size</u> Le fi	•			
late Disea GCM La 🗖 hete Otym Sell Courses	•			
tyb Olge #974	600			
lyb Oige Makeburg	80			
late Oligo Sequence (•			
Birt Dimen 1				

Pick Pilmons Report Form

1.19 Methods For Finding Tandem Repeats

All available nucleotide sequences of sorghum from the NCBI (<u>http://www.ncbi.nlm.nil/gov/entrez</u>) were downloaded in FASTA format on to a local database.

1.19.1 Steps Involved After Downloading The Sequence Of Sorghum From NCBI.

1.19.1.1 Fasta Format

The FASTA format looks like this,

gi|18063613|gb|BM324719.1|BM324719 PIC1 34 G05.b1 A002 Pathogen-infected compatible 1 (PIC1) Sorghum bicolor cDNA. mRNA sequence GCACGAGGCACACACACACACTGACGTACGATATCGAACACACTGACACAA CACAACCGCCGGGCGCCGGCCTTTGTCGTCAACGTCACCATTCCTTAACAC ACTCTCTACATACACGTCGCAGCCGGCCGGCAGCTCATCTTCTGTATTTAAAC GCGCTCCATCCATTCCGGTCAGGGCACGGAGTCGTCACTCCTAGCTCCGCGC GCACCGAGCTCCCACACACTCTCGCAGCAATGGCACGCGGCGGCTCTACCGC GGCTCCTCCACCCCTCCCGCCGCTGCTGCTGCTGCTGGTGGTAGTAGTGGTGT CCTCGGCCTTTTTCGCTCGCTGCCACGCCGCCGCCGCCGGGTGCACTGGTG TACGCCGGGTACGTGACGGTGGACGAGGCGCACGGCAGGAGGCTCTTCTACT ACCTGGTGGAGCCGAGCGTGACCCCGCCAAGGACCCCGTCGTGCTGTGGCTC AACGGCGGGCCTGGCTGCTCCAGCTTCGACGGC

1.19.1.2 Enter The Sequence For Finding Tandem Repeats.

Each of the sequence was searched for tandem repeats region and motif using the Tandem Repeat finder at <u>http://c3.biomath.mssm.edu/trf.advanced.submit.html</u>. This program was developed by the department of biomathematical sciences, Mount Sinai School of Medicine (<u>http://www.mssm.edu/school.html</u>). Advanced options were used to be find the tandem repeat region.

Parameters are set according the requirement. Alignment parameters (match, mismatch, indels) are set to (2,3,5). Minimum alignment score to report repeat is set to 30 and option period size to 10.

£Ε	ಕ್ಷರಗ	1.24	Fainter	PCCN:	10

Admin 💭 ma yoʻ sosmatri misuri ada hisaratincesi sasmitikari.



8

SUBMISSION OF SEQUENCE TO ADVANCE TANDEM REPEAT FINDER

1.19.1.3 Submit Sequence

By clicking the submit sequence button which is on the advanced tandem repeat finder we will get program analysis of DNA sequence on screen

r for far Vere farrates (see the)	
nåtten (Li läp Nei bomst mesma havg kost Mia og	51 4 K
Tandem Repeats Finder	250.02
The Tandem Repeats Finder Server.	
Please cite: 6. Remson, "Trandem repeats finder: a program to analyze DWA sequences," Nucleic Roids Research (1999) Vol. 27, No. 2, pp. 573-530.	
Click haze to:	
If you save the summary table and alignment files, use the defailt tam supplied by the browser to preserve the automatic cross-linking	

1.19.1.4 Tandem Repeats Report

Click on the tandem repeats reports button on the program to analyze DNA sequence. We

will get the summary table

```
for Eall View Favorites (1906 Hep)
                                                                              ÷1
Addeur (23 http://www.americanseria.com/eria/to/27/state27/state235/86/16/8/16/11/state
Tanden Repeats Finder Frontan vritten by:
                                   Gary bennon
                        Repartment of Geomethematical Sciences.
                          Mount Sinal School of Medicine
                                   Version 1.22
Please dite:
 G. Benson,
 "Tanden repeats finder: a program to analyze DNA sequences"
 Nuclear Acid Research (1996)
 Vol. 27, No. 2, pp. 573-580.
 Sequence: 123
 Parameteis: 2 3 5 80 10 10 10
 Length: 2324
 Tables: 1
 This is table 1 of 1
 Click on induces to view alignment
               Penoo
                      Copy Consensus Percent Percent Score A C IS & Entropy
     hodes.
                Size Mumber Size Matches adals
                                      37 0
                                                    应 求 許 主义
                      13.7
                 2
 The End!
```

SUMMARY TABLE

1.19.1.5 Summary Table

By clicking the column of indices of table explanation, which is present in summary table, we will get the alignment explanation.

kan Edit Jana Reported Jacob Haip

Allen Control close to mean exact the length of the State State and the State Stat

8°.

Tandem Perests Finder Success written by: Garn Betern Separtment of Signathematical Sciences Mount Sinei Sphool of Zedicine Vecesor 1.80 Sequence: 123 Parameters: 2 3 5 00 10 30 13 Ptest Re-81.8%, Pindel-0.100 taple pines 0,4,5,7 tuple distances 0, 29, 159, MARDISTRATE Sempth: 1114 Elefebrant: Arbibl, Croile, Orling, Dicity Found of itill drighted ditely final size() Indites: 1119--1159 Score: 61 Period gife: 3 Capynember: 1:17 Consenity size: 4 1109 CRTCACLOCG v + 1 I GAR CLE GAR GAR CAR CAR CEN GAR CAL THE GAR GAR CAR CAR CAR 11 FOR GETER PROPERTY Statigtors. Matches: 17, Erznatches: 1, Indefe: 0 3,97 0.00 0.00 Matches are distributes among these distances: 5 37 1.96 ACGTOCUME: AND.56, C:0.14, C:0.10, F:0.00 Consensus pattern (5 hc); 622

ALIGNMENT EXPLANATION TABLE

1.19.1.6 Alignment Explanation Table

Alignment explanation table explains the number of times the repeat is repeated, for Ex-In the above case the repeat CAA is repeated 9 times

1.19.1.7 Primer3

PRIMER3 program to used design the primers in sequences where repeat region were found. The sequence and its Id. Were pasted in the sequence text box in FASTA format. Tandem repeats regions were masked by placing them in brackets []. The parameters were set according to the requirement. The selected options for product size Min imum100,Optimum 200 ,maximum 400 . the primer anealing temperature Minimum 59.0,optimum 60.0.Maximum 61.0 degree celcius. Primer GC% should be Minimum 20.0 and Maximum 80.0.primer size should be minimum 18bp.Optimum 20bp maximum27bp.Then click the **pick primer** option.

	Ship Pathach Huma I want adayt gi Ginypranav(paraar 5 ju		-
Primer3		AND THE REAL PROPERTY	കരാം
		ALLAND BOOLSELLIGHT LUEAL	Ersen.
nate source sequences and the sequences of the second seco	ance below (5'->3', st unbers and blanks igno	ring of ACGIMECGIN other is red). FASTA formet ok. Fleese	Htter H-our
		Bo, etc.) or upp a Michelmung TCTAACCGTACAGACCATCACAGCCCAGCAGCAGCAG	
CTRCTRCTRCTRCTRCTRCT	BCARCAA]		
		The the anean and the the area and the second states and an area area and an area area area area area area area a	
TOCALCAGE AND AND THE THE	CCTRCERCERCERCERTUTCERCECE CCARTCERCERCEGOCTCTUDCEEEC	TO TRACTORCOCTOR FUEL ACTINGUE A TRACE A BACTON ACTORCOCATACTITICAAC ANCIAC A BETTACTITOODY	TTAATO
🖉 a ser en a ser en s	e . 1	9	
ick eft	Pick hybridization	Pàcht Làght	
orimer or	probe (internel	primer	
ife aft	oligo) er	use	
Finer	List 6	right primer	
HELOW.	oligo below.	belew (5°-	
		>3° 98	
		opposite strend).	
	t Form		
1 41 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A string to it	dentify your output.	
	¥ g 50.2 req	uires primers to surround the :	2 2-4-8
EAS MAS AL	Esawegee with	1 and 1: e.g ATCT[CCCCITC	AT
Eng and Ris 🚦	Esquere with means that pr. E.s. 401.7 68	1 and 51. Or mark the second solution is a second solution of the second solution of the second solution of the second solution solution of the second solution soluti soluti solution soluti so	AT
includes a	E.g. 401,7 68 the 7 bases s	50 and 51. Or mark the a_{12} a_{23} a_{23	sin at 6
includes a	E.g. 401,7 68 the 7 hases s Or mark the ATCT <cocc></cocc>	50 and 51. Or mort the start i [and]: e.gATCT[CCCCTC] imers must flank the central G .3 forbids selection of primers	rin ∎to ∖vi
Registers	8.g. 401,7 60 the 7 bases s Or mark the 	50 and 51. Or mark the only in 1 and 1: e.gATCT[CCCCTC] imetrs must flank the central C .3 forbids selection of primer tarting at 401 and the 3 bases with 4 and 2:	rin ∎to ∖vi
Rectauna Rectaunaa Rectauna Risa Mini	E.g. 401,7 64, the 7 bases s Or mark the 	50 and 51. Or mark the second provided and the second and the second and the second prime second and the second second and the second and	rin ∎to ∖vi
Sulausi Registari Registari Registari Selan Mini Namtor To Anturas Kin Mangistari	#.g. 401,7 64. the 7 bases at Or mark the y	50 and 51. Or mark the variable of and 51. Or mark that the variable of the second sec	rin ∎to ∖vi
Rochulai Rochanai Rachanai Rachanai Rachanai Rochanai Rochanai	#.g. 401,7 64. the 7 bases at Or mark the y	50 and 51. Or mark the variable of and 51. Or mark that the variable of the second sec	rin ∎to ∖vi
Suliy Js. 2 Regisenzi Regisenzi Regisenzi Norton To Asthuro San Marton To Asthuro San Marton San	#.g. 401,7 64. the 7 bases at Or mark the y	50 and 51. Or mark the variable of and 51. Or mark that the variable of the second sec	rin ∎to ∖vi
SALIS JE 2 REGISERI REGISERI RANTOR VO RELIVO FRANCISCO RELATION FRANCISCO RELATION	8.9. 401.7 64. 0.1.7 64.7 7 hases si 0.1.7 74.7 7 hases si 1.1.7 7 hases si<	10 and 51. Or mark the second line of the second se	rin ∎to ∖vi
ANGLAUSE REGISERI NARTOF FO GACAUTO NARTOF FO GACAUTO NAR MORT SAAAAL PAART SAAAL RAATA SAAAL RAATA SAAAL	S.g. 401,7 66, Steer at 0,7 76, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10	50 and 51. Or mark the operation of and 51. Or mark the operation of a second state	rin ∎to ∖vi
NALIAUNA REGISERZI REGISER	S.g. 401.7 66. S.g. 401	50 and 51. Or mark the operation of and 51. Or mark the operation of a second state	rin ∎to ∖vi
Antibulan Kardanan Kardanan Kardanan Kardan Kardanan Kardanan Kardanan Kardanan Kardan	S.g. 401,7 66, S.g. 401,7 66, Or mark tha Or mark tha Ort. 200 DO D	1 and 51. Or mark the one-of- 1 and 1: e.g	rin ∎to ∖vi
SALJAURA REGISTRI REGISTRI RAN DEL VOI ARCHIVE PIC PUNES REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL REAL	S.g. 401.7 66. S.g. 401	1 and 51. Or mark the one-of- 1 and 1: e.g ANTERCECTIC 2 forbids salection of primer tarting at 402 and the 3 hansa Max: 400 Max: 400 Max	rin ∎to ∖vi
Chaile Unit Resolution State Mines Resolution State Mines	B.g. 401,7 66 Cor anth the 7 Or mark the 7 Cor anth the 7 <t< td=""><td>1 and 51. Or mark the one-of- 1 and 1: e.g ANTERCECTIC 2 forbids salection of primer tarting at 402 and the 3 hansa Max: 400 Max: 400 Max</td><td>rin ∎to ∖vi</td></t<>	1 and 51. Or mark the one-of- 1 and 1: e.g ANTERCECTIC 2 forbids salection of primer tarting at 402 and the 3 hansa Max: 400 Max: 400 Max	rin ∎to ∖vi
Antibular Register: Stor Hins New Yor No. Reductor New Yor You Reductor New York Provide Property New Piece Reserve. The	S.g. 401,7 66, G. 7 Mark 716, 7 G. 7 Mark 71, 7 G. 7	S0 and S1. Or mark the overlap 1 and 1: e.g	s in ut d tunt
RALINDER REFERENCE REFEREN	8. ч. 401.7 С.6. • • • • • • • • • • • • • • • • • • •	30 and 51. Or mark the control 10 and 10 and 10 and 10 and 20 and 2	s in ut d tunt
RALINDER REFERENCE REFEREN	8. ч. 401.7 С.6. • • • • • • • • • • • • • • • • • • •	50 and 51. Or mark the one-information of the state o	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
Antinung Rechtenze Neutrone Stone Hins Neutron Volkelauro Hick Propers Bes Benared Primers Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter	Sup 401,7 66 S	50 and 51. Or mark the one-information of the state o	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
Antinung Rechtenze Neutrone Stone Hins Neutron Volkelauro Hick Propers Bes Benared Primers Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter Rechter	В. ч. 401.7 С.6. • • • • • • • • • • • • • • • • • • •	S0 and S1. Or mark the owner, it and it or, it and it owner, it and the owner, it and the owner, it is the owner, it is and the owner, it is an owner, it is	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
CALLADAS REGISTIONS RE	B	50 and 51. Or mark the one-of- land 1: e.g	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
Antibular Rectanza Re	B	50 and 51. Or mark the one-of- land 1: e.g	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
CALLAURA Registration State Mins: Registration State Registration Registration State Registration Registratio	В. ч. 401.7 С.8.	50 and 51. Or mark the one-of- land 1: e.g	5. 19 19 5. 5 5 19 5. 5 5 19 5. 5 5 19 5 19 5 5 10 19 5 5 19 5 5 10 1
CALLER SATE MINE ELANGLICA MINE ELANGLICA MINE ELANGLICASE MINE MAN ALASE CALANGLICASE MINE MAN ALASE MAN ALASE MAN ALASE ALASE MANY, REAMAN ALASE CONSALLEN ALASE CONSALLEN ALASE LINEARA ALASE LINEARA ALASE LINEARA ALASE LINEARA ALASE LINEARA ALASE	B. G. 401,7 660,7 B. G. 70,7 660,7 C. 70,70,7 D. 70,7 D. 200,7 D. 200,7 <td>50 and 51. Or mark the one-of- land 1: e.g</td> <td>s in at 6 - U. Cant M Salah Adata - Salah Salah Salah</td>	50 and 51. Or mark the one-of- land 1: e.g	s in at 6 - U. Cant M Salah Adata - Salah Salah Salah

Perstarn.

Semanoe Quality

Sequence Quality

Pf 2 33 Min Ros . Bushapper GRANTE FARME LANALANN SAMAS MOO HANA 6 Seguence Stautors CWALNEY: Objective Function Penalty Weights for Primers La La: Do Ga: Do LIT LE FO GE: FO Set Le: Do Seit Complementaries 0.0 3' Sylf Complementarity 00 2011.02 0.0 MASDELWAND 6.0 SHAMPING LONDARY 60 Red Requeste Guality 50 PARALLANT. PARALLY 0.0 Gnd stability 6.0 Objective Function Penalty Weights for Primer Pairs Fraduct Size Lt: DOS Gt: DOS Endors In Lt: DO Gt: DO The South Strategies 0.0 ADV 空中教育委員師會的教育的主义文字 6.0 2' COMPLEMENTARITY 60 60 SALE MARSELLARD HEAMAR FORMELLY MALANT 6.0 Hit Nils Penalty Weacht Do Pick Primers Reset Form Byb Oligo (Internal Oligo) Per-Sequence Imputs AND GLIGS EXCLUSED BUGLONS Byb Oligo (Internal Oligo) General Conditions Opt 20 Syb Olige Size: Min D8 HAX 27 HAL SLAG IN: Nin 590 Opt \$0.0 HAX 610 Hyb CLIDE SALL Min: DOD Opt: [Max: 80.0 STAR SEAM FAILS HIG PLANE MAR I' SHAS 12.00 12.00 Complementality; 6 MAX BUSS HYE OLIGO MAN POLY X: 5 HUN BLIGS MARNYD LIDNATY: MONE . HYD PILLER MAX MARMAN 62.00 HTTE MAAGE MAD REAMENCE QUALITY: 6 <u>888 . 024 ac</u> 3412 500 HVL DILLOS MIN CONCENT PARTYON 50.0 Concentrations Pick Primers Reset Form Objective Function Penalty Weights for Hyb Oligos (Internal Oligos) HYP OLLING THE LA: P.O GE: D.O Byb Oligo size Lt: 10 Gt: 10 HYD DIA 20 GET LE: D.O GE: D.O Hyb Oligo Solf Complementarity Do

Hyb. S.Lize Hi's Do Byb. S.Lize Mishwhing Do Exh. S.Lize Mashwhing Do Exh. S.Lize Magyers Statisty Do Mick Phimers Reset Form

1.19.1.7 Out Put Of Primer3

Primer3 Output

No mispriming library specified Using 1-based sequence positions OLIGO 55.00 4.00 3.00 GGCCCATCTAACCGTACAGA 1088 20 59.96 LEFT PRIMER 59.50 45.00 /.00 1.00 TGCAACTGTTGTTGTTGCTG 1285 20 RIGHT PRIMER SEQUENCE SIZE: 2124 INCLUDED REGION SIZE: 2124 PRODUCT SIZE: 198. PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 2.00 TARGETS (start, len)*: 1131,27 1 GAATTCGAGCTCGGTACCCAGACCTCCCAACCCATGCTCGCCACGTTTGTTAGGCCAAGG 121 TTGCCTCACGTGAATGTCAAGATTGTTTCACCATATTATTATGACAGAGGACTTGATAA 181 TTTTTTTCTTGTAATCAAAGTTTAAATTAAAACTTTGTCAAATTTACACAATACTAAAGCA 241 ATTGTAATCAAGAACACAGGGAGTGCCCTGTGATAATAAGCTATAAGCATTCTAATTTGTAC 301 ATTCTATTTGTGTGCATACATCTGTACATACTGGGATTTCAATTTGTCTTGATCTTGTAG 361 CATTTTTCAATCATTGATGAACAACTTCATCTAACTACGTTGCAAGACAAATAGTACAGT 421 ACTACAACAAAGTCCTTTGATAAAGGCTTTGATATACATGAGCAAGTCATAACTTTACTT 481 GCACATCATGTCTGTAAAGTGGAACATTTCTGATGTGGCTAAGGCTATAACATGTGTAAA 541 GGTGAAGTGATGTCACTCCTCATTTATCGAAAAGTTCCAATAGAAAATGACAACTTTTCC 601 TTGTAAGTAGTGGAAATTGTCTTTCCTACACAGACCATATAATCCAATAAAATTGATAAC 661 TAAATGTCAAAATCGACTAGGTGCCATGTCATCTATAGTTTATCTGTTGTTCGCAAAAGC 721 CAANATCTAAACAGATATCTATGAGCTCTCACTCATATAAATAGGCCCCCAAATCAGTAGT 781 TAAACCATCGCCCATAACATTGAGAGGAATTAGAAAAATACCAAGTGAAACGAACTAGCA

