JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
JUIT
विद्या तत्व ज्योतिसमः

Jaypee University of Information Technology
Solan (H.P.)
LEARNING RESOURCE CENTER

Acc. Num. SP03029 Call Num:

General Guidelines:

- Library books should be used with great care.
- Tearing, folding, cutting of library books or making any marks on them is not permitted and shall lead to disciplinary action.
- Any defect noticed at the time of borrowing books must be brought to the library staff immediately. Otherwise the borrower may be required to replace the book by a new copy.
- The loss of LRC book(s) must be immediately brought to the notice of the Librarian in writing.

Learning Resource Centre-JUIT
SP03029

DEVELOPMENT OF A PATTERN MATCHING TOOL FOR DETECTING INORGANIC CO TRANSPORTER FAMILIES IN PLANTS

By

PRERNA MISRA-031526
KUHU GUPTA-031528

Learning Resource Centre JUIT Waknaghat, Solan (H.P.)

MAY-2007

Submitted in partial fulfillment of the Degree of Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY
AND BIOINFORMATICS
JAYPEE UNIVERSITY OF INFORMATION
TECHNOLOGY-WAKNAGHAT

CERTIFICATE

This is to certify that the work entitled, "Development of a Pattern Matching Tool for Detecting Inorganic Co Transporter Families in Plants" submitted by Prerna Misra and Kuhu Gupta in partial fulfillment for the award of degree of Bachelor of Technology in Bioinformatics of Jaypee University of Information Technology has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr.R.S Chauhan
(H.O.D)
Department of Biotechnology and Bioinformatics

ACKNOWLEDGEMENT

It has been a great pleasure working under the able guidance of faculty & staff in the Department of Biotechnology and Bioinformatics at Jaypee University of Information Technology, Waknaghat, Solan during our study as a B.Tech student. Thanks a lot to our project advisor and committee chair for his able guidance and motivation in completing this project successfully without which this research would not be possible.

Many people have contributed to this project in a variety of ways over the past few months. To the individuals who have helped us, we again express our appreciation. We also acknowledge the many helpful comments received from our teachers of the concerned department. We are indebted to all those who provided reviews & suggestions for improving the results and the topics covered in this project and extend our apologies to anyone we may have failed to mention.

Prerna Misra
PRERNA MISRA

KUHU GUPTA

TABLE OF CONTENTS

ABSTRACT

The present study was carried out to develop an in silico pattern matching tool to identify whether a particular protein belongs to a transporter family. For this purpose the protein sequences belonging to different families were downloaded from the NCBI. Then a consensus sequence was found using an online tool called PRATT from Prosite scan. The consensus sequence was obtained by giving the protein sequences of a single family as input in FASTA format. This consensus sequence was present in all the protein sequences of a single family. Consensus was used as a determining factor to classify the query protein sequence into its respective family using pattern matching.

Pattern matching refers to searching of a sequence of characters within a character string. A tool was developed using C# and Visual Basic.NET to match the pattern of the consensus sequence with the query protein sequence which