841	ACGTCCTTCCAACAATGGCTACCAAGATATTCGCCCTCCTTGCGCTCCATGCTCTTTAG
901	TGAGCGGTACAACTGCGGCCATTATTCCACAGTGCTCACTTGCTCCTAATGCTATTATTC
961	CACAGTTCATCCCACCTGTTACTGCTTTAGGGAATGAACACCTAGCTGTGCAGGCCTATC
1021	CTGGACAGCAGGTGCTTTCGCCGAGCATCTTACAACAACCAATTGCCCAATTGCAGCAGC
1081	AATCCTTGGCCCATCTAACCGTACAGAGCATCACAGCGCAGCAGCAGCAGCAACAACAAC >>>>>>>>>>
1141	AACAACAACAACAACAGTTCTTGTCATCACTCAGTGCACTAGCCGTGGCGAACCAGG
1201	CCGCCTACTTGCAACAACAGCTGCTTACATCCAACCCACATTCTCTGGCTAATGCTGCTG
1261	CATACCAGCAACAACAACAGTTGCAACTAGCCATGGCGAATCCAACCGCCTACGTACAAC <<<<<<<
1321	AACAACTGCTTCTATCCAACCGCACAGGCTGCCACCAATGCCGCTACATACCTACAACAAC
1381	AACAATTTCAACAGATCTTACCGGCGCTCAGTCAACTACGCATGGCAAACCCTACCGCTT
1441	ACTTGCAACAGCAACAGTTGCTTCCAATCAACCAACTGGCTCTGGCAAACACTGACGCAT
1501	ACTTGCAACAACAACAGCTGCTTCCGGTTAATCCACTGGTAGTTGCCAATCCATTAGTTG
1561	CCGCCTTCCTACAGCAACAACAATTGTCGTCATTCAACCAGATATCTTTGGTTAACCCTG
1621	CCTTGTCGTGGCAGCAACCCATCATTGGTGCGTGCCATCTTCTAGATTACATATGAGTTGT
1681	TGFACTTGATAACAAAGGTCTCATACCGGCATGGGCAACTTFCCTAAAATAATCAATAA
1741	TTGGTTGAGATTTATTTGTCTTCGTAATTATCTTCTTCATATATGCGATTGAAAACA
1801	TCACATCATAATTAAAGACACATGCTTGGTTAATTTGTGGACAATAACATAAATACTTCAT
1861	CAATCTAAAGATGTGCCTGCCCGACCTGAATATTCTAACTTGGGTGTGTGT
1921	CGATAGATGCACTATTGGAATCGCGCGCTTTGCCTAGTGTCTAGGGCACTCGGCAAACGG
1981	TAAAAAAAAAACACTCGGCAAACCGTTTGCTGGGTGCCACACTCGGAAAGCGGAGACAGCAA

2041 CAAAAATAACGGCAAAGCAAAACATAGTCAAATCAGATGCATGTATGGGGATCCTCTAGA

2101 GTCGACCTGCAGGCATGCAAGCTT

```
KEYS (in order of precedence):
****** target.
>>>>> left primer
<<<<<< right primer</pre>
```

ADDITIONAL OLIGOS

- 1 LEFT PRIMER 1081 20 60.15 50.00 5.00 0.00 AATCCTTGGCCCATCTAACC RIGHT PRIMER 1285 20 59.50 45.00 7.00 1.00 TGCAACTGTTGTTGTGCG PRODUCT SIZE: 205, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00
- 3 LEFT PRIMER 1001 20 60.00 60.00 6.00 2.00 CCTAGCTGTGCAGGCCTATC RIGHT PRIMER 1188 20 59.06 55.00 8.00 3.00 CGGCTATCCACTGAGTGAT PRODUCT SIZE: 186, PAIR NAY COMPL: 5.00, PAIR 3' COMPL: 3.00
- 4 LEFT PRIMER 1002 20 60.00 60.00 6.00 2.00 CTAGCTGTGCAGGCCTATCC RIGHT PRIMER 1188 20 59.06 55.00 8.00 3.00 CGGCTAGTGCACTGAGTGAT PRODUCT SIZE: 187, PATR ANY COMPL: 5.00, PATR 3' COMPL: 3.00

Statis	stics con sid ered	too many Ns	in tar get	in excl reg	bad GC::	no GC clamp	tm too low	tm too high		high 3' compl	poly X	high end stab
ok Left 771	10645	0	0	U	448	0	5809	3518	5	22	14	58
Right.	8773	0	0	0	142	0	3498	4419	11	24	24	78
Pair S consid	Stats: dered 11 :3 relea	,	•	table p	produ	et size	10518	8, hig	h end (compl 1	17, ok	: 539

(primer3 www results.cgi v 0.2)

1.20 Results

All available (approximately 110055) nucleotide sequences were analyzed for the presence of tandem repeats up to 80bp (maximum) length of repeat motif and no penalty gaps or indels were allowed. Although such methods may not detect some repeats containing insertions or deletions, it was considered that such a conservative approach might not be affecting the results very dramatically. The summary of the total number of potential SSRs in each of the nucleotide length group is presented in table 1. As descried in the above procedure the nucleotide sequences are allowed to find tandem repeats by submitting them to Tandem repeat finder. Possible primers were found to these sequences which had repeat regions using Primer3. All the sequences with tandem repeats do not yield primers either due the sequence being short and less than minimum product size that is prescribed in the loss as a flanking region. The results of Primer3 output were collated and entered into a database table, with right primer sequence, left primer temperature, right primer temperature and total sequence as shown in Table 2.

Of the total 9232 a little over (6232) were from tri-nucleotide groups.the tetra-nucleotide groups that potentially form the polypeptide repeats were the next largest groups followed by di-nucleotide and penta-nulceotide. The SSRs summary by various nucleotide groups for a minimum repeat size length of 10bp and the units that are part of the group, are presented in Table3. The largest group is the tri-nucleotide group 'CCG' consisting of 3312 accessions. Of these, the largest sub-group were from the 'GCG' and 'GCC' group which both code for alanine. In the di-nucleotide repeats the most common repeat was the 'AG' group which translated to di-oligopeptide of Arginine and Glutamine.

 Table 1 . Summary of the repeat motif by nucleotide unit length of more than 10bp

 repeat length

Number of SSRs
962
6232
1468
569
9232

Table 2

ld	Accessio	Details	Sequence	Repeat	Leftprimer	Rightprimer	Lefttem	Rightte.	Totalsequence
1	6674175	AW284211	AAGAGGAGCCAAAGAA	AAG(5)	GAGGAGCCA	ATTTTATTGC	59.13	60.63	112
			AAAGAAGAAGAAGGAG			CCCGCCTTT			
			AAGAAGGCCAAAAAGG		AAGA				
			CCAAGAAGAAGAAGAA						
			GTCCAAAAAGGATTGA						
			GGTGGTCAGGCAAAG] [
			GCGGGGCAATAAAATC						
			C		001700170				400
2	7554456	AW680658	GCACGAGGATCGATCA	CA(6)		CCGCCATTG	60.78	59.04	126
			GCGTGCATCCCTGACA		AGCGTGCAT	TTAATCCTTC			
			AAAAAAAGCACACACA			1			
			CACATACATACACCTA						
			CTAGTCTTCGTGGCTG						
1			GAAGGATTAAGCTCGA						
1			CGGTAGTCTCGCGG						
3	8087874	AW922049	TCAATGGGTGGACCAA	TGC(4)	ATGGGTGGA	CTGATGCTGT	59.95	59.36	105
ľ	000/0/4	10022040	ACACAATTGCAAACCC	100(4)	CCAAACACA	TGTTGCATAG			
			TCAAGCTGCTGCTGCT		AT	G			
			GCTCAACTGACCCTCT			-			
			TACAGCAGCAACGGAT	1	1				
			CCTGCCTATGCAACAA						
			CAGCATCAG						
4	8089381	AW923556	CGCTTTAATGTCGCGA	CTG(6)	CGCGAGATT		59.53	60.66	122
			GATTGATAGCAAGGTT			ATGCACACC			
			CCTGCTGCTGCTGCTG		G				
			CTGCTCTGGAACAGTG]			
1			GGAAACACTGCTGTCT						
			TGCTCCGTTGTGTCTA						
1			GGGCCGGGGTGTGCA						
			TCGCTGGATTT						

E	0047044	BE592838		TOODAW	100070000	OCOTOTOT	50 77	60.63	116
2	9647911	BE292838	GCACGAGGCTCGGCC AAAAATCTGCGGCGAG		AGGCTCGGC		59.77	00.63	110
					САААААТСТ	стсстстсст			
			CTCCGCTCCGATCCGA			G			
			TCCGATCCGACTGCTC						
			CCCACGCCGTTGAGG						
			CCGAGCGGGGGGTCCA						
			GGAGAGGAGAGGAGA						
			GGCGCGGC						105
6	11680518	BF588194	GCACGAGGGGATTGCT	GAGGA(4)	CACGAGGGG		59.17	60.16	125
			TTCCCCACCCCACCTC		ATTGCTTTC	GCTGTCCTCT			
			CGTCTCCGTCTCTTGC			C			
1			CGACCGAAGCGACTCC		1				
			ACCTGAGGAGAGGAG						
			AGGAGAGGACAGAGG						
			AGAGGACAGCAAGCTA						
			CCGCCCCCGTCCCCG						
7	11920692	BF655560	GCACGAGGCTCTCTCT	GGAG(4)	GCACGAGGC		59.03	59.53	112
			стстстстстстстстс		TCTCTCTCTC	CCCTCTCTCT			
			TCTTCTCTCGAGAATG		Т	C			
			GTTGTGTAGGCGGCG						
			GGCCCAGGAGGGAGG						
			GAGGGAGGAGAGAGA						
			GAGGGGAGCTAGGGT						
			ттт						
8	12500452	BG049068	TACTACTAGTAGGACA	TGCT(4)	AGGACAGGA		59.13	60.81	128
			GGAGTGCTCCATTATT		GTGCTCCAT	GAGAGGGTT			
			AATTAATTATTCTATTA		TATT	С			
			TAAAGTGCTTGCTTGC						
			TTGCTGCTAATCCATG						
			CACCATTAGGAAGAGA						
			ACCCTCTCTCCGTCGT						
			GCATGCAGCTTGTAC						

112	120
60.47	59.93
90 [.] 08	60.03
AGCAGGAGA GGAAAGGAG GA	
CTTTGGTTTC AGCAGGAGA GTTTCCCTG GGAAAGGAG A A CCCCTG GAAAGGAG GA	
ATC(6)	AGCG(3)
9 13588582 BG559584 GAACTITGGTTTGTTT ATC(6) CCCTGATCATCATCATCATCATCA ATCATCATCATCATCATCATCAT CATGTGTCGAAGGGATTT CATGTCGCCAAGGGGATTT CCGTCCCCAGGGGGATTT CCGTCCTCCTTTCC TCTCCTGCTCTTCT	GCACGAGGCTCTAGC AGCG(3) GCCCAGCAGGCAGCTCTCG GCCCAGCAGGGGGGG AGCGTGCCGAGGGGG TGACAGCGGGGGGGG TGACAGCGGGGCTGCA AGTCGCCGGGGCTCGA AGTCGCCGGGGCTCGA CCTTCGTTGTCGAGAG TGAAGGAGAGGGG
BG559584	
13588582	10 14570584 Bi099002
თ	10

•

Table 3. Summary by nucleotide group and members of the group

Click on the group or the inidividual unit to retrieve data

Group	Group SSR Count	SSR Units in Group
AC	255	र्स्ट ट्रम् डा रह
AG	357	RG CI GA IC
AT	260	<u>AT TA</u>
ଥ	91	छ छ
AAC	39	AAC ACA CAA GIT IGI IIG
AAG	294	AAG AGA CIT GAA ICI IIC
AAT	29	AAT ATA ATT TAA TAT TTA
ACC	281	ACC CAC CCA GGI GIG IGG
ACG	1378	ACG AGC CAG CGT CTG GAC GCA GCT GTC TCG TGC
ACT	278	ACT AGT ATC ATG CAT CTA GAT GTA TAC TAG TCA TGA
AGG	621	AGG CCT CTC GAG GGA TCC
000	3312	<u>CCC</u>
AAAC	31	AAAC AACA ACAA GATI IGII IIGI IIIG
AAG	29	AAAG AAGA AGAA CITI GAAA ICIT IIICI IIIC
AAAT	20	AAAT AATA ATTI TAAA TATI TIAI TITA
AACC	27	
AACG	46	AACG AAGC ACGA AGCA CAAG CGAA CGTT CTTG GAAC GCAA GCTT GTTC ICGT IGCT ITCG

46

AACT	25	AACT AATC ACTA AGTT ATCA ATTG CAAT CTAA GATT GTTA TAAC TAGT TCAA IGAT TTAG TIGA
AAGG	29	AAGG AGGA CCTT CTTC GAAG GGAA TCCT TTCC
<u>AAGT</u>	15	AAGT AATG ACTT AGTA ATGA ATTC CATT CITA GAAT GTAA TAAG TACT TCAT IGAA TTAC TICA
<u>AATT</u>	7	AATT ATTA TAAT TTAA
<u>ACAG</u>	7	ACAG AGAC CAGA CTGT GACA GTCT ICTG IGTC
<u>ACAT</u>	77	ACAT ATAC ATGT CATA GTAT TACA TATG TGTA
ACCC	12	ACCC CACC CCAC CCCA GGGT GGTG GTGG TGGG
<u>ACCT</u>	108	ACCI AGGT ATCC ATGG CATC CCAT CCTA CTAC GATG GGAT GGTA GTAG TACC TAGG ICCA IGGA
<u>ACGC</u>	97	ACGC CACG CGCA CGTG GCAC GCGT GTGC IGCG GCCA GCTG GGTC IGGC ACCG AGCC CAGC CCAG CCGA CGGT GACC
ACGG	84	ACGG AGGC CAGG CCGT CCTG CGGA CGTC CTGC GACG GCAG GCCT GGAC GGCA GICC ICCG IGCC
<u>ACGT</u>	181	ACGT ATGC CATG CGTA GCAT GTAC TACG TGCA AGTC CAGT GTCA TCAG TGAC
<u>ACTC</u>	71	ACTC AGTG CACT CTCA GAGT GTGA TCAC TGAG
<u>AGAT</u>	194	AGAT ATAG ATCI CTAT GATA TAGA TATC TCTA
<u>AGCG</u>	54	AGCG CGAG CGCT CTCG GAGC GCGA GCTC TCGC
<u>AGCT</u>	228	AGCI ATCG CGAT CTAG GATC GCTA TAGC TCGA
AGGG	52	AGGG CCCT CCTC GAGG GGAG GGGA TCCC
<u>CCCG</u>	74	<u> </u>
AAAAG	69	AAAAG AAACA AAATG AAGAA ATACA ATTAC CATAT GITTI TAATG ICATT IGTAA IGTTA IGTTI TICTI TITCI TITIGI TITTG TITTC ICTTI

AAAAT	12	AAAAT ATAAA TAAAA ITATI
<u>AAAGC</u>	155	AAAGC AAGCA AAGCT ACATC ACTGT AGAGA AGCTA AGGAA AGTAC AGTGT ATGGA ATTCG CAGAT CATCT CATLG CCAAT CTAGA CICAT CIGAT CTTCT GAATC GAGAA GAGTA GATCA GATCT GATIG GCAAA GGTTI GIGTA GTGTI GTIGT GTTIC TACTG TAGCA TAGGA TAGTC ICAGA TCCAI TCCTA TCGAA TCTTC TGATC IGCAA IGCTI IGTAC IGTIG TTCAG ITCTC TIGCA TIGCT IIGIG TTTCG IGATI
AACGG	227	AACGG ACAGG ACCGA AGAGC AGAGG AGCAG AGCGA AGCIG AGGGA AGGIG AGIGG AGIGG AGIGG AICCC AICGG AIGCG CACCI CACGC CAGCI CAGIG CCACA CCCII CCIAC CCIGA CCIIC CGAIG CGAIG CGGII CGICI CGIGI CIAGC CICCI CICGI CICIC CICIG CIGAG CIGCI AGAGA GAGAG GAGGA GAGGA GAGGI GAICG GCACA GCACI GCGII GCICA GCICI GCIGG GGAGA GGGAI GGIGA GIGCA GIGIG ICCGA ICCGI ICGGA ICGGI ICICC IGCGI IGCIC IGGGI IGGIG IICCC
ACCGC	91	ACCGC AGGCC CACCC CAGGG CAGGC CCCCT CCCGT CCCTG CCGTC CCTGC CGCCT CGCGA CGGAG CGGCA CGGCT CTCCC CTCCG CTGCC CTGCC CTGCG GAGGG GCCGT GCGGA GCTCC GCTGC GGGAC GGGAA GTGCG TCCCC TCCGC TCCGG TCGCC TGCCC TGCCG
<u>cccc</u>	15	20200 20000 20000 20000 20000 20000 20000 20000 20000

and a second secon

1.21 Discussion:

De novo generation of micro-satellite markers through laboratory-based screening of SSR- enriched genomic libraries is highly time consuming and expensive an alternative is to screen the public database of related model species where abundant sequence data is already available, beyond cost savings, this approach also offers the possibility of identifying rare micro-satellite motifs, which would be uneconomical to identify through laboratory protocols the availability of massive amounts of nucleotide sequence data has led to the development of innovative ways to examine these data as reflected in their functions. Various types of DNA markers have been used in plant breeding and of these the most extensively used are the micro-satellite markers. The reasons for their extensive use are due to their mode of transmission, which is bi-parental-nuclear with few loci and many alleles per locus. Mode of gene action being co- dominance with exception of null allele at some loci, show large variation within populations and are generally found in non-coding regions, which may contribute to the genome stability. Genome sequence and protein sequence information is publicly available for large -scale analysis from Gene bank at (NCBI) and European Molecular biology Laboratory (EMBL). Today, the search for a gene of interest starts with sequence

Information, including expressed sequence tags (EST). Genome related public database are an invaluable part of the scientific community and most notably the model organism database, have two major consumers: the focused scientific community actively studying that system, and the large scientific community interested in relating this specialized information to and from other systems. The thrust of any high-throughput facility is the creation of large, well-organized, rigorous datasets. The model system database can be mined by other related crop specialist to design markers for marker-assisted selection and -aided introgression methods. such an approach can save valuable resources both in terms of time and funds.

2. Phylogenetic studies in cereals

2.1 Introduction

Phylogeny is really about Evolution and is used to reconstruct evolutionary events at a molecular level through analysis of molecular sequences, namely proteins & nucleic acids. Phylogeny is a diagram (a phylogenetic tree or cladogram) that depicts the evolutionary relationships among organisms. Comparative morphological, anatomical, embryological, molecular, behavioral, physiological, chemical, geographical, and fossil data can all be used, together or separately, to construct the phylogeny. Phylogeny provides the historical perspective from which to interpret the evolution of characters. patterns and processes of diversification, rates of evolution, historical biogeography, and co-evolutionary phenomena, such as the relationships between hosts and parasites or plants and herbivores. Phylogeny is used to classify organisms on the basis of their inferred evolutionary relationships (the phylogenetic approach to classification). Phylogeny is a hypothesis based on the best interpretation of the data at hand and subject to further evaluation (and possibly change) as new data become available.Phylogenetic focuses on the construction of descent relationships of species or groups of species and on how to incorporate these relationships into classification systems. Probably the first steps on establishing relationships and classifications were by Carolus Linnaeus in the 18th century. He is the founder of taxonomy, the biological branch of study dedicated to naming and classifying life forms. Linnaeus first established the binomial naming system, which continues today. In fact, many of the names he developed are still being used as scientific names. He also established a filing hierarchy, but avoided making his classifications based on evolutionary relationships. He was a natural theologian and believed that species were permanent creations that existed according to God's plan.

2.2 Phylogenetic terms

Adaptation	Change in a organism resulting from natural selection; a
	structure which is the result of such selection.
Anagensis	Evolutionary change along an unbranching lineage:
	change without speciation.
Ancestor	Any organism, population, or species from which some
	other organism, population, or species is descended by
	reproduction.
Basal group	The carliest diverging group within a clade; for instance,
	to hypothesize that sponges are basal animals is to
	suggest that the lineage(s) leading to sponges diverged
	from the lineage that gave rise to all other animals.
Character	Heritable trait possessed by an organism; characters are
	usually described in terms of their states, for example:
	"hair present" vs. "hair absent," where "hair" is the
	character, and "present" and "absent" are its states.
Clade	A monophyletic taxon; a group of organisms which
	includes the most recent common ancestor of all of its
	members and all of the descendants of that most recent
	common ancestor. From the Greek word "klados",
	meaning branch or twig.
Cladogenesis	The development of a new clade; the splitting of a single
	lineage into two distinct lineages; speciation.
Cladogram	A diagram, resulting from a cladistic analysis, which
	depicts a hypothetical branching sequence of lineages
	leading to the taxa under consideration. The points of
	branching within a cladogram are called nodes. All taxa
	occur at the endpoints of the cladogram.

Convergence Similarities, which have arisen independently in two or more organisms that are not closely related. Contrast with homology.

Crown group All the taxa descended from a major cladogenesis event. recognized by possessing the clade's synapomorphy. See: stem group.

Derived Describes a character state that is present in one or more subclasses, but not all, of a clade under consideration. A derived character state is inferred to be a modified version of the primitive condition of that character, and to have arisen later in the evolution of the clade. For example, "presence of hair" is a primitive character state for all mammals, whereas the "hairlessness" of whales is a derived state for one subclade within the Mammalia.

Diversity Term used to describe numbers of taxa, or variation in morphology.

Endosymbiosis When one organism takes up permanent residence within another, such that the two become a single functional organism. Mitochondria and plastids are believed to have resulted from endosymbiosis.

Evolution Darwin's definition: descent with modification. The term has been variously used and abused since Darwin to include everything from the origin of man to the origin of life.

Evolutionary tree A diagram, which depicts the hypothetical phylogeny of the taxa under consideration. The points at which lineages split represent ancestor taxa to the descendant taxa appearing at the terminal points of the cladogram.

Extinction When all the members of a clade or taxon die, the group is said to be extinct.

Gradualism	A model of evolution that assumes slow, steady rates of
	change. Charles Darwin's original concept of evolution
	by natural selection assumed gradualism. Contrast with
	punctuated equilibrium.
Homology	Two structures are considered homologous when they are
	inherited from a common ancestor, which possessed the
	structure. This may be difficult to determine when the
	structure has been modified through descent.
Hypothesis	A concept or idea that can be falsified by various
	scientific methods.
In-group	In a cladistic analysis, the set of taxa which are
	hypothesized to be more closely related to each other than
	any are to the outgroup.
Lineage	Any continuous line of descent; any series of organisms
	connected by reproduction by parent of offspring.
Monophyletic	Term applied to a group of organisms, which includes the
	most recent common ancestor of all of its members and
	all of the descendants of that most recent common
	ancestor. A monophyletic group is called a clade.
Outgroup	In a cladistic analysis, any taxon used to help resolve the
	polarity of characters, and which is hypothesized to be
	less closely related to each of the taxa under
	consideration than any are to each other.
Paraphyletic	Term applied to a group of organisms, which includes the
	most recent common ancestor of all of its members, but
	not all of the descendants of that most recent common
	ancestor.
Parsimony	Refers to a rule used to choose among possible
-	cladogram, which states that the cladogram implying the
	least number of changes in character states is the best.

53

Phylogenetic Field of biology that deals with the relationships between organisms. It includes the discovery of these relationships, and the study of the causes behind this pattern.

Phylogeny The evolutionary relationships among organisms; the patterns of lineage branching produced by the true evolutionary history of the organisms being considered.

Plesiomorphy A primitive character state for the taxa under consideration.

Polarity of characters The states of characters used in a cladistic analysis, either original or derived. Original characters are those acquired by an ancestor deeper in the phylogeny than the most recent common ancestor of the taxa under consideration. Derived characters are those acquired by the most recent common ancestor of the taxa under consideration.

Polyphyletic Term applied to a group of organisms, which does not include the most recent common ancestor of those organisms; the ancestor does not possess the character shared by members of the group.

Primitive Describes a character state that is present in the common ancestor of a clade. A primitive character state is inferred to be the original condition of that character within the clade under consideration. For example, "presence of hair" is a primitive character state for all mammals, whereas the "hairlessness" of whales is a derived state for one subclade within the Mammalia.

Pseudoextinction The apparent disappearance of a taxon. In cases of pseudoextinction, this disappearance is not due to the death of all members, but the evolution of novel features in one or more lineages, so that the new clades are not recognized as belonging to the paraphyletic ancestral

54

group, whose members have ceased to exist. The Dinosauria, if defined so as to exclude the birds, is an example of a group that has undergone pseudoextinction.

Punctuated equilibrium A model of evolution in which change occurs in relatively rapid bursts, followed by longer periods of stasis.

 Radiation
 Event of rapid cladogenesis, believed to occur under conditions where a new feature permits a lineage to move into a new niche or new habitat, and is then called an adaptive radiation.

Rank In traditional taxonomy, taxa are ranked according to their level of inclusiveness. Thus a genus contains one or more species; a family includes one or more genera, and so on.

Relatedness Two clades are more closely related when they share a more recent common ancestor between them than they do with any other clade.

 Reticulation
 Joining of separate lineages on a phylogenetic tree, generally through hybridization or through lateral gene transfer. Fairly common in certain land plant clades; reticulation is thought to be rare among metazoans.

SelectionProcess which favors one feature of organisms in a
population over another feature found in the population.
This occurs through differential reproduction -- those
with the favored feature produce more offspring than
those with the other feature, such that they become a
greater percentage of the population in the next
generation.Sister groupThe two clades resulting from the splitting of a single

The two clades resulting from the splitting of a single lineage.

A period of little or no discernible change in a lineage.

55

Stasis

Stem group	All the taxa in a clade preceding a major cladogenesis	
	event. They are often difficult to recognize because they	
	may not possess synapomorpies found in the crown	
	group.	
Synapomorphy	A character which is derived, and because it is shared by	
	the taxa under consideration, is used to infer common	
	ancestry.	
Systematics	Field of biology that deals with the diversity of life.	
	Systematics is usually divided into the two areas of	
	phylogenetics and taxonomy.	
Taxon	Any named group of organisms, not necessarily a clade.	
Taxonomy	The science of naming and classifying organisms.	
Vicariance	Speciation, which occurs as a result of the separation and	
	subsequent isolation of portions of an original population.	

2.3 Phylogenetic classifications

Integrating the results of a phylogenetic analysis into a classification is an extremely contentious issue in systematics. Basically, taxa are recognized on the basis of monophyly.

- 1. Linnaean system
 - a) Named groups are monophyletic
 - b) Not all groups are necessarily named
 - c) Ranks are arbitrary
- 2. Phylogenetic or "rankless" systems
 - a) Groups given unranked names
 - b) Groups defined by ancestry i.e. phylogenetic tree
 - c) Groups described/diagnosed by a character(s) on the branch of the monophyletic

2.4 Methods of phylogenic analysis

There are three methods

- 1) Cladistic
- 2) Phenetic
- 3) Multiple alignment

2.4.1 Cladistic Method

Cladistics is a particular method of hypothesizing relationships among organisms. Like other methods, it has its own set of assumptions, procedures, and limitations. Cladistics is now accepted as the best method available for phylogenetic analysis, for it provides an explicit and testable hypothesis of organismal relationships. The basic idea behind cladistics is that members of a group share a common evolutionary history, and are "closely related", more so to members of the same group than to other organisms. These groups are recognized by sharing unique features, which were not present in distant ancestors. These *shared derived* characteristics are called synapomorphies. The Cladistic approach takes into account both ancestral relationships between organisms and current data stored in their sequences. Here, morphological similarities/differences are also considered as part of the evolutionary history of organisms being compared. The software employed in this approach includes parsimony and maximum likelihood methods.

2.4.2 phenetic Method

Phenetics, also known as numerical taxonomy, involves the use of various measures of overall similarity for the ranking of species. There is no restriction on the number or type of characters (data) that can be used, although all data must be first converted to a numerical value, without any character "weighting." Each organism is then compared with every other for all characters measured, and the number of similarities (or differences) is calculated. The organisms are then clustered in such a way that the most similar are grouped close together and the more different ones are linked more distantly. The taxonomic clusters, called phenograms, that result from such an analysis do not necessarily reflect genetic similarity or evolutionary relatedness. The lack of evolutionary

significance in Phenetics has meant that this system has had little impact on animal classification, and as a consequence, interest in and use of Phenetics has been declining in recent years.

2.4.3 Multiple Alignment Method

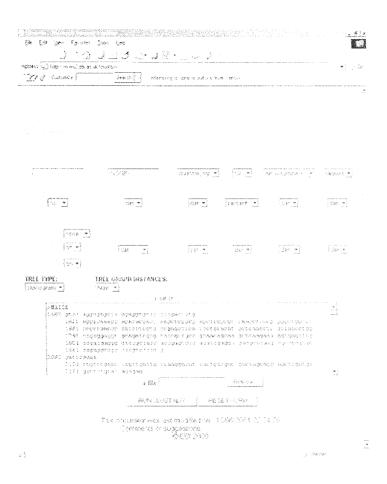
The most practical and widely used method in multiple sequence alignment is the hierarchical extentions of pair wise alignment methods.

- Multiple alignments are built by successive application of pair wise methods:
- Compare all sequences pair wise; (for N sequences there are N.(N-2)/2 pairs or scores)
- Perform cluster analysis on pair wise scores to generate a hierarchy for alignments;
- Build the multiple alignment by first aligning the most similar pair of sequences, then the next most similar pair and so on;

Once an alignment of 2 sequences has been made, then this is fixed. Thus for a set of sequences A, B, C, D having aligned A with C and B with D the alignment of A, B, C, D is obtained by comparing the alignments of A and C with that of B and D using averaged scores at each aligned position.

2.5 Clustalw

ClustalW is a sequence alignment tool, which is used to produce Multiple alignments of a protein sequence./nucleotide sequence. This web tool available at http://www.ebi.ac.uk/clustalw/



2.5.1 upload a file

Upload a file from the computer which containing a valid protein nucleotide sequence in any format (GCG, FASTA, PIR, etc.) by using this option. This option only works with Netscape Browsers'

2.5.2 Sequences

The sequences have different names, as the first 30 characters of the name are significant. If clustalw finds two or more sequences with the same name it will fail!

The following are the points to follow not to get the fatal error.

A set of sequences in different formats has been used as input. For example:

>Sea1 MRVLLAALGLUFLGALRAFPODRPFEDTCHGNPSHYYDKAVRRCCYRCP MGLFPTOOCPORPTDCRKOCEPDYYLDEADRCTACVTCSRDD >Sea2 PVAEERGLMSQPLMETCHSVGAAYLESLPLQDASPAGGPSSPRDLPEPRV STEHTNNKIEKIYIMKADTVIVGTVKAELPEGRGLAGPAEP ELEEELEADHTPHYPEOETEPPLGSCSDVMLSVEEEGKEDPLPTAASGK >Seq3 MSALLTAAGL LFLGMLOAFP TDRPLKTTCA GDLSHYPGEA ARNCCYQCPS GLSPTQPCPR CTACVTCLPG GPAHCRKOCA PDYYVNEDGK LVEKAPCSGN SPRICECOPG MHCCTPAVNS CARCKLHCSG KKDTICELPS EEVVKSPGTA SGSGPNCSNP GDRKTLTSHA TPOAMPTLES

All input sequence must be in the same format. The sequence separators every 10 as in seq3 are not allowed in the f asta format specification. The program does not allow this kind of input.

- · Sequences with spaces (delctions) will have the same affect as above.
- Sequences should not have illegal characters as for example.

>Badsequence

*SNERNVCN*WFVYWADQL*WIELKCFA*HICNSG*LQRYNKPKNNQPI WGGIKRRCSS*

• Need at least two sequences to generate an alignment.

Illegal characters in the input should be removed. This could be due to cuting & pasting from a word processing program..

Remove any empty space or empty lines from the beginning of the input. put format for the alignment.

Example:

In FASTA format:

2.5.3 search title

To identify the database search results.

2.5.4 cpu mode

This option allows the choice the version of clustalw based on cpu mode(Central processing unit). There are two options:

Clustalw_mp this is a parallelised version of clustalw that runs on multiprocessor SGI systems.

Clustalw this is the generic (single CPU) version of the program. Use this version if we are having prolems with the MP version of the program.

2.5.5 alignment

choose a full alignment or using a stringent algorithm for generating the tree guide or a fast algorithm.

2.5.6 Output

Output format for multiple alignment results display. The options are ALN, GCG, PHYLIP, PIR and GDE.

2.5.7 Jalview

A new experimental option has been added to the results page, which involved using a Java Applet called JalView. This is a fully featured MSA editor which allows not only to edit the alignment but also, to exchange the alignment formats. For documentation click on the JalView Hyperlink.

2.5.8 Outorder

Decide, which order the sequences, should be printed in the alignment.

2.5.9 color

To get the alignment in color this option should be selected. This option only works when we choose ALN or GCG as the output format.

2.5.10 Fast paiwise alignment options

Ktup	This option allows you to choose which 'word-length' to use when
	calulating fast paiwise alignments.(note: make sure you have
	choosen 'fast' in the ALIGNMENT.
Window	Use this option to set the window length when calculating fast
	pairwise alignments.(Note: make sure you have choosen 'fast' in
	the ALIGNMENT.
Score	This option allows you to decide which score to take into account
	when calculating a fast pairwise alignment. (Note: make sure you
	have choosen 'fast' in the ALIGNMENT.
Topdiag	Select here how many top diagonals should be integrated when
	calculating a fast pairwise alignment.(Note: make sure you have
	choosen 'fast' in the ALIGNMENT.
Pairgap	Select here to set the gap penalty when generating fast pairwise
	alignments.

2.5.11 multiple sequence alignment options

matrix This option allows you to choose which matrix series to use when generating the mulitple sequence alignment. The program goes through the choosen matrix series, spanning the full range of amino acid distances.

62

blosum (henikoff)	These matrices appear to be the best available for carrying out
	data base similarity (homology searches). The matrices used are:
	Blosum80, 62, 40 and 30.
pam (dayho <u>f</u> f)	These have been extremely widely used since the late '70s.We
	use the PAM 120, 160, 250 and 350 matrices.
gonnet	These matrices were derived using almost the same procedure as
	the Dayhoff one (above) but are much more up to date and are
	based on a far larger data set. They appear to be more sensitive
	than the Dayhoff series. We use the GONNET 40, 80, 120, 160,
	250 and 350 matrices. We also supply an identity matrix, which
	gives a score of 10 to two identical amino acids and a score of
	zero otherwise.

2.5.12 Gapopen

Set the penalty for opening a gap.

2.5.13 Endgap

Set the penalty for closing a gap.

2.5.14 Gapext

Set the penalty for extending a gap.

2.5.15 Gapdist

Set the gap separation penalty.

2.5.16 Phylogenetic Tree

To use this option need to input a sequence alignment, if this alignment is in PIR or PHYLIP format. ALN and GCG MSF files are not supported so this has to convert MSF files to PIR format.

This option allows choosing the following output formats for the tree:

Neighbor-Joining

Phylip

Distance

2.5.17 kimura correction of distances

This option allows setting on distances correction (correction for multiple subsitutions). This is because, as sequences diverge, more than one substitution will happen at many **Node:** represents a taxonomic unit. This can be either an existing species or an ancestor. Branch: Defines the relationship between the taxa in terms of descent and ancestry. **Topology:** The branching patterns of the tree.

Branch length: Represents the number of changes that have occurred in the branch. **Root:** The common ancestor of all taxa.

Distance scale: scale that represents the number of differences between organisms or sequences.

Clade: a group of two or more taxa or DNA sequences that includes both their common ancestor and all their descendents.

Operational Taxonomic Unit (OTU): Taxonomic level of sampling selected by the user to be used in a study, such as individuals, populations, species, genera, or bacterial strains.

A phylogenetic tree is composed of nodes--each representing a taxonomic unit (species, populations, and individuals)--and branches, which define the relationship between the taxonomic units in terms of descent and ancestry. Only one branch can connect any two adjacent nodes. The branching pattern of the tree is called the topology and the branch length usually represents the number of changes that have occurred in the branch. This is called a scaled branch. Scaled trees are often calibrated to represent the passage of time. Such trees have a theoretical basis in the particular gene or genes under analysis. Branches can also be unscaled, which means that the branch length is not proportional to the number of changes that has occurred, although the actual number may be indicated numerically somewhere on the branch. Phylogenetic trees may also be either rooted or unrooted. In rooted trees, there is a particular node, called the root--representing a common ancestor--from which a unique path leads to any other node. An unrooted tree only specifies the relationship among species, without identifying a common ancestor, or evolutionary path.

65

2.7 Methods For The Phylogenetic Studies Of Cereals

Cereals i.e., sorghum bicolor, Hordeum vulgare, Zea mays, Triticum aestivum, orizya sativum are taken and co – related the phylogenetic relationship between these cereals by taking a nuclear enzyme, a mitochondrial enzyme and a chloroplast enzyme.

2.8 Reasons For Taking Enzymes In Phylogenetic Studies

2.8.1 Nuclear Enzyme

- 1) Nuclear enzymes have bi-parental inheritance.
- 2) Nuclear enzymes are in abundance in a Genome.
- 3) Amplification and sequencings are easy.
- 4) They have mosaics of highly conserved variable region. The conserved regions have been informative for resolving relationship at higher taxonomic levels. Alignment of the variable region is often problematic.
- 5) They exhibit a wide range of evolutionary rate in phylogenetic utility.
- Many nuclear genes may contain large Introns that necessitate reverse transcriptase PCR.
- The nuclear enzyme which we are taken are alcohol-dehydrogenase due to its role in metabolism a key functional pathway

2.8.2 Mitochondrial Enzyme

- 1) Mitochondrial enzymes are maternally inherited.
- These enzymes are used to construct a PhylogeneticTree to display the Evolutionary relationships between Individual sequences.
- 3) The structure of this gene tree contains information which in conjunction with a calibrated mutation rate for the DNA sequence under study, can be used to estimate a time-scale for events in evolutionary prehistory.
- 4) Sites that have frequently undergone mutations are less conserved among species compared to those where the consensus is more the sequence is highly

conserved. Evolutionary changes are found the non-conserved regions of the sequence.

- 5) These will provide the phylogenetic evolution of a given mitochondrial gene.
- 6) The mitochondrial enzyme investigated in this study is Malate-dehydrogenase a key enzyme in the respiratory pathway.

2.8.3 Chloroplast Enzyme

- Chloroplast gene Restriction-Site Variation has been shown to be well suited for studies of genetic relationships at or below the family level.
- The chloroplast genome consists of a large and a small region of Single-copy DNA separated by a pair of identical but inverted repeat sequences.
- Restriction-pattern differences between taxa may be interpreted as site changes caused by single base substitutions or single insertion deletion events.
- 4) By relating variation in chloroplast DNA restriction-fragment patterns to specific mutations, either base substitution or indels data sets suitable for phylogenetic reconstructions using Parsimony analysis may be produced.
- Restriction-site variation is used to estimate total sequence divergence between taxa. Such distance measures may used to reconstruct phylogenies.
- Chloplast Enzymes found in plants only. These enzymes are mostly related to C3 and C4 pathways of photosynthesis.
- 7) These enzymes are maternally inherited.
- The chloroplast enzyme, which we taken is Phosphoenol pyruvate the key enzyme of the C4 photosynthetic pathway.

2.9 Multiple Alignment Method

The most practical and widely used method in multiple sequence alignment is the hierarchical extensions of pair wise alignment methods.

2.10 Steps Involved In Multiple Alignment Method

2.10.1Select The Most Conserved Enzymes:

Select the most conserved and functional mitochondrial, chloroplast and nuclear enzymes of cereals. In cereals the most conserved functional enzymes are as follows: -

Mitochondrial Enzymes:

1.	NADH dehydrogenase	1.6.5.3
2.	ATP synthase	3.6.3.14
3.	Succinate dehydrogenase	1.3.99.1
4.	Malate dehydrogenasc	1.1.1.82
5.	Citrate synthase	4.1.3.28

Chloroplast Enzymes:

1.	Fructose 1,6 bis phosphatase	<u>3.1.3.11</u>
2.	Phosphoenolpyruvate carboxylase	4.1.1.31
3.	Glyceraldehyde- 3-phosphate dehydrogenase	1.2.1.9

Nuclear Enzyme:

1.Methyl transferase	<u>2.1.1.37</u>
2.Alcohol dehydrogenase	1.1.1.1
3.Cysteine synthase	4.2.99.8

From the all above-mentioned enzymes we selected the following enzymes, which are more convenient for our practical purpose.i.e Malate dehydrogenase (mitochondrial enzyme)

Phosphoenol pyruvate carboxylase (chloroplast enzyme)

Methyl transferase (nuclear enzyme)

2.10.2 Search The Sequences From NCBI:

Scarch the sequences of the Malate dehydrogenase (mitochondrial enzyme), Phosphoenol pyruvate carboxylase (chloroplast enzyme), and Methyl transferase (nuclear enzyme) From NCBI.(National Centre For Biotechnology Information)

Then note the accession number of the above enzymes of interest of cereals.

Chloroplast Enzyme (Phosphoenol Pyruvate Carboxylase)

Cereal	Accession number
Maize	E17154
Rice	AF271995
Sorghum	AF399915

Paste accession numbers of the cereals for a Chloroplast enzyme (Phosphoenol pyruvate carboxylase) in the NCBI. The sequences for the accession number of the cereals for ex:-(maize, oryza sativa and sorghum bicolor) will display.

 All the second states and the s		않았는 것 못 하는 것 같아. 가 가 가 가 가 다 가 다 가 다 다 다 다 다 다 다 다 다 다	
Ene Ene Yers Spranes Ioon 147			
-2+ J_	14 コロタート 14 日日 ひと		
Agana (
NO Leane		1817 B	
	- Network in the second sec		
	National Center for Biolectus dogs	lands anna an lanan	
		a daar na caalaa ay ah ne b	
2012/246622			
Search Museor			
	What does NCBI do?		
·	Finalesteron, 1465 en el center com una com	1 Leiner Secure	
	– inste svarskology in omratise (1908) ordstea	A CROBELS OF THE T	
	द्या २४४ (प्रहास) हाइम्स्स १९४१ व्याप्त १९४ ४६ १२२४ (१९४१)	1 Jugarens en	
	- ಸರ್ವಾಧವರ್ಷದಲ್ಲಿ ಹಿನಸ್ಥಿ, ಅದುಕಾಂಧ್ರ ನಿಲ್ಲೇತಗಳ "ಸಂಗ್ರಾಮಿಗ ನಡುತ್ತದಲ್ಲಿ ತಾರ್ಮಗಳು ಮನಕ್ಕ ನಗರ	tett i Regional qui su c	
	deservated transactal membrane area	Coltan Friday	
- And	Merketer moerstwaard, Armsteader Armsteader		
	iningsiges with include that with states and wise with	 The second s	
		· Dens excession	
		1.8 × 1.8 × 1.	
	「「「「「「」」」では、「」」では、」」では、	Constant and Annalise	
	or use the borner gets and service th		
	32/52 FH	ningen gestanden. Verstaanse	
	and all a set of a set of a		
	Refseqs for viral genomes!	$(22\pi^{2})^{2} = (2\pi^{2})^{2} = (2\pi$	
	Trans are non-control 000 mod educeria control control officer a control control	A Section of the sect	
	- 「「「「「」」「「」」」「「「」」」」」」」」」」」」」」」」」」」」」」	No. Constant Constants	
	Analyze in sphile genotrial of Martes and inapies the Analyze	11000	
	VisiCamerica wasarin Mina	i shigewa basari i k	
	a complete the set	and a second	
	NCBI in the News	1.87 5-22	
	医静脉管 化合物合物合物 建水合物化合物 化分析工具		
	ues places das vije	 Miterature Simulation and projects 	
	, me Encez databas e dystem i sigen do tro ore ogicist stevance of MCE's molecular		
	なななな語られた、美心ななならならならったらいた。	i filmanan Teraput ing	
	and say aspects such as taken only medicine		
	appletantent, and the subvector (* 1755) Sciences Marcin (1952)	化合成第一角化化 是一个人 ^{们们们} 这是种学生(这个会主要公共	
fer for groot		- Andre Construction - Construction	44 ·
		and a second	
	- Heread June Tol (2001	- True	
		and Br	
		1、2111年1月1日	
		10.00	
		1 - 1 - 1 - 2 - 2 - 2 - 2 - 2 - 2 - 2 -	
A COLOR OF A COLOR			

Elle Edd Year Frankins Totis Esp •1 × Gr 10 g Owner interior in interior to the state of the second Follows moventia proteia devana attractive Probled Sanacher Caller Grade Scarch Hunlecode + 6 ET 754.4F271995.4F29291 and there Disc by Beauty • Seventier List Aca 12em: 1-3-37 3. T 2 - $\Gamma +$ Mane characteric lyse water with a splace gets gentistija gentes (gentasjin när) 52 Organistica phosphore depression activity is mobile, a support of the #18284444144F071995(188F071995(2020444) 17 3: Soughon bushes phosphoenelpyroute and applies house (FFTE) within complete staalbo22404ge/AF379215 LAF379215[15025404] Reviet Scholer 14 1911 Ľ.

2.10.3 Accession Number

First click on accession number of marze then oryza sativa followed by sorghum bicolor Sequences of Marze, oryza sativa and Sorghum bicolor phosphenol pyravate gene will obtain.

NEW CONTRACTOR DAMAGES AND REAL TO A CONTRACTOR fie for your famores forth help 1 *・ しいはんじゅう *** ほうしく ** And the fighting there with an the government from the found of the well of an antical and second and the base ₩. -×13a Search 71 & Carminion [effected and refine the factors of the standard Findward Michaeling Pringen Salerana Strendung Fersbal Zanasarmy (Called Broad) search Nudeobde - for Gol Clear Dispay defeat - Save Tax Add to Claphoerd 1: E17154, Maize phosphoenol. [ei:5711837] LOCUS E17154 6781 bp DNA hnear PAT 28-JUL-1999 DEFINITION Maize phosphoenolpyruvate carboxylase gene ACCESSION E17154 VERSION E171541 GE5711837 KEYWORDS JP 1998248419-A 2 SOURCE unidentified ORGANISM nnclassified REFERENCE 1 (bases 1 to 6781) AUTHORS: Matsuoka M. Tokutomi M. TokuS and Moris, S.K. TITLE C3 PLANT BODY EXPRESSING PHOTOSYNTHETIC ENZYME OF C4 PLANT. IOURNAL Patent JP 1998248419-A 2 22-SEP-1998 NORIN SUISANSYO NOGYO SEIBUTSU SHRÆN KENKYU SHO COMMENT OS Zea mays (maize) PN JP 1998248419-A.2 PD 71.SEP-1998 PE 11-MAR-1997 JP 1997056742 PE MATSPOKA MAROTO, TOKUTOMEMETSME, TOKESERTH, PE MORIS SUN-BEN KU PC A01H5 00, A01H1 00, C071P1 04, C12N5 10, C12N9 00, C12N15 09, PC (C12N5.19) PC C12R1/91/(C12N15/09/C12R1/91): CC strandedness: Double. CC topology: Linear:

Location/Qualifiers FH Kev FH FT source 1..6781 /organism='Zea mays' FT /cultivar='Golden Cross Bantam' FT intron FT1477 1586 FT exon 1587 1981 1982..2091 FT intron FT exon 2092..2176 FT intron 2177..3073 3074 3296 FT exon 3297..3745 FT intron FT exon 3746..3849 FT intron 3850..3974 3975..4065 FT exon FT intron 4066..4202 FT exon 4203..4357 FT intron 4358..4453 FT exon 4454 5452 FT intron 5453..5549 FT exon 5550..5936 FT intron 5937..6024 FT CDS 17T join(1297..1476,1587..1981,2092..2176, FT 3074.3296. FT 3746..3849.3975..4065.4203..4357, FT 4454..5452.5550..5936. FΤ 6025..6318) FT /product='phosphoenolpyruvate coaboxylase'. FEATURES Location/Oualifiers source 1..6781 /organism="unidentified" /db_xref="taxon:32644" BASE COUNT 1687 a 1854 c 1548 g 1692 t ORIGIN 1 tetagagatg taatggtgtt aggacaegtg gttagetaet aatataaatg taaggteaaa 61 attegatggt ttatttteta tttteacita eetageatta teteatitet aattgtgtga 121 taacaaatgc attagaccat aattetgtaa atacgtacat ttaagcacac agtetatatt 181 ttaaaattet tettitigig tggatateee aacceaaate cacetetete eleaateegt 241 gtatetteae egetgecaag tgecaacaac acategeate gtgeaaatet ttgttggttt 301 gtgcacggtc ggcgccaatg gaggagacac ctgtacggtg cccttggtag aacaacatcc 361 tratecetat atgratgete cettegtaga tggeaccett atecetacaa tagecatgta 421 ttgcatacca agaattaaat atactttttc ttgaaccaca ataatttatt atagcggcac 481 ttettgttet ggttgaacae ttatttggaa caataaaate eegagtteet aaceaeaggt 541 teactititt teettateet eetaggaaae taaattttaa atteataaat ttaattgaaa 601 tettaatgaa aacaaaaaaa ttatctacaa agacgactet tagecacage egecteactg

661 cacceteaac cacateetge aaacagacae eetegecaca tecetecaga ttetteeete 721 egatgeagee tactigetaa eagaegeeet eteeacatee tgeaaageat teeteeaaat 781 tettgegate cecegaatec ageattaact getaagggae geceteteea cateetgeta 841 cccaattage caacegaata acacaagaag gcaggtgage agtgacaaag caegtcaaca 901 geacegagee aagecaaaag gagcaaggag gagcaageee aageegeage egeagetete 961 caggteccet tgegattgee gecageagta geagacacee etetecacat eccetecgge 1021 cgctaacagc agcaagccaa gccaaaaaga agcctcagcc acagccggtt ccgttgcggt 1081 taccecceat cacatececa agecegeec titecaaace cegaggeeg ecegiteceg 1141 tgcacageca cacacace egecegecaa egacteecea teetatttg aacecaceeg 1201 cgcactgcat tgatcaccaa tcgcatcgca gcagcacgag cagcacgccg tgccgctcca 1261 acceletege treectgett agetteeege egegeeatgg egtegaceaa ggeaceegge 1321 cetggegaga ageaceacte categaegeg cageteegte agetggteee aggeaaggte 1381 techangacg acaageteat egagtacgat gegetgeteg tegacegett ceteaacate 1441 ctccaggace tecaegggee cageettege gaatttgtaa etaaceaeeg eegeegeeca 1561 agtgettact actgtaatge atgeaggtee aggagtgeta egaggtgteg geegactatg 1621 appreciating appreciate and the second 1681 cegeegacge catectegtg gegageteea teetgeacat geteaacete gecaacetgg 1741 ccpaggaggt gcagategcg cacegeegee gcaacageaa geteaggateg 1801 ccgacgaggg ctccgccacc accgagtccg acatcgagga gacgctcaag cgcctcgtgt 1861 eegaggtegg caagteecee gaggaggtgt tegaggeget caagaaceag acegtegace 1921 testetteae escecatece acseastees eccecete setectseas aaaaacseca 1981 ggtatatatt teteaatgge ttgategata tgetaeteae gttatataee ettaagtett 2041 aaccattatt attatttttg ataaataaaa atgteggtet tgtegetgea ggateeggaa 2101 ttgtetgace cagetgaatg ceaaggacat caetgacgae gacaageagg agetegatga 2161 geotetgeae agagagatae gtacatatta cattteacae cagggaatge aagaacttta 2221 teaagagaca tteattettt gatagagata gaatagaaca catgeacagt acaegtggac 2281 teatgagett geagacate gageacgaca egtptaagtt agtgegeeag agaaatette 2341 aatttatatg teaagteagg teaggttete ceattaaaac acatataaat aaatatteat 2401 tattatcaag ctaaggtaat aaacaaccaa actittccac tatttaaact gtctttgcaa 2461 actecaaagt agaaactaac ctaatcagga aagaactaga ctgcacattt atgttttaac 2521 aatgeaatga gagaactget acatgtataa cagaattatt tatatgagge egaettgaet 2581 taagatteaa tgttgaagae caettgatga aaactacaet gaattattta tatgetatte 2641 tecagetgtg etcaaageat ttteetttta ettaaaaaag ateattttgt acaaagatet 2701 ettactcata tagagecatt tgagtagaac tteggtacea cagatgeatt aatggtttag 2761 ttgtaatcaa gttgttgtac tcatcattat atttteetaa caagtaggge atceagttte 2821 teettgatga ggaateaaac etagatagee ttaacteeac acceteaatt agetaggeta 2881 tgeteaagtt eetagtgtta caaattteag aegagteata atgteateae tgageaeteg 2941 gtaaagageg teteteteat ggtgcatata tatgatgcag accacetgag aagtttactg 3001 etteaageea eeaaagtggt attitigtig titgggtigt tiagttetaa tteettitet 3061 tgggtgttca cagatecaag cageetteag aacegatgaa atcaggaggg cacaacecae 3121 cccccaggac gaaatgcgct atgggatgag ctacatccat gagactgtat ggaagggggt 3181 gectaagtte ttgegeegtg tggatacage eetgaagaat ateggeatea atgagegeet 3241 tecetacaat gittetetea tieggitete tiettggatg ggtggtgaee gegatggtae 3301 attictgcct accettitea ataaagtgge aggagetete tgtettteag ettgagagaa 3361 accttectge titactetga etgeaataga tgiteagaaa aactagteta teatitegag

3421 ctctcaggag ctagaatttt aaaattgaaa ttatttagta cacctcacta ataaaaattt 3481 atcatccata catgetagea caacatataa geataattta ateaaatett tatattgeaa 3541 cctggaacet aacttgttga attttttata teacagaatt atacgtgtag tattatttta 3601 tatatcaaag agtgettata ttatatcagt acttgteetg teaatattea aggetaaegt 3661 ttttetttte tegeeagaaa attatatata cagaattata tgttttttet aageetgtat 3721 atetttgeaa tetategeta tataggaaat eeaagagtta eeceggaggt gacaagagat 3781 gtatgettee tggecagaat gatggetgea aacttgtaca tegateagat tgaagagetg 3841 atgtttgagg tactgtacat ceatactgea gatttgtttg attgaatget etatgatttt 3901 ttgettgece tgttttttge tgteteeggt ceataceaga acteteatge atgeategte 3961 tgatatatet gtagetetet atgtggeget geaacgatga gettegtgtt egtgeegaag 4021 ageteeacag ttegtetggt tecaaagtta ceaagtatta cataggtaac cacaaacaga 4081 ageatttatg titgettaat tittgeetge egtacagget titgeaaaag teteeaetag 4141 tgttttcaaa ttaatttgag ggetettttg geatetttte tgaagtgtat ttgetggege 4201 agaatteteg aagcaaatte eteeaaacga geeetaeegg gtgataetag geeatgtaag 4261 ggacaagetg tacaacacac gegagegtge tegecatetg etggettetg gagtttetga 4321 aattteageg gaategteat ttaceagtat egaagaggta aatategtea tgtatatatt 4381 atatatatte atagtatgae ateageaetg caactaacaa aaaaaaaate actactgteg 4441 tgcatgcatg cagtteettg agceaettga getgtgetae aaateaetgt gtgaetgegg 4501 egacaaggee ategeggaeg ggageeteet ggaceteetg egecaggttt teaegttegg 4561 getetecetg gtgaagetgg acateeggea ggagteggag eggeacaeeg aegtgatega 4621 egecateace acgeaceteg geategggte gtacegegag tggteegagg acaageggea 4681 ggagtggetg etgteggage tgegaggeaa gegeeegetg etgeeeeegg acetteeeea 4741 gacegaggag ategeogacg teateggege gtteeacgte etegeggage teeegeega 4801 cagettegge cectacatea tetecatgge gaeggeeeee teggaegtge tegeegtgga 4861 geteetgeag egegagtgeg gegtgeggee ageegtgeee gtggtgeege tgttegaaag 4921 getggccage etgeagtegg egecegegte egtggagege etettetegg tggaetggta 4981 catggaccgg atcaagggca agcagcaggt catggtcggc tactccgact ccggcaagga 5041 egeeggeege etgteegegg egtggeaget gtaeagggeg eaggaggaga tggegeaggt 5101 ggccaagcgc tacggcgtca agctcacctt gttccacggc cgcggaggca ccgtgggcag 5161 gggtggeggg eccaegeace ttgecateet gteecageeg eeggacaeca teaaegggte 5221 catcegtgtg acggtgcagg gegaggtcat egagttetge tteggggagg ageacetgtg 5281 ettecagaet etgeageget teaeggeege eaegetggag eaeggeatge accegeeggt 5341 eteteeeaag eeegagtgge geaageteat ggaegagatg geggtegtgg eeaeggagga 5401 gtaccgetee gtegtegtea aggaggeeeg ettegtegag taetteagat eggtatgetg 5461 ccattgccca ttgctttgtg acgategaat tcatccatgt cgategttet tttcattcat 5521 tegagegttt gtgegteaet eactateagg etaeaeegga gaeegagtae gggaggatga 5581 acateggeag ceggeeagee aagaggagge eeggeggegg cateaegaee etgegegeea 5641 teccetggat ettetegtgg acceagacea ggtteeacet eccegtgtgg etgggagteg 5701 gegeegeatt caagttegee ategacaagg aegteaggaa etteeaggte eteaaagaga 5761 tgtacaacga gtggccattc ttcagggtca ccctggacct gctggagatg gttttcgcca 5821 agggagacce eggeattgee ggettgtatg aegagetget tgtggeagaa gaacteaage 5881 cetttgggaa geageteagg gacaaataeg tggagacaea geagettete eteeaggtae 5941 caaaaccage actgeactgt acgatatgaa taaaagtetg ttgtetgget eetgategat 6001 gactgactac tecattttgt geagateget gggeacaagg atattettga aggegateea 6061 tteetgaage agggaetggt getgegeaae eeetacatea ceaecetgaa egtgtteeag 6121 gectacaege tgaageggat aagggaeeee aactteaagg tgaegeeeea geegeegetg

6181 tecaaggagt tegeegaega gaacaageee geeggaetgg tenagetgaa eceggegage 6241 gagtaceege eeggeetgga agacaageee geeggaetgg tenagetgaa eceggegege 6301 agegaaca etggetagge ggettetett caeteacetg eugagtgeat egeegeetge 6301 agetteegga tggtggegt tigteagtti tggatggaaa tgeegaetig geagegtetg 6421 titteeetat geatatgitaa titeetgeet etitatute aetetietig tetageteeaa 6421 titteeetat geatatgitaa titeetgeet etitatute aetetietig tetageteeaa 6431 gitggaaaate tiggetagta taeetaatig taataataaa eetigetig tetagetee 6441 gitggaaaate tiggetatut ataeetaatig taataataaa eetigetig tetageteeaa 6451 gitggaaaate tiggetatut ataeetaatig taataataaa entegtagaa tetgetiget 6541 gittigtaat aattaataa tateeeagee entiggatigg aetigtitae eatggtigtia 6601 etteageeae eetetettag tigggetaaa eanttetiga tiggiatut tittistaga 6661 gaacetagt geattaett aagagagatg atatetagtg genetagtga tiggitgea 6721 agattgagaa etigttaete geteetagag gitaacaeta geaagtgatt ggagettagg 6781 g

 Im Law Same Space Same Inc. Inc.

 Im Law Same Space Same Inc. Inc.

 Im Law Same Space Same Inc.

 Im Law Same Space Same Space State Stat

1: AF271995 Oryza sativa phos . [gi:9828444]

LOCUS AF271995 3307 bp mRNA linear PLN 16-AUG-2080 DFFINITION Oryza sativa phosphoenolpyravate corboxyfase mRNA, complete eds

ACCESSION AF271995

VERSION AF271995.1 619828444

KEYWORDS

SOURCE — Oryza sativa (japonica cultivar-group)

ORGANISM Department of payment and frames

Enkaryota, Vindiplantae, Streptophyta, Embryophyta; Fraelteophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Phrhartoideae, Oryzeae; Oryza;

REFERENCE | (bases 1 to 3307)

AUTHORS Yamamoto, N., Kurita, A., Masumura, L., Sagumoto, T., Moota, S., Shiraishi, N., Oti, Y. and Tanaka, K.

${\rm THE}-{\rm Root}$ type of phosphoenolpyruvate carboxylase in developing nee seeds

- JOURNAL Unpublished
- REFERENCE 2 (bases 1 to 3307)
- AUTHORS Yamamoto, N., Kurita, A., Masumura, T., Sugimoto, T., Morita, S., Tanaka, K., Shiraishi, N. and Oh, Y.

TITLE Direct Submission

JOURNAL Submitted (24-MAY-2000) Lab. Plant Nutrition, Faculty of Agriculture, Kobe University, Nada, Kobe 657-8501, Japan

FEATURES Location/Qualifiers

/protein_id="<u>AAG00180.1</u>" /db_xref="GI:9828445"

source 1..3307 /organism="Oryza sativa (japonica cultivar-group)" /cultivar="Nipponbare" /db_xref="taxon:39947" /de_stage="developing seeds" (<u>CDS</u> 188..3067 /codon_start=1 /product="phosphoenolpyruvate carboxylase"

/translation="MERHQSIDAQLRLLAPGKVSEDDKLVEYDALLVDRFLDILQDLH GPHLREFVQECYELSAEYENDRDEARLDELGRKLTSLPPGDSIVVSSSFSHMLNLANL AEEVQIAHRRRIKLKRGDFADEASAPTESDIEETLKRLVTQLGKSREEVFDALKNQTV DLVFTAHPTQSVRRSLLQKHGRIRNCLRQLYAKDITADDKQELDEALQREIQAAFRTD EIRRTPPTPQDEMRAGMSYFHETIWKGVPKFLRRIDTALKNIGINERLPYNAPLIQFS SWMGGDRDGNPRVTPEVTRDVCLLARMMAANLYFSQIEDLMFELSMWRCSDELRIRAD DLHCSSRKAAKHYIEFWKQIPPNEPYRVILGGVRDKLYYTRERTRHLLTTGVSEIPEE ATFTNVEEFLEPLELCYRSLCACGDKPIADGSLLDFLRQVSTFGLALVKLDIRQESDR HTDVLDAITTYLGIGSYAEWSEEKRQDWLLSELRGKRPLFGPDLPQTEEIADVLGTFH VLAELPADCFGAYIISMATAPSDVLAVELLQRECHIKQPLRVVPLFEKLADLEAAPAA VARLFSIDWYMNRINGKQEVMIGYSDSGKDAGRLSAAWQMYKAQEELVKVAKHYGVK L

TMFHGRGGTVGRGGGPSHLAILSQPPDTIHGSLRVTVQGEVIEHSFGEEHLCFRTLQR FTAATLEHGMHPPISPKPEWRALMDEMAVVATKEYRSIVFKEPRFVEYFRSATPETEY GRMNIGSRPSKRKPSGGIESLRAIPWIFAWTQTRFHLPVWLGFGGAFKHIMQKDIRNI HTLKEMYNEWPFFRVTLDLLEMVFAKGDPGIAALYDKLLVAGDLQSFGEQLRNNFEET KQLLLQVAGHKDILEGDPYLKQRLRLRESYITTLNVCQAYTLKRIRDPSFEVMSQPAL SKEFVDSNQPAELVQLNAASEYAPGLEDTLILTMKGIAAGMQNTG" BASE COUNT 816 a 814 c 880 g 797 t

ORIGIN

1 eccacactet etacececaa aettgecete caatetatet eegeegeege egeegeetee 61 geogegagag ggogtegeog agetgegegg gaggetegee teoegeagea tetggeogga 121 ttegtegtag tegaggeage tgtaagegag agagagggag atggeegegg eegeegggaa 181 ggeggegatg gageggeace agtegatega egegeagetg eggetgeteg egeeggggaa 241 ggtateegag gaegacaage tegtegagta egaegeeete etegtegaee getteetega 301 catcetecag gacetgeacg geceteacet cegegaatte gtgeaggagt getacgaget 361 gteggeggag taegagaacg acagggaega ggegeggete gaegagetgg ggaggaaget 421 caccageetg eegeeggggg acteeategt egteteeage teettetege acatgeteaa 481 cetegecaac etegecgagg aggtgeagat egegeacege egeeggatea ageteaageg 541 eggggaette geegaegagg eeteegegee eacegagtee gacategagg agaegeteaa 601 gegeetegte acceageteg geaagtegag ggaggaggtg ttegatgege teaagaacea 661 gacegttgac etegtgttea eegegeatee aaegeagtee gteaggaggt eeetgeteea 721 gaagcacggg aggateegga attgeetgag geagttatat gegaaagaca teactgetga 781 tgacaagcag gagettgatg aggeeetgea gagggagatt caagcagett teagaactga 841 tgaaateege aggaeteete ceaeteetea ggatgaaatg egegeeggga tgagttattt 901 teatgaaact atatggaagg gtgtaccaaa gttettgege egeattgata etgetetgaa 961 aaatattggg attaatgage gtetteetta caatgeteet eteateeagt teteateetg 1021 gatgggtggt gaccgtgatg gaaacecaag agttacacea gaggttacaa gagatgtatg 1081 ettgttggca agaatgatgg etgetaacet gtacttetet cagatagaag atetgatgtt 1141 tgagetetet atgtggeget geagtgatga aettegaatt egtgeagatg atttgeattg 1201 etectecaga aaagetgeaa ageactatat agaattetgg aageaaatte etecaaatga 1261 geettategt gteataettg gtggtgteag ggataaattg taetataege gtgaaegtae 1321 tegecateta ttgacaactg gagtttetga aatteeagag gaggeaacgt ttaetaatgt 1381 tgaagagttt etggageege ttgagetgtg etacagatea ttatgtgett gtggtgacaa 1441 acctataget gatggaagee ttettgattt ettgegteaa gtateaaett ttgggetgge 1501 tettgtaaaa ettgacataa ggcaggagte tgategacac actgatgtee ttgatgccat 1561 aactacatat ettgggattg gatettaege tgaatggtet gaggagaaac gecaggattg 1621 gettttgtee gaactaaggg geaaacgtee tttgtttggt eetgatette eteagaetga 1681 agaaattget gatgttttag gaacatttea egteettget gageteeegg cagattgttt 1741 tggtgettae ateateteaa tggeaactge aceatetgat gtgettgetg tggagettet 1801 acagegggag tgecatataa aacageetet gegagtigti cecetatiig agaaacitge 1861 agatettgaa geageteeag eageggtgge acgaetattt teaatagaet ggtacatgaa 1921 taggattaat ggcaagcagg aggtcatgat tggatactca gactctggca aggatgctgg 1981 tegtetttet geageatgge aaatgtacaa ageacaagag gagettgtea aggtggeaaa 2041 gcattatggt gtaaagttga caatgttcca tggaagaggt ggaacagttg gcagaggagg 2101 tggteecagt catettgeca tattatetca accaccagae acgatacatg gateactteg 2161 tgtaacagta cagggcgagg ttattgaaca ctcgtttgga gaggaacact tgtgctttag 2221 aactetgeag egetttaetg eagetactet tgageatgga atgeateete eaattteeee 2281 caaaccagaa tggcgtgete taatggatga gatggetgtt gtggcaacaa aggaatateg 2341 atcaategte tteaaggage caeggtttgt tgaataette egateggeaa caeetgagae 2401 agaatatggc aggatgaaca ttggtageeg accateaaag agaaageeta gtggtggcat 2461 tgaategete egtgeaatte ettggatttt tgettggaca caaactaggt tteatettee 2521 tgtatggctt ggatttggtg gagcgttcaa acatataatg cagaaggata tcaggaacat 2581 ccatacactg aaagaaatgt acaatgagtg gccattette agggteacet tggaettget 2641 tgagatggtt ttegetaagg gagatecagg aattgetget ttatatgaca aattgettgt

270) ggcaggtgal etgcaateet tiggagagea getgagatae aacttigaag agacaaaaca 276) getaeteett caggtigeeg glcacaagga tattetggaa ggggateett acetgaagea 2821 gegtetgegg etgcgtgagt egtacateae tacetigaac gittigeeag eetaeteett 288) gaageggala agggaceta gettegaggt galgtegeag ceggeetigt egaaggagtt 2941 egtegacage aaceageetg eegagetggt ecaaetgaate getyeengee gaaggagtt 2001 tiggeetggag gacaceetea teetgaegat gaagggeatt getgeengeen teenaaaaete 3001 tiggeetggag gacaceetea teetgaegat gaagggeatt getgeergeen teenaaaaete 3001 aggetagatt gecacaaga atgetteeag eaggtgenat caateaatea teaataaete 3061 aggetagat gagatggtga atatgittaa etagtaegit egeegatget agtgaattei 3181 gaggaettit tittgetege aegittett tiggitgigt aetigatga aaagtcaagg 3241 ggggaettat tiggettig eageeegate caaaataaaa tatggaceaa teattetaeg 3301 aaaaaaa

/ Provide law pro
/ Provide law provide law pro
/ Provide law provide law pro
/ Provide law provide law provide law provide law pro
/ Provide law provide law provide law provide law pro
/ Provide law provide law provide law provide law provide law provide law pro
/ Provide law provide law

1: AF399915. Sorghum bicolor p. (gr.15029404)

LOCUS AF399915 1389 bp mRNA linear PLN 30-JUL-2001 DEFINITION Sorghum bicolor phosphoenolpyravate carboxylase kinase (PPCK) mRNA.

complete eds.

ACCESSION AF399915

VERSION AF399915.) GE15029404

KEYWORDS.

SOURCE sorehum

ORGANISM Elizabert Studies

Eukaryota; Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, Lihopsida, Poales; Poaceae, PACC elade, Panicoideae, Andropogoneae; Sorahum

REFERENCE 1 (bases 1 to 1389)

AUTHORS Hartwell, J., Gill, A., Nimmo, G.A., Wilkins, M.B., Jenkins, G.L. and Nimmo, H.G.

TTLE Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression

JOURNAL Plant J. 20 (3), 333-342 (1999).

MEDLINE COSCO76

PUBMED 10571893

REFERENCE 2 (bases 1 to 1389)

AUTHORS Fontaine, V., Hartwell, J., Jenkins, G.I. and Nimmo, H.G.

TITLE Direct Submission

JOURNAL Submitted (12-JUL-2001) IBLS, University of Glasgow, University Avenue, Glasgow G12 8OO, UK

FEATURES Location/Qualifiers

source 1..1389

/organism="Sorghum bicolor" /db_xref="taxon:4558"

gene 1..1389

/gene="PPCK" <u>CDS</u> 21..944 /gene="PPCK" /note="protein kinase" /codon_start=1 /product="phosphoenolpyruvate carboxylase kinase" /protein_id="ΔΔ<u>K81871.1"</u> /db_xref="GI:15029405"

/translation="MSAELKRDYEIGAEIGRGRFGVVHRCTSRATGEAFAVKSVDRSQ

LADDLDRELAELEPKLAQLAGAGNPGVVQTHAVYEDETWTHTVMDLCAGPDLLEWVR L

RRGAPVPEPLAAAIVAQVAQALALCHRRGVAHRDVKPDNILIDTAPAGSDDGEDEEED

SGEAEEAETAPRARLADFGSAAWVGAGGLGRAEGLVGTPHYVAPEVVGGGEYGAKAD V

WSAGVVMYALLSGGALPFGGETAAEVLAAVLRGSVRFPPRLFSGVSPAAKDLLRRMIC RDEWRRFTAEQVLAHPWIVSGGGARAIERPT" BASE COUNT 263 a 393 c 507 g 226 t

ORIGIN

- 841 tgatetgeeg egacgagtgg eggaggttea eegeegagea agteeteget eaceegtgga
- 901 tegtgagegg eggaggagea egggeaateg ageggeeaae etgaggatga ggagggeata
- 961 ccataccage agaggaggag accatatgge gggeggaegt gtattgeegt agagacacag
- 1021 acaeggeaag gaateeegae gtegttatgt ataeggtttt gtatgtatgt atgtatgtae
- 1081 gtacggacca gagagtagta gagtacggta getgetgeeg aatgatgagt ageateateg
- 1141 attecgeget gtgtaggege ageagageag ategategge aaagaagett gtataggett
- 1201 etggttttee ttaetteegg gtttttggga tgetetttae etageagtge agageaecet
- 1261 ttttttggat gtgcctggta agtggtaget ggatgtatat acactgaccg tcttgtacaa
- 1321 atggagtgga gttggataaa taaataaaat cgagcetttg tacttecaaa aaaaaaaaaa
- 1381 aaaaaaaaa

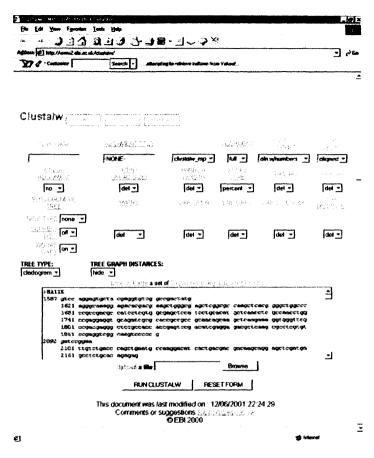
2.10.4 Exon Regions

Note the Exon regions of maize, oryza sativa and sorghum bicolor pyruvate carboxylase

from the NCBI, which is given above.

2.10.5 Multiple Alignment

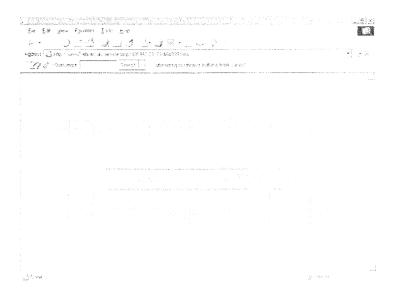
Make a Multiple alignment by putting the sequences (only Exons) of Maize, oryza sativa and Sorghum bicolor phosphenol pyruvate carboxylase in Clustal W.



ClustalW Submission Form

2.10.6 Run Clustalw

Click RUNCLUSTALW button, which is present on CLUSTALW. Job output will be displayed which is given below.



2.10,7 Jahiew

Jab iew will obtain after the job output by chicking on RUNCLUST VI Wilston.

Ele Edi View Favorites	Icol: Line	
	346 2.38.900%	
Address () http://www.2 ebs ac.u	ik/servicestmp/976177 634959-82798 html	
Y? & - Customize	Search	•
		<u>♦</u>
		-
Your Clustal	W Poculto:	

Circ JalView: JaNiew SUBMIT ANOTHER JOB

Pairwise Scores:

CLUSTAL W (1.01) Multiple Sequence Alignments

CLUSTAL W (1.81) Multiple Sequence Alignments

Sequence format is Pearson 2339 bp Sequence 1: MAIZE Sequence 2: ORYZA 2878 bp Sequence 3: SORGHUM 924 bp Start of Pairwise alignments Aligning... Sequences (2:3) Aligned, Score: 11 Sequences (1:2) Aligned, Score: 74 Sequences (1:3) Aligned, Score: Guide tree tile created: /net/mfs0/voll/production/w3nobody/tmp/976177.634959-82798.dnd1 Start of Multiple Alignment There are 2 groups Aligning... Score: 35531 Group 1: Sequences: 2 Group 2: Delayed Sequence:3 Score:8699 Alignment Score 16642 CLUSTAL-Alignment file meated [/net/nfs0/vol1/production/w3nobody/tmp/976177.634959-82798.ain) Your Multiple Sequence Alignment:

CLUSTAL W (1.81) multiple sequence alignment

MAIZE		
ORYZA	ATGGAGCGGCACCAGTCGATCGACGCGCGCGCGCGCGCTGCTCGCGCGCG	60
SORGHUM		
MAIZE		
ORYZA	GACGACGACAAGCTCGTCGAGTACGACGCCCCCCCCCGACCGCCTCCTCGACATCCTC	120
SORGHUM		
MAIZE	GTCCAGGAGTGCTACGAGGTGTCGGCCC	
ORYZA	CAGGACCTGCACGCCCTCACCTCCGCGAATTCGTGCAGGAGTGCTACGAGCTGTCGGCG	180

SORGHUM		
MAIZE ORYZA SORGHUM	GACTATGAGGCAAAGGAGACACGACGAAGCTGGGCGAGCTCGGCGCCAAGCTCACGGG GAGTACGAGAACGACAGGGACGAGGCGUGGCTCGACGAGCTUGGGAGGAAGUTCACCAGC	87 240
MAIZE ORYZA SORGHUM	CTGGCCCCCGCCGACCCTCCTGCGCGCGCGCCCCCTCCCGCCCCGCCCACCCCCCCC	147 300
MAIZE ORYZA SORGHUM	AACCTGGCCGAGGAGGTGCAGATCGCGCCACCGCCGCGACAGCAAGACGCAAGAAGGT AACCTCCCCGAGGAGGTGCAGATCGCGCACCGCCGCCGGATCAAGCTCAAGCGCGGG	
MAIZE ORYZA SORGHUM	GGGTTEGECGACGAGGGGCTECGCCACCAGCGCGACATUGAGGAGACGCTCAAGCGC GACTTEGEUGACGAGGCCTECGCGCCCACCGACTEGGACATEGAGAGACGCTCAAGCGC	267 417
MA12E ORY2A SORGHUM	CTCGTGTCCCAGGTCGGCAAGTC	
MA (ZE ORYZA SORGHUM	CCCCG- GTTGACCTCGTGTTCACGCCGCATCCAACGCAGTCCGTCAGGAGGTCCCTGCTCCAGAAG	
MAIZE ORYZA SORGHUM	GATCCGGAATTGTCTGACUCAGCTGAATGCCAAGGACATCACTGACGACGAC CACGGGAGUATCCCGGAATTGCCTGAGGCAGTTATATGCGAAAGACATCACTGCTGATGAC	347 597
MAIZE ORYZA SORGHUM	AAGCAGGAGCTCGATGAGCCTCTGCACAGACCGAAGCAGCCTTCAGAACCGATGAA AAGCAGGAGCTTGATGAGGCCCTGCAGAGGGGAATTCAAGCAGCTTTCAGAACTGATGAA	
MALZE ORYZA SORGHUM	ATCAGGARGGCACAACCCACCCCCCGGGACGAAATGCGCTATGGGATGAGCTACATCCAT ATCCGCACGACTCCTCCCCACTCCTCAGGATGAAATGCGCGCCGGGATGAGTTATTTCAT	
MAIZE Oryza Sorghum	GAGACTGTATGGAAGGGCGTGCCTAAGTTCTTGCGCCGTGTGGATACAGCCCTGAAGAAT GAAACTATATGGAAGGTGTACCAAAGTTCTTCCCCCCGCATTGATACTGCTCTGAAAAAT	527 777
MAIZE ORYZA SORGHUM	NTCGGUATUANTGAGCGCCTTCCCTACANTGTTTCTCTCATTCGGTTCTCTTTGGATG ATTGGGATTANTGAGCGTCTTCCTTACAATGCTCCTCCATUCAGTTCTCATCCTGGATG	
MAIZE ORYZA SORGHUM	GGTGGTGACCCCGATGGAAATCCAAGAGTTACCCCGGAGGTGACAAGAGATGTATGCTTG CGTGGTGACCCTGATGGAAACCCAAGAGTTACACCAGAGGTTACAAGAGATGTATGCTTG 	
MAIZE ORYZA SORGHUM	CTGGCCAGAATGATGGCTGCAAACTTGTACAATGGATCGAAGAGCTGATGTTGAG TTGGCAAGAATGATGGCTGCTAACCTGTACTTCTCTCAGATAGAAGATCTGATGTTTGAG	707 957
MAIZE ORYZA Sorghum	CTCTCTATGTGGCGCTGCAACGATGAGCTTCGTGTCGTG	

MAIZE OPYZA SOPGHUM	TCTGGTTCCAAGTTACCAAGTATTACATAGAATTCTGGAAGCAAATTCCTCCAAACGAG TCCAGAAAAGCTGCAAAGCACTATATAGAATTCTGGAAGCAAATTCCTCCAAATGAG	827 1074
MAIZE ORYZA SORGHUM	CCCTACCGGGTGATACTAGGCCATGTAAGGCACAAGGTGTACAACACGACGCGGGGGGC CCTTATCGTGTCATACTTGGTGGTGGTGAGGGATAAATTGTACTACGCUTGAACGTAGT	
MAIZE ORYZ A SORGHUM	CGCCATCTGCTGCTTCTGGAGTTCTGGAATTTCAGCGGAATCGTCATTTACCACTATC CGCCATCTATTGACAACTGGAGTTTCTGAAATTCCAGAGGAGGAACGTTTACTAATGT	
MAIZE ORYZA SORGHUM	GAAGAGTTCCTTGAGCCACTGAGCTGTGCTACAAATCACTGTGTGACTGCGGCGACAAG GAAGAGTTTCTGGACCGCTGAGCTGTGCTACAGATCATTATGTGCTTUTGGTGACAAA	
MA1ZE ORYZA SORGHUM	GCCATCGCGGGAGGGGGGCCCCCCCGGGCCGGCGGGGGTTTCACGTCCGGGCTCCC CCTATAGCTGATGGAAGCCTTCTTGATTTCTTGCGCCAAGTATCAACTTTGGGCTGGGT	
MAIZE ORYZA Sorghum	CTGGTGAAGCTGGACATCCGGCAGGAGTCGGAGCGGACACCGACGTGATCGACGCCATC CTTGTAAAACTTGACATAAGGCAGGAGTCTGATCGACACACTGATGTCCTTGATGCCATA	
MAIZE ORYZA SORCHUM	ACCACGCACCCCGGCATCGGGTGTACCGCGAGTGGTCCGAGGAGAAAAGCGCCAGGAGTGG ACTACATATCTTGGGATTGGATCTTACGCTGAATGGTCTGAGGAGAAAACGCCAGGATTGG	
MAIZE Oryza Sorghum	CTGCTGTCGGAGGCTAGCGCGCGCGCGCGCCGCGGGCCTCCCCCGGACCTCCCCGGACCGAG CTTTTGTCCGAACTAAGGGGCAAACGTCCTTTGTTTGGTCCTGATCTTCCTCAGACTGAA	
MA12E ORYZA SORGHUM	GAGATCGCCGACGTCATCGGCGCGTTCCACGTCCTCGCGGAGGTCCCGCCGGACAGCTTC GAAATTCGTGATGTTTTAGGAACATTTCACGTCCTTGCTGACGTCCCGGCAGATTGTTT ATGATTGCGGACCTGAAGAGGACTACGAGAAGGCGCGGAGATCGGTCGCGCGCG	1554
MAIZE ORYZA SORGHUM	GGCCCCTACATCATCTCCATGGCGACGGCCCCCGGGACGTGCCCCGGGGGGCCCCTG GGTGCTTACATCTCAATGGCAACTGCACCATCTGATGTGCTTCGTGGGACGTTCTA GGGGTGCTTCCACCGCTGCACGTCCCGCGUCACCGGCGACGCGTTCGCCGTCAAGTCCGTG	1614
MATZE ORYZA SORGHUM	CAGCGCGAUTGCGGCUTGCGGCCAGCUTGCCCG-TGUTUCCGCTGTTCUAAAGGCTG CAGCGGGAGTGCCATATAAAACAGCUTCTGCGAG-TTUTTCCCCTATTTUAGAAACT GACUUTCGCAGCTGUCCGACGACCTGUACCGGAUCTCGCGGAGCTGUAGCCCAAGCTG	1671
MAJZE Oryza Sorghum	GCCAGCUTGCAGTCUGCGCCCCGCGTCCGTGGAGCGCCTCTTCTCGGTGGACTGGTACATG GCAUATCTTGAACCAGCTCCAGCAGCGGGGGCGCCGCGCACTTTTTCAATAGACTGGTACATG GCGCAGCUTCGCCCCGCGGGGAACCCCCGGCGTGUTGCAGACGCACUCGGTG-TACGAG	1731
MAIZE ORYZA SORGHUM	CACCEGATCAAGGGCAAGCAGGCAGGTCATGGTCEGCTACTCCEACTCUGGCAAGGACGC- AATAGGATTAATGGCAAGCAGGAGGGTCATCATTGGATACTCAGACTCTGGCAAGGATGC- GACEAGAGCGTGGACCCACAGGTGATGGACCTCTGCEGGGGCCCGGACCTGCTCGAG	1790
MATZE ORYZA SORGHUM	CGGCUCCCTGTCCGCGGCGGCGCGCGTACAGGGCGCAGGAGAGAGA	1850

MAIZE ORYZA SORGHUM	CAAGEGETACGGEGTEAAGETEACETTGTTECAEGGEGEGGGAGGEACEUTGGGEAGGG AAAGEATTATGGTGTAAAGTTGAEAATGTTECATGGAAGAGGTGGAACAGTTGGEAGAG- GEAGGTEGEECEAGGEGUTEGEGETETGEEACEGEEGGEGGUGEECAEEGUEAEGT	1909
MAIZE ORYZA SORGHUM	GTGGCGGGCCCACGCACCTTGCCATCCTGTCCCAGCCGCCGGACACCATCAACGGGTCCA GAGCTGGTCCCAGTCATCTT GCCATATTATCTCAACCACCAGACACGATACATGGATCAC AAAGCCCGACACATCCTCATCGACACCGCCCCCGCAGGGAGGG	1969
MAIZE ORYZA SORGHUM	TCCGTGTGACGGTGUAGGGCGACGTCATCGAGTTCTGCTTCGGGGAGGAGCACCTGTG TTCCTGTAACACTAUAGGCCGAGGTTATTGAACAUTCGTTTGGAGAGAGAACACTGTG GGAAGAGGACAGCGGCGACGCGAGGACGGCGGGCGCGCGCGCGGCG	2027
MAIZE ORYZA SORGHUM	CTTCCAGACTCTGCAGGCTTCACGGCCGCCACGCTGGAGCACGGCATGCACCCGCCGGT CTTTAGAACTCTGCAGGCTTTACTGCAGCTACTCTTCAGGATGGAT	2087
MAIZE ORYZA SORGHUM	CTCTCCCAAGCCCGAGTGGCGCAAGCTCATCGACGAGATGGCGTCGTGGCCACGGAGGA TTCCCCCAAACCAGAATGGCTGCTCTAATCGATGAGAGGCGCTGTTGTGGCAACAAAGA GACGCCCCACTACGTGGCGCCCCGAGGTGGTGGGCGGCGGCGAGTACGGC-GGGAAGGC	2147
MAIZE ORYZA SORGHUM	GTACCGCTCCGTCGTCGACGACGACGCCGCCTTCCGCGACTACTTCAGATCGGCTACACC ATATCGATCATCGACGTCTTCAGGACCACGCTTTGTGATACTTCCGATCGACGACGACACAC GGACGTGTGGACGCCGCCGTGGTGATGTACGCGTGCTCTCCGGCGGCGGCGCCGCCGTGT	2207
MAIZE ORYZA SORGHUM	GGAGACCUAGTACGUGAGGATGAACATCGGCAGCUGGCAGCCAAGAGGAGGCUCGGCG TCACACAAATATGGCAGGATGAACATTGGTAGCUCGACCATCAAGAGAAAGACAACGCCAGTGG CGCGGGCGAGAACGCCGGCUGAGGTGCTAUCGGCGGCCAUCGTCCGGGTCC	2267
MAIZE ORYZA SORGHUM	CGGCATCACGACCCTGCGGCCATCCCCGGATCTTCTCGTGGACCCAGAC-CAGGTTCC TCGCATTGAATCGCTCCGTGCAATTCCTTCCATTTTGCTTGGACACAAAC-TAGGTTC CCCCAAGGCGTGTTCAGCGGGGTGTCCCCCCGGCCAAGGACCTGTTGAGGCGCATGATCT	2326
MAIZE ORYZA SORGHUM	AUCTCCCCGTGTUGGTGGGGGGGGGGGGGGGGGGGGGGGG	2386
MAIZE ORYZA SORGHUM	GGAACTTCCAGGTCCTCAAAGAGATGTACAACGAGTGCCCATTCTTCAGGGTCACCTGG GGAACATCCATACACTGAAAGAAATGTACAATGCAGTGGCCATTCTTCAGGGTCAACCTGG TGAGCGGCGGAGGAGCACGGGCAATCGAGCGGCCAACCTGA	2446
MAIZE ORYZA SORGHUM	ACCIGCTGCAGATGGTTTTCCCCCAGGGAGACCCCCCCCATTGCCGCCTTCTATGACGAGC ACTIGCTTCAGATGGTTTTCCCCTAAGGGAGACTCCAGGAATTGCTGCTTTATATGACAAAT	
MAIZE ORYZA SORGHUM	TGCTTGTGGGGGAGAAGAACTCAAGGCCTTTGGGAAGCAGCTCAGGGACAAATACGTGGAGA TGCTTGTGGGGGGGTGTCTGGCAATCCTTTGGGGAGGAGCAGCTGGAGAAGCAACTTTGAAGAGA 	
MAIZE ORYZA SORGHUM	CACAGCAGCTTCTCCCCCAG	
MAIZE ORYZA SORGHUM	TGAAGCAGCGTCTCCGGCTGCGTGAGTCGTACATCACTACCTTGAACUTTTCCCAAGCCT	2686

MAIZE ORYZA SORGHUM	ACAUCETGAAGEGGATAAGGGACCETAGUTTECGAGGTGATGTEGGGAGEESGUETTGTEGA 2746
MAIZE ORYZA SORGHUM	AGGAGTTCGTCGACAGCAACCAGCCTGCCGAGCTGGTCCAACTGAACGCTGCGAGCGA
MAIZE ORYZA SORGHUM	ACCCACCTGGCCTGGAGUACACCCTCATCCTGACGATGAAGGGCATTGCTGCCGGCATGC 2866
MAIZE ORYZA SORGHUM	AUAACACGCT 2876

Your guide tree:

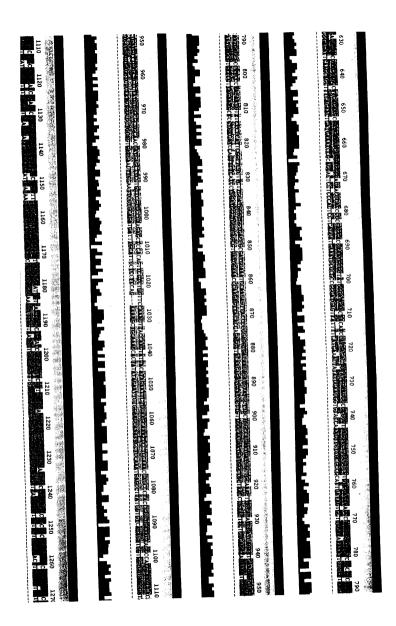
(MA1%E:0.03716, ORYZA:0.01465, SORGHUM:0.669551;

5

2.10.8 Alignment Graph

An alignment graph among Maize, oryza sativa and Sorghum bicolor phosphenol carboxylase gene will obtain.

90 90 100 120 130 140 159 E-TE-TO-WETA-GE-CETTETTO-ELALATECTTCA-ELALATEGE-CALATER 6 700 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 200 300 310 2.00 2.00 2.00 2.00 2.00 2.00 300 310 2.00 2.00 2.00 2.00 2.00 300 310 2.00 2.00 2.00 2.00 2.00 300 310 2.00 2.00 4.00 4.00 4.00 4.00 4.00 4.00 2.00 4.00 4.00 4.00 4.00 4.00 4.00 4.00 4.00 4.00	320 330 340 350 350 370 380 390 400 400 420 430 440 450 460 470 320 320 340 350 350 370 380 390 400 400 420 420 420 440 450 450 470 320 320 320 340 350 350 370 380 390 390 400 400 400 400 400 400 400 400 400 4	Hahrer algment edno Fre Edi Fort View Caku Cakude Algn Hab 10 20 30 40 55 50 70 80 90 100 110 120 130 140 150 ALEZAA 1982 ALEZAA 1982 ALEZAA 1982 ALEZAA 1982
100 110 120 CALCOUTTECTER ACATECTER AND CALCOUTTECTER ACATECTER AND CALCOUTTECTER ACATECTER AND CALCOUTTECTER ACATECTER AND CALCOUTTER AND		80 90 Errorokara. Geocrecoro

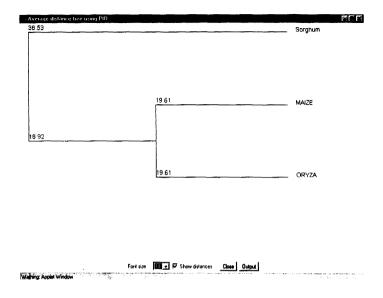


A MAGERAMETAENTATGETEGGTACTECEGATECEGEAGEACEC-CEECECETETECEGEGEGEGEGEGEGEGEGEGEGEGEGEG	1750			150 1600 1510 1500 1500 1500 1500 1500 1		1430 1440 1470 1470 1480 1590 2020AJAKTROGUTUTUTUTUKAADUSCAALOTAUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUS			алороные система и торитории и из сооторонности от	
TAXO					Ŀ	GGAG			GAGE	
22	5		EXECTION OF T	16 ACGT ATGT AGGC		1440 TGGCTG TGGCTT		2ma	1996	
5	50	1000	Constant.	GIGCIO GIGCIO GIGCIO	E	1160		5	116	
1				19000 19000		1010		-		
í.	1770		CONSULT TRANSPORT	1610 STGGAC STGGAC		1450 20AGC 70AAC			etton etton ostri	į.
ŝ				0 ACTI ACTI	*	CIBC			0070	
1			inin Kanada			6000			1000	
2	1780			1620 PGCAGE CACAGE		1460 360AA			NALVIC VALVIC	
Į.			ACCESSION OF A			1001			ANCT CALCO	
ġ.	1790	Control of the	Annes	1630 AGTO AGTO CGCAGC		1470 000010 001110			ACTILIA ACTILIA	
100	8		and a second	00110 00 00	E	1611			1000)
5				GCC01 ATA1		1601		171	0000	
5	1800			1640 Toceor Taana Sacgar		1430 1430	5		1330 20070	ŝ
2			1000	1000 1000 1000		3ACC		12	GIGA	
	H		10.35	1 DOTO 1 DOTO		1000		9	- 260 D	ŝ
n'n	1810			1550 1550 157660		1490 00040 07046		5.5	30 100 100 100 100 100 100 100 100 100 1	ŝ
222			and the second	8633		Acro		1	1330 Autrosontaleoces Autrosontaleoces Autrosontaleoces	
576	1820			1660 1660 1660 1660	an a	1500 MGGAG MAGAA			10000 10000 1114	i.
CAGE	0		MERCENS	5000 5000 5000	i i i i i i i i i i i i i i i i i i i	SO TO ANT			AGGA	Ì
đ	_	-		DIGI	an (1990)	1900 1900			GICT CT	
CAG	1830	-	San de la constantino Montantino	1670 TTCGA TTTGA	Condition (1)	1510 DGACGI DGATGI BGAGCI		25	CTGATO	ŝ
GCG		And a local de la constante		ANGG GANA CANG		TGAT			64 6 4	
CAGG	1340					0990 1090 1090		and a second	CACO LA	34
AGGA	40			1680 Teccac Teccec		1520 GAACAD GGGACT		in.	TAD.	
GATG				NGCT)		TTCC.		3.C3.	STOX	
6060	1850			1690 30AGT0 10AAG0 2600		1530 ACCIC ACCIC		13218	GATOGACO CETTOATO	1370
AGG			and the second	1090 1090 1090	and a second	CCLUC CCLUC			1600	
1996	-			C C C C C C C C C C C C C C C C C C C		6000 6000			ATA	
CAAG	1860			1700 DCCGCG DCAGCA DGGGCA		1540 DGAGCT DGAGCT DGAGAT		(jso	LCTAC	286
CGCI			10 Berry	ACCO	RECESSION	0000		i dana	ATA ATA	
200%	1370	-	i interest	1710 MiggAg MiggAg		1550 BCCCGA BCCAGA				1390
ŝ	ő		a and a second	0100		SACAN SACAN SGCCI		10.72		
AAG				STOCIC ACTA	1000	jaj.			100	
TCA	1680		a constant	DAGACI DITCI 1730	Coppers.	1360 Friger Friger		i in	ALCI	50
CELL			a seren	CGGI	805-029 578750044	olocu CCCL			TÁCC	
9	13			L7 AGAC	a de la comercia de l	1573 ACATCA ACATCA FCCACC		13 Base	CTGA	
CACO	0681			1733 Achiga Achiga Sachiga		CATC		R.	ATGO	¢
9000			100	FACA FACA		TCL		de la	700	
GCG6	0067			1740 TGGACI TGAATI AGGACI	and the second	1490 150 <td></td> <td></td> <td>5466. 3466.</td> <td>1400</td>			5466. 3466.	1400
AGGC	0	1.1	2793	i0 CCGG TAGG	Suppose Supposed	0 CGAC CAAC		100200 1007973	AGAA	
ACCO	1			AC - D	A STREET	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			1007	1
TOGU	0161			1750 CAAGG FAATG	Nonana	1590 CACCAT CACCAT			CNOC	8

2070	CAC6 TACT		2236 67ACG ATATG 6-ACC		2390 CAAGG. CCAAGG.	NEW COLUMN	2550 666A) 666A)	l
	000000 1002400 1000400		SACCGAL	Received and	CATCGA	ag Ma	rccrrr rccrrr	
2060	CTTCAC CTTTAC GGTGGG	REDORMAN RECORDER	2020 CCG66A(CCT6A(2.355 attract c.attata atto:-tr		2540 CTCAAGC CTGCAAT	
	SCAGCG SCAGCG SCCGTG		0 6CTACA GCAACA CTCCCC	Institution of the second	C TTCAAG TTCAAA	E.	30 62A6AA 66T6AT	÷.
2050	ACTUT		2210 2210 6470660 06470660		3570 Sécecatt Seacott	Marine Association	2530 37660AGAU	
2040	TTCCA6 TTTAGA TTCCGG	NUMBER OF	2200 GTACTTC ATACTTC		0360 SATTT660 SATTT660	BARRA BARRA	2520 607060770 0770607770	
20	CTGTGC TTGTGC 60664C		21 21 21 21 21 21 21 21 21 21 21 21 21 2	ARCON CO.	21 17699A0 17769A1 17769A1	Bark	ACGAG C ACAAAT	5
3030	GAGCAC GAACAC CGGCTG	and the second	2190 607710 6077710 86077710		2350 ristosi riktrosi	5.	3510 TGTATG	
	SGGGAG SGAGAG CGCGCG	ESELEX/ALEXA ESELEX/ALEXA ESELEX/ALEXA	AGGCCC		TCCCCG		006667	
2020	Decritor Contrati Contrati	623	2180 2180 TCAAGGA TCAAGGA		2040 Docaco Treator Doteool	BREAM	S500 S500 SATTGCT	
õ	AGTTCI AACACT AGACGG	FU	No noerror noerror		00 -CAGGT -TAGGT 9CATGA	Martin Barris	2490 6470066	
2010	TCATCG TTATTG AGGCCG	Real Property lies	2170 2170 SCTCGTCC MTCANTCC	alize a	2300 CAGAC-CA CAAAC-TA GAGOCOCA	HALAN CONTRACTOR	24 666A6A 666A6A	7
2000	oceaco oceaco cceaco		2160 2160 AGTACCC AATATCC CGGACGT		2300 GTGGACC GTGGACC	and the second se	2480 CGCCAA	-
51	CAGG CAGG GCGAGG		CGAGGA AAAGGA AAAGGA GAAGGC	H.	2 Theres Thres	Ballan Ballan	9611111 9611111	
1990	ACGGTG ACAGTA ACAGCG		2150 2150 660A00 660A00				2470 2470 56A6AT0 16A6AT0	
~	06TIGTIGI LETETAL LAGAGG		0 GGTCGT TGTTGT GAGTAC		0 ATTOCC ATTOCC	Electron (State	0 CTGCT	
1950	CCATCO CACTTO CACTTO		2140 2476666 34766676		2300 2606CCAN 06T6CCAN	BASA BARANANANAN BARANANANANANANANANANANANANANANANANANANA	2460 CTGGACC	ų
1970	ACGGGT ATGGAT CGAGGA	Constant of the second	2130 7166ACGAI 7166ATGAI		2250 26ACCETG MATEGETCI	u BEERR	2450 Bistrace Bistrace	
2	CCATCA CGATCA CGATCG	NAME OF TAXABLE	2. SCTCATI CTAATI SCT0GT(. 21 0104010 01104411		2 TTCAGG TTCAGG	No.
1960	DGGACA	angle [1]	3120 500040 5000607	All I	3280 6000028 6000028	Bernardar occup Bernardar occup Bases	2440 DCATTC DCATTC DCAACC	Cencius Section
	NGCCGC ANDCAD CDCAGD	E CONTRACTOR	GAGTGG GAATGG GTGGCG		CCGGCD CCGGCD CTAGTC CGGTTC	ALC: NO.) 5ÅGTGG(5ÅGTGG(
1980	ercco. Arctro. 60000	tituesta bio	2110 2110 AAGCCCG AAACCAG		2270 56A6600 6AAA6000 A600T00	10000 (0000) 10000 (0000) 10000 (0000)	2430 2430 ACAACG	
ç	CATCCI CATATT GACAC	R. C.	2100 41570000 117700000		so CCAAGA CGAAGA CGGGGC		2420 AGAGATOT AGAGATOT AGAAATOT	
10.00	CCTTOC TCTTOC CTCATC	No. 1	21) 2006/101 0004/117 516666		2260 SGCCAGGC SACCATCAU STGCTGCG	and the second s	24 TCAAAG TCAAAG	
1020	CACGCA		2090 EACCC6 EATCCT0 GCCT60		0.2250 BGCAGCCC BGTAGCCCC		2410 AGGTCC ATACAC	konstar Silili
-	GGGGCC GGGCCC CCGGAC		GCATGC GCATGC GAATGC CGGAAC	- 1	CATCOC		KCTTCC VCATCC SCGCG	Status and a state
1920	orboeckeet arteeceeecökkeekerrarrateekarteekeekeekeekeekeekeekeekeekeekeekarrateekarrateekeerroookeekeekarrateekeekerrateekeerrateekeerrateekeekeekeerrateekeekee Barroeckaekeekareerrateekeekarratrateekeekeekeekeekeekeekeekeekeekeekeekeek		2210 2000 2009 2009 2100 2110 2110 2120 212		230 2240 2250 2260 2270 2280 2290 2290 2290 2300 2310 2310 2320 2330 2340 2350 2360 2370 2390 2370 2395 илтовекователатовескотоликоваевсесовоескатолиских: стороваетсествелитороваескала 240000000000000000000000000 и Котовекователа сторовае соста на обеската стороваетсествелитороваетсеска соста 2350 2360 2370 2390 2390 239 И - Котовекователа соста на соста соста стороваетсествелитороваетсествели стороваетсе соста и 2350 2360 2370 239	Constant References References	2360 2400 2410 2420 2430 2433 2446 2450 2450 2470 245 420 2400 2510 2500 2510 2520 2520 2540 2551 2551 2551 2551 2551 2551 2551 255	
	GCAGGG GCAGAG GCAGAG		Decres, Section		Ceseko Descker	ABLES CONTRACTOR	G GGACGT GGATAT GTGGAT	
1010	Serres Kerres Xerres	Listance and	2070 2070 600A06 600A06 600A06	BRANCE OF CONTRACTOR	2230 NGTACI NATATI	tenetre ^{est} in andre statistica	2396 AGAAGG AGAAGG	Betweeting

J11. Yr.) J17. J27.	
June Set June Set June Set June Set June Jun	
Lunio Studi Lunio Lunio <th< td=""><td></td></th<>	
Lunio Studio Studio </td <td></td>	
0 212 34.4 -411 -442 255 267 0011040 0 0 0 0 0 0 0 0 212 2770 275 2730 2600 2310 2600 0 212 2770 275 2730 2600 2310 2600 0 212 2770 275 2730 2600 2310 2600 0 2120 2770 275 2730 2600 2310 2600	
0000 011 041 041 041 041 0001040 000104 041 041 041 041 0001040 000104 0401 041 041 041 0001040 000104 0401 041 041 041 0 012 0141 0140 041 041 0 012 0141 0140 041 041 0 012 0141 0140 041 041 0 012 0141 0140 041 041 0 012 0141 0144 044 041 0 012 0141 0144 044 044 0 012 0141 0144 044 044 0 012 0141 0144 0144 0144 0 0144 0144 0144 0144 0144	
00100 011 054.0 411 421 259.1 001000 0011 054.0 411 421 259.1 011000 0011 1010 1011 1010 1011 0 112 2110 2134 1013 1010 0 112 2110 2134 1013 1010 0 112 2110 2134 1013 1010 0 112 2110 2134 1013 1013	
0 1150 011 1544 - 111	
0011040 - 2110 - 2514 0011040 - 2110 - 2514 0 - 2132 - 2370 - 2710 - 2132 - 2370 - 2710	
551 304000454 Alacturates 0011011	
2700 AMTR/0710640 AMTR/0710640 AMTR/0710640 AMTR/0601 2720 001101 100000 110000	FI
ия. 211 211 211 211 211 211 211 211 211 21	
2510 06,6462A0240 06,6462A024 06,060414400 2510 2510 2510 2510 2510	

2.10.9 Phylogenetic Tree Construct the phylogenetic tree among Maize, oryza sativa & Sorghum bicolor phosphenol pyruvate carboxylase.



2.10.10 Tree Analysis

Check the two species which having maximum alignment By tree Analysis. Take the species, which have maximum alignment. By Tree Analysis it has been found that there is maximum alignment between Maize and oryza sativa phosphenol pyruvate carboxylase gene.

2.10.11 Design Primers For The Sequences Of Maize:

Design two sets of Left and Right Primers for the exon regions of maize sequences by using PRIMER3. One set of primers are sufficient for PCR, but the second set is taken as additional primer in the event the first fails hybridize.

2.10.12 Selection Of First Set Of Primers

Design left primer from one Exon of the sequence and right primer from another Exon of the sequence. The distance between the two primers should be about 500 to 1000 base pairs. Take the sequence from the species, which has both Exons and introns.i.e we have taken Maize sequence because it has both Exons and introns. But rice sequence for this enzyme has only coding regions i.e.Exons.so we can't design primer from rice for sequencing of DNA but can be used as a marker.

2.10.12.1 Primer3 For Left Primer

Take one Exon sequence of the maize and then paste it in PRIMER3 and pick primer. The region of the exon from which the left primer is designedare 4454-5452. The left primer is picked from the Exons, which is right of the exon from which the right primer is picked. The parameters for both the primers should be same.the product size should be Min: 100.Opt: 200,Max: 1000.Primer size should be Min: 18bp.Opt: 20bp.Max: 27bp.Primer annealing temperature should be Min: 57.Opt: 60,Max: 63 degree celcius. The GC% should be Min: 20and Max: 80.

2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	
Primers Antonication procession and a	
Primero and Antoritory pick of the a	
ామానించి వాయాలి మీడామారి మాడిప్రి చాడికారి ప్రారంగా 10	
The sequence consists to be a first of addition of a first sequence of the first sequence of a first sequence of the first sequence of a first sequence of the first sequence	
化化化合物 化分配 化化化化化合物 化化化化合物 化化化合物 化分析 化合物 化合物 化合物 化合物 化合物 化合物 化合物 化分析 化分析 化合物	
ELA temporative response for an an information of a first organization of a first organization of a second seco	
tiφ - − − − − − − − − − − − − − − − − − −	
šuok Pick šuve. 1ett Nyktodzestan oljant	
originar product primer	
้งที่ โระที่เคียงต้ ไว้ร่ เสียง เป็นสู่สุดิ เชื่อ	
ತಿಂದರೆ ನಾವು ಸಂಪ್ರದೇಶದ	
学生主席(Mar Angle) (学生)(1997) 全主義(Mar Angle) (1995) (1997)	
And the second	
p	
$\frac{1}{1 - \frac{1}{2} \left(\frac{1}{2} \int \frac{1}{2^{2} - \frac{1}{2} \left(\frac{1}{2} \int \frac{1}{2^{2} - \frac{1}{2^{2} - \frac{1}{2} \int \frac{1}{2^{2} -$	
Conservation and a second se	
4. 3 a	
- Control of the second sec	
χ_{1} is the second	
$\hat{F}_{\mu\nu} = e^{2\pi i \frac{1}{2}} \hat{F}_{\mu\nu} = e^{2\pi i \frac{1}{2}} + e^{2\pi i \frac{1}{2}} e^{2\pi i \frac{1}$	
General Primer Ficking Conditions	
Maria Print Ara Print Basa Print	
$\frac{1}{2} \sum_{i=1}^{n-1} \frac{1}{2} \sum_{i=1}^{n-1$	
and a second	
. 1 957 (1991)	
s 1 - Jär	
Production and the second and the se	
Other Per-Sequence inputs	
Special Till 2011 product presenter unter Paris (* 1997) Reductor alternational paris principal (* 1997) (* 1997) Reductor alternative (* 1997) Reductor alternative (* 1997) Reductor alternative (* 1997) (* 1997) Reductor alternative (* 1997)	8-18-11 V
g - 2 2 2 - 2 2 2 2 4 2 4 4 4 4 4 4 4 4 4	

Objective Function Penalty Weights for Primers

```
L to 2 (1777) Att 5 (1777)

to 7 (1777) Att 5 (1777)

L to 2 (1777) Att 5 (1777)

Att 5 (1777) Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5 (1777)

Att 5
```

Objective Function Penalty Weights for Primer Pairs



Hyb Oligo (Internal Oligo) General Conditions

an The New Trans 28 Miles Mean Alfred 3 No. 11 Marco Miles	
1	
the second se	
NT X	

End Entron Sciences Contest Sciences

Objective Function Penalty Weights for Hyb Oligos (Internal Oligos)

	;	(C) (77) (77)	 <u>, </u>	
				304
				1. 1992 - 2004 2012 - 2012 - 2014 2014 - 2014 2014 - 2014
C. 10. 5 . 5				

2.10.12.2 Primer3 Output For Left Primer.

Put the sequence of Exon of the maize for the left primer in the primer3 output and by clicking pick primer the left primer and right primer both will be obtained. But, left primer only has to be taken into cosideration. Primer3 output for left primer are given below.

Primer3 Output

WARNING: Numbers in input sequence were deleted. No mispriming library specified Using 1-based sequence positions OLIGO 663 20 60.83 55.00 7.00 1.00 cgtcaageteacettattee LEFT PRIMER RIGHT PRIMER 843 20 60.33 55.00 7.00 0.00 etgeagagtetggaageaca SEQUENCE SIZE: 999 INCLUDED REGION SIZE: 999 PRODUCT SIZE: 181, PATR ANY COMPL: 4.00, PATR 3' COMPL: 3.00 1 tteettgagecacttgagetgtgetacaaatcaetgtgtgactgeggegacaaggecate 61 geggaegggageefeetggacolootgegeeaggtttteacqltegggeteteeetgqtg 121 aagetggaeatcoggeaggagteggageggeacacogaegtgategaegeeatcaceaeg 181 cacciegcateggglegtaccgegggtggtccgaggacaagcggeaggaglggetgetg 241 teggagetgegaggeaagegeegetgetgeeegggacetteeeeagaeegaggagate 301 geogacyteateggegegttecacytectegggageteecgeeegacagetteggeeee 361 tacatestetecatggcgacggccccctcggacgtgctcgccqtggagctcctgcagcgc 421 gagtgcggcgtgcggccagccgtgcccgtggtgccgctgttcgaaaggetggccagcctg 481 cagteggegeccgcgtccgtggagegeetettetcggtggactggtacatggaceggate 541 aagggcaageageaggtcatggteggetacteeggcaaggacgeeggeetg 601 teegeggegtggcagetgtacagggcgcaggaggagatggcgcaggtggccaagegetac 661 ggcgtcaagetcacettgltccacggccgcggaggcaccgtgggcagggg1ggcgggccc 721 acgcaccttgccatcctgtcccagccggccggacaccatcaacgggtccatccgtgtgacg

**************** 841 cagegettcacggccgccacgctggageacggcatgcacccgccggtetetecccaageee <<< 901 gagtggegeaageteatggacgagatggeggtegtggeeacggaggagtacegeteegte 961 gtcgtcaaggaggcccgcttcgtcgagtacttcagatcg KEYS (in order of precedence): >>>>>>>>>>> left primer <<<<< right primer ADDITIONAL OLIGOS <u>1941 – 1940 – 1970 – 1970 – 1980</u> 1. 1 LEFT PRIMER 662 20 60.99 55.00 7.00 3.00 gcgleaageteaeettgtte 843 20 60.33 55.00 7.00 0.00 etgeagagtetggaageaea RIGHT PRIMER PRODUCT SIZE: 182, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 3.00 2 LEFT PRIMER 662 20 842 20 60.99 55.00 7.00 3.00 gegteaageteacettgtte 60.33 55.00 7.00 1.00 tgcagagtctggaagcacag RIGHT PRIMER PRODUCT SIZE: 181. PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 3.00 3 LEFT PRIMER 665 20 58.85 50.00 7.00 1.00 teaagetcacettgttecae 843 20 60.33 55.00 7.00 0.00 etgeagagtetggaageaca RIGHT PRIMER PRODUCT SIZE: 179, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 3.00 21 59.76 52.38 7.00 1.00 gtcaagetcacettgttccac 4 LEFT PRIMER 664 843 20 60.33 55.00 7.00 0.00 ctgcagagtctggaagcaca RIGHT PRIMER PRODUCT SIZE: 180, PAIR ANY COMPL: 5.00, PAIR 3' COMPL: 3.00 Statistics in in con too tm tm high high high no sid many tar excl bad GC too too any 3' poly end get reg GC: clamp low high compl compl X stab 0 0 770 0 147 6961 0 85 0 75 0 0 770 0 85 7041 0 32 0 82 ered Ns ok Left 8334 0 276 Right 8279 0 269 Pair Stats: considered 1342, unacceptable product size 828, high end compl 82, ok 432 primer3 release 0.9

(primer3_www_results.cgi v 0.2)

note down the left primer from the above PRIMER3 output.

2.10.12.3 Primer3 For Right Primer

Pick the next Exon region sequence after the left exon region from which the left primer was designed for the right primer. The right primer from this region is designed as descbried about for the left primer in Primer 3 but only the right primer is to be considered from the output for right primer are given below. The parameters are similar to that of PRIMER3 for left primer. the exon region, which are taken to design the right primer are550-5936.

a a se a se constante en la constante en la constante de la constante de la constante en la constante en la con La constante en 器 Hanna (1) from the source of t Firine r 3 gr 4 Lucki for Rado Lucki forek (de Norgensler Crean) Mille engenten Rade engenten Rad Полности и полности по по по полности по по по по 19 2010 2010 2010 19 新たした 第五日代に 新五日代に 新五日代に 「小小小 「小小小 「小小小」 「小小小 」 ्रियो केल्ड्स हिंदा 2 11 PS-243 A start of the second s Seneral Primer Picking Conditions . . Sther Per-Sequence Insute

Objective Function Penalty Weights for Primers

 $\begin{array}{c} \Delta t : \left\{ \overline{\beta}, \overline{\beta} \right\} & \forall \sigma : \left[\overline{\Delta t} \right] \\ \forall r : \left[\overline{\beta}, \overline{\beta} \right] & \forall \sigma : \left[\overline{\beta}, \overline{\beta} \right] \\ \forall t : \left[\overline{\beta}, \overline{\alpha} \right] & \overline{\sigma} : \left[\overline{\beta}, \overline{\alpha} \right] \\ \hline \overline{\beta}, \overline{\alpha} \\ \overline{\alpha}, \overline{\alpha} \\ \overline{$

Objective Function Penalty Waights for Primer Pairs

No: Con Sector 4.5

Hyb Oligo (Internal Oligo) Fer-Sequence Imputs

Hyb Oligo (Internal Oligo) General Conditions

Каз ^{3.4} Иля (бр.). Маха (бр.)	Sec 201	Max 3 x	
	5- 5-		Politik Prime Politik Politik
	р 74 ж		gent of the state of 17 of 19

Flesh Filteration of French Fleron

Objective function Pessity Weights for Hyb Oligos (Internal Oligos)

2.10.12.4 Primer3 Output For Right Primer

By clicking pick primer we will get the right primer and left primer both. but right primer only has to be consider.primer3 output for right primer are given below.

Primer3 Output

WARNING: Numbers in input sequence were deleted. No mispriming library specified Using 1-based sequence positions 1 20 1 OLIGO 60.28 60.00 4.00 3.00 getacaceggagacegagta LEFT PRIMER RIGHT PRIMER 197 20 59.70 50.00 7.00 3.00 tggaagtteetgaegteett SEQUENCE SIZE: 387 INCLUDED REGION SIZE: 387 PRODUCT SIZE: 197, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00 61 cooggeggeggeateacgaccetgegegeeateccetggatettetegtggacceagace 121 aggliceacctcoccgtgtgggtgggagloggcgcogcattcaagttcgccalogacaag 181 gacgtcaggaacttecaggtcctcaaagagatgtacaacgagtggccattettcagggtc eccercecceccecce 241 accetggacctgctggagatggtttttcgccaagggagaccccggcattgccggcttgtat 301 gacgagetgettgtggcagaagaactcaagceetttgggaagcagetcagggacaaatac 361 gtggagacacagcageitetectecag KEYS (in order of precedence): >>>>>> left primer <<<<< right primer ADDITIONAL OLIGOS and the second 1 20 60.28 60.00 4.00 3.00 getacaceggagacegagta 1 LEFT PRIMER RIGHT PRIMER 196 20 59.70 55.00 7.00 1.00 ggaagtteetgaegteettg PRODUCT SIZE: 196, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00 174 20 60.30 55.00 6.00 1.00 cgacaaggacgtcaggaact 2 LEFT PRIMER RIGHT PRIMER 377 20 59.65 55.00 5.00 2.00 agetgetgtgtetecaegta PRODUCT SIZE: 204, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 2.00 20 60.28 60.00 4.00 3.00 gctacaccggagaccgagta 3 LEFT PRIMER 1 193 20 60.30 55.00 6.00 2.00 agttectgacgtecttgteg RIGHT PRIMER PRODUCT SIZE: 193, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 2.00 4 LEFT PRIMER 186 20 60.23 55.00 7.00 3.00 caggaacttccaggtcctca 186 20 377 20 59.65 55.00 5.00 2.00 agetyctgtgtctccacgta RIGHT PRIMER PRODUCT SIZE: 192, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 0.00

Statist	LICS												
	con	too	in	in		no	tm	tm	high	high		high	
	sid	many	tar	excl	bad	GC	too	t.00	any	3'	poly	end	
	ered	Ns	get	reg	GC -	clamp	low	high	compl	compl	х	stab	ok
Left	2541	0	0	0	86	0	144	1972	0	10	0	41	288
Right	2438	0	0	0	7	0	224	1728	1	8	0	50	418
Pair St	tats:												
conside	ered 22	4, unad	cepta	ble pro	oduct	size 1	59, hi	gh en	d_comp	1 13, o	k 52		
primer:	3 relea	se 0.9		-									
•													

(primer3_www_results.cgi v 0.2)

Note down the right primer from the above PRIMER3 output.

2.10.12.5 Calculation Of Product Size

Left primer sequence size is	999
Left primer starting point is	663
Left primer total length is	336
Introns length is	98
Right primer sequence size is	387
Right primer ending point is	197
Total right primer length is	190

Product size = 1.eft primer total length is (336) +introns(98) +Right primer ending point is (197) = 631

2.10.13 Selection Of Second Set Of Primers

Repeat the same process as explained in the first set of primers. Set the parameters for designing the primers for PCR.

2.10.13 .1 Primer3 For Left Primer

The parameters set to design the left primer for second set of primers are: the product size should be Min: 100,Opt: 200,Max: 1000.the primer annealing temperature is Min: 49, Opt:51, Max: 54.the GC% isMin : 20 and Max: 80.the primer size is Min: 18bp,Opt: 20bp,Max: 27bp. The exon regions which are taken to design the left primer are 3074-3296.

A subjective of these is even in the second se		
the Batt Cons Reported Ser	0. 249	
		an an an the state
	the second s	and the second resident the
$\label{eq:constraint} = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} \sum_{j$	n an	and a set of the set
Frine r?		
The second factor and the	general and a state of the second	
The second se		
TREFAC R. BLOCK COLORNAL	n provinsi o nici si s Referensi den ne	na de cara en la secola como conserva en la compositiva. Antes en la grada en la compositiva en
		1988 C. C. C. C. Marker, C.
1924 of a way comparing the second se	AND TRANSPORT	
	<i>e</i> -	_
an Alta	6 Si.	
i ta z ta	2013) 2013) 2014) 2014) 2014) 2014)	e note E la giber
2 ST = 19 4 3	the states	S-11-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-1-
	(40535555555	0.4
1.1.1.1.1.1	0 I 3 8 4 9	1.1
(2-11-1) (2-11-1) すべ) 覧が -		12、12、12年2月 21、日 は 11(11)-11
and a strategy	11100	and a second sec
	الريحان فالمستر	22.03 Stat 6.7 Stat
		€
		(1) 12 年代) (1) 12 年代) 12 年代) (1) 12 年代) 12 年代)
		and at which is a
generation of the second se		
Contraction of the second seco	an communication of the second second	1 de 14 1 ^d e - Recht - Staret
на на селото селото на селото н Посто на селото на село	 A. G. S. S.	<pre>quello glamma provincia de la construction ante ante ante ante ante ante ante ante</pre>
1. I		and the second sec
, stand by	 All the second se	
$\frac{dx_{1} \cdot e}{dx_{2} \cdot e} = \frac{b_{d_{1}} \cdot e_{1} \cdot e_{2}}{dx_{2} \cdot e_{1} \cdot e_{2}} = \frac{b_{1}}{dx_{2}} \frac{b_{2} \cdot e_{2}}{dx_{2}} + \frac{b_{1}}{dx_{2}}$		
Seneral Frizer Fic	king Conditions	
		an analysis a C
2017 BEAG 1863	in a Charles Care	STR 1 TO STR
	the second second	
		A construction of the cons
. ::::::::::::::::::::::::::::::::::::	Carlo F States	
	Service Formation and Service	
	Carlo F States	control of the second sec
a da anti- ta anti-	and a second s	
a ta an	and a family of the second s	
n an	inger e Frankensen Ser in Standard Barrison Ser in Standard Barrison	
entra de la composition de la	and a family of the second s	
entra de la composition de la	inger e Frankensen Ser in Standard Barrison Ser in Standard Barrison	

σ...

Τ. τ. Οίνου δια την μισκαταστική την ποληγοριατική την ποληγοριστη την ποληγολη την ποληγοριστη την ποληγοριστη την ποληγορι

Objective Function Penalty Weights for Primers

Objective Function Penalty Weights for Primer Pairs



Hyb Oligo (Internal Oligo) Fez-Sequence Imputs

1

Nyb Oligo (Internal Oligo) General Conditions

862 × 163				
Mart Plan				
River (Str. 7				
				1200
	. 1997 a 🖭			ê wî
	, where $\tau_{\rm e}$,			
	A. A.			

Objective Function Feasity Weights for Hyb Oligos (Internal Oligos)

			<u> 1</u>		
		25:	<u>8</u> 5.2	27.2	1.27
		2.5.4	g.5-1	375 S	$\overline{c_{2n}}$
					S
					E.S.
					and a second sec
ŵ.sçā	40, -(<i>a</i> , <i>a</i> , · · ·	4.5048			

2.10.13 .2 Primer3 Output For Left Primer. Primer3 Output

WARNING: Numbers in input sequence were deleted. No mispriming library specified Using 1-based sequence positions OLIGÓ 18 50.69 38.89 5.00 2.00 cttcagaaccgatgaaat LEFT PRIMER 18 50.69 38.89 5.00 2.00 ctteagaacegatgaaat 20 51.23 35.00 5.00 3.00 cdaatgagagaaacattdta RIGHT PRIMER 191 SEQUENCE SIZE: 223 INCLUDED REGION SIZE: 223 PRODUCT SIZE: 180, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 0.00 1 atccaagcagcottcagaaccgatgaaatcaggagggdacaacccacccccaggacgaa 61 algogotatgggatgagotacatecatgagactgtatggaagggegtgeetaagttettg 121 cgccgtgtggatacagccctgaagaatatcggcatcaatgagcgcctteectacaatgtt 181 teteleatteggttetettettggatgggtggtggtggeggatg eccencere. KEYS (in order of precedence): <<<<< right primer ADDITIONAL OLIGOS sta<u>s l</u>eas<u>s</u>ta <u>se</u>to <u>se</u>to se tores 1 LEFT PRIMER 12 18 50.69 38.89 5.00 2.00 cttcagaaccgatgaaat RIGHT PRIMER 185 20 51.05 40.00 5.00 0.00 agagaaacattgtagggaag PRODUCT SIZE: 174, PAIR ANY COMPL: 4.00, PAIR 3' COMPL: 0.00 2 LEFT PRIMER 13 18 51.55 38.89 5.00 3.00 ttcagaaccgatgaaatc 51.55 38.89 5.00 3.00 ttcagaaccgatgaaatc 51.23 35.00 5.00 3.00 cgaatgagagaacattgta RIGHT PRIMER 191 20 PRODUCT SIZE: 179, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00 3 LEFT PRIMER 13 18 51.55 38.89 5.00 3.00 ttcagaaccgatgaaatc RIGHT PRIMER 185 20 51.05 40.00 5.00 0.00 agagaacattgtagggaag PRODUCT SIZE: 173, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00 4 LEFT PRIMER 12 18 50.69 38.89 5.00 2.00 etteagaacegatgaaat 51.69 40.00 5.00 0.00 algagagagagattgtaggg RIGHT PRIMER 188 20 PRODUCT SIZE: 177, PAIR ANY COMPL: 3.00, PAIR 3' COMPL: 1.00 Statistics con too in in no tm tm high high high sid many tar excl had GC too too any 3' poly end ered Ns get reg GC% clamp low high compl compl Х stab ok Left 1069 0 Ü 0 0 0 18 983 4 27 22 2 13 1206 0 Right 0 0 0 0 17 1107 0 1 0 4 77 Pair Stats: considered 303, unacceptable product size 149, high end compl 4, ok 150

primer3 release 0.9

(primer3_www_results.cgi v 0.2) note the left primer from the above PRIMER3 output.

2.10.13 .3 Primer3 For Right Primer

The exon region which are taken to design the rigth primer are 3746-3849.the parameters are similar to that of left primer.

new v () 2) - Microsoft Intern For Edd Show Castleter Hole Hole and the second s Sup Petresh Andres to the second se A m Priner3 days to inter-PETERSENS Strengthered times and room to be and room to be a set of the set o October Parker (2014) Park (2014) Parker (2014) A sequences (2014) Parker (2014) Pa Fich hybridization probe dinternal te Ras, k Elas, k Elas inter galant rasses Falat parament of summer 041 1920 1920 11.440 320.1000 100.1000 10.200 20.200 20.200 20.200 petinor Delteror OCI Paniers | Prest Form | A second state of the s Fuch Primers | Reset Form General Frimer Picking Conditions Constant Filler File File Constant States and States and File File Bar Bar States and File File File Bar Bar States and File File Bar File Bar File States and File File Bar File Bar File File States and File File Bar File Bar File Bar File File Bar File States and File File Bar Fi Contraction of the second s The second sand provide the state of the stategy
$$\label{eq:product} \begin{split} F_{\rm eff} &= \frac{1}{2} \left\{ \begin{array}{c} 1 & 1 \\ 1$$
 $\left\{ e_{1}^{(1)} e_{2}^{(1)} = e_{1}^{(1)} e_{2}^{(1)} e_{2}^{(1)} + e_{2}^{(1)} e_{2}^{(1)} e_{2}^{(1)} + e_{2}^{(1)} + e_{2}^{(1)} e_{2}^{(1)} + e_{2}^{(1)} +$
 ₩
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 1
 Other Per-Sequence Inputs

Sequence Quality

Sequence Quality

La construction de la construcción de la construcci

Objective Function Penalty Weights for Primers

Mark Part (1998)
 Mark Part (1998)

Objective Function Penalty Weights for Primer Pairs

Hyb Oligo (Internal Oligo) Fer-Sequence Inputs

Tarte in the state of the second because of the

Hyb Oligo (Internal Oligo) General Conditions

- Ale sea de Anna Ran Ran 1911 - Sea Sea Man Ran 1918 - Lig Statut, Ban Poet	apa Frank Mara Frank	
Higg (1994) Maria Cossada ang tigo Maria Katao Maria Katao Maria (1994)	Proventing Provide the state of the state o	El an
Received point to be table as end to be description de	 Bernard Market and Mar Market and Market and Market Market and Market and Market Market and Market and Marke	

Objective Function Penalty Weights for Hyb Oligos (Internal Oligos)

An order of the second state of t

2.10.13 .4 Primer3 Output For Right Primer

Primer3 Output

concernence previous conditions of the

note down the right primer from the above PRIMER3 output.

2.10.13 .5 Calculation Of The Product Size

Left primer sequence size is	223
Left primer starting point is	12
Left primer total length is	211
Introns length is	450
Right primer sequence size is	104
Right primer ending point is	101
Total right primer length is	101

Product size = Left primer total length is (211) +introns(101) +Right primer ending point is (450) = 762

2.10.14 Mitochondrial And Nuclear Enzyme

Mitochondrial and nuclear enzyme selected are Malate dehydrogenase and methyl transferase. The procedures of multiple alignments for these two enzymes are similar to that of chloroplast enzyme, which are explained above.

2	Edzynie	Accessionnumber	Coroelname	and a second
1	malatedehydrogenase	M31965		603-783,873-963, 1461-1547,1643-1727, 1984-2165,2420-2490, 2570-2660,2749-2823, 2912-3013,3100-3222, 3303-3371,3456-3524, 3619-3672,3765-4638
Γ	malatedehydrogenase	AF353203	Oryza sativa	90-1088
Γ	maiatedehydrogenase	AF007581	zea mays	78-1076
2	methyl transferase	AF063403		1854-3316,3481-4960, 5107-5271,5361-5553, 5634-5745,5819-5966, 6069-6266,6384-6642, 6723-6884,6960-7178, 7259-7417
Γ	methyl transferase	AF063403	zea mays	5107-5271,5361-5553
Γ	methyl transferase	AB028870	hordeum vulgare	83-3349
Γ	methyl transferase	AF387790	sorghum bicolor	110-1198
	methyl transferase	U76384	Triticum aestivum	52-1173
	methyl transferase	AF042332	Oryza sativa	187-1227
3	Phosphenolpyruvate carboxylase	EI7154	maize	1587-1981,2092-2176,3074- 3296,3746-3849,3975-4065,4203- 4357,4454-5452,5550-5936
	Phosphenolpyruvate carboxylase	AF399915	sorghum bicolor	21-944
	Phosphenolpyruvate carboxylase	AF271995	oryza sativa	188-3065

The summary sheet of the three enzymes is given in the table below

2.11 Results

For our study we have taken Phosphoenol pyruvate carboxylase a chloroplast enzyme. Malate dehydrogenase a mitochondrial enzyme in sorghum bicolor, maize and oryza sativa, and methyl transferase a nuclear enzyme in zea mays. Hordeum vulgare, sorghum bicolor, Triticum aestivum and oryza sativa. In case of Phosphoenol pyruvate carboxylase, Malate dehydrogenase and methyl transferase, there is a maximum alignment is between oryza sativa and maize, maize and sorghum bicolor, zea mays and Hordeum vulgare.the maximum alignment can be studied by alignment graph. The distance between the two species can be studied by phylogenetic tree. The exons of the above mentioned enzymes are used for designing primers for PCR.The summary sheet of the primers for the above enzymes are given in the table1.

Table1:

- HT			i cipa	Exems legion	Leftprimer	Left	Rightprimer.		k 10
	Phosphenolpyruvate carboxylase	EI7154	zea mays		cgtcaagctcaccttgttcc	60.83	tggaagtteetgaegteett	59.7	631
	Phosphenolpyruvate carboxylase	EI7154		3074-3296, 3746-3849	cttcagaaccgatgaaat	50.69	aaacatcagctcttcaatct	51.53	762
2	malatedehydrogenase	M31965		2912-3013, 3100-3222	gttcctgatttcttgaatgc	54.82	ccctgtggagaaccagtc	56.33	312
3	methyl transferase	AF063403		5107-5271, 5361-5553	ttgatgtgcctgtggatatga	59.94	tgcagcccttctgataaacc	60.21	417
	methyl transferase	AF063403		5634-5745, 5819-5966	tatcttttacaaaatgggcgattg	61.30	acacggaggaccaccattta	60.23	317

2.12 Discussion:

Phylogeny is about evolution and is used to reconstruct evolutionary events. It is now possible to construct phylogenetic evolution at a molecular level through analysis of molecular sequences, namely proteins & nucleic acids.

To construct phylogenetic tree among grass family, the sequences of conserved enzymes from mitochondria, chloroplast and nucleus are probed using bio-informatics tools. The scheme for such study is the following

- Identify exon regions for the enzyme to be investigated.
- An exon region of the particular enzyme is used to design the primers.
- Confirm the presence the particular sequence of the enzyme (exon) in the species of interest using wet lab techniques.
 - Isolation of chloroplast, mitochondrial and nuclear DNA
 - ✓ Amplification of DNA by using PCR
 - Hybridization techniques(southern blotting)
 - DNA sequencing by chemical and enzymatic methods.
 - Analysis of sequence based on mitochondrial and chloroplast to determine maternal inheritance.
 - ✓ Analysis based on nucleus to determine paternal inheritance.
 - ✓ Comparison of sequences using multiple alignment tools

Determine the relationship among the species under study

References

Benson, G. (1999) Tandem repeats finder program to analyze DNA sequences Nucleic acids Research 27:573-580.

GerardF.B.(2001). The use of the Monsanto draft rice genome sequence research. Plant physiology 125:1164-1165.

Katti, V.M.;Sami-Subbu,R.;Ranjekar,P.K. and gupta. V.S.(2000). Amonoacid repeat patterns in protein sequences:Their diversity and structural and functional implications. Protein science 9:1203-1209.

Lench,N.J.;Norris,A;Bailey,a.;Booth,A. and Markham,A.F.(1996). Vectoreete PCR isolation of microsatellite repeat sequences using anchored dinucleotide primers. Nucleic acids research 24:2190-2191.

Macas, J.; Meszaros, T. and Nouzova, M. (2002). Plantsat: a specialized database for plant satellite repeats. Bioinformatics 18:28-35.

Mahalakshmi, V. and Ortiz, R.(2001). Plant genomics and agriculture : from model organisms to crops, the role of data mining for gene discovery. Electronic journal of biotechnology.

Dinakar bhattramakki, jianmindong, ashok K.chhabra, and Gary E.Hart. An integrated SSR and RFLP linkage map of Sorghum bicolor(L.) Moench

Pearson, C.E., nad Sinden, R.R. (1998). Trinucleotide repeat DNA structures: Dynamic mutations from dynamic DNA. Current openion in Structural biology 8:321-330 Yuan, q.; Quackenbush, J.; Sultana, R.; Pertea, m.; Salzberg, S.L. and Buell, C.R. (20010. Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. Plant physiology, 125:1166-1174.

References

BASU,D.,K.DEHESH,H.J.SCHNEIDER-

POETSCH.S.E.HARRINGTON, S.R.McCOUCH and P.H.QUAL.2000.rice PHYC gene: structure, expression.map position and evolution. plant mol bio44:27-42.

DJE, Y., M. HEUERTZ, C. LEFEBVRE and X.VENKEMANS.2000.Assessment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatelite markers. Theoretical and applied genetics.100:918-925.

DOEBLEY, J. 1989. Isozymic evidence and the evolution of crop plants. Pp. 165-191 in D. E. soltis and P. S. soltis, eds. Isozymes in plant biology. Dioscorides press, Portland, Oregon.

DGGETT,H. 1976.Pp. 112-117.Evolution of crop plants. Longman, Essex, UK.

GAUT, B.S. and M. T. CLEGG. 1993a. Molecular evolution of the Adh1 locus in the genus Zea. proc Natl Acad Sci U S A 90: 5095-5099.

HARLAN, J. R. and A. B. L. STEMILER. 1976. The races of Sorghum in Africa. Pp. 465-478 in J. R. Harlan, J. M. de wet, and A. B. L. STEMLER, eds.origins of African plant Domestication. Mouton Press, The Hague.

HILTON, H. and B. S. GAUT. 1998. Speciation and domestication in maize and its wild relatives: evidence from the globulin-1 gene. Genetics 150:863-872.

HUDSON, R. R., M. KREITMAN and M.AGUADE.1987. Atest of neutral molecular evolution based on nucleotide data. Genetics 116: 153-159.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences Mol Evol 16: 111-20. KOLUKISAOGLU, H.U., S. MARX., C. WIEGMANN, S. HANELT and H. A. SCHNEIDER-POETSCH.1995. Divergence of the pyhtochrome gene family predates angiosperm evolution and suggests that selaginella and Equisetum arose prior to psilotum. J Mol Evol 41: 329-337.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN and M. NEI, 2001 MEGA2: Molecular Evolutionary Genetics Analysis software, Pp., Arizona state University, temp, Arizona, USA.

Nei,m.1987. molecular Evolutionary Genetics.Columbia University Press,New York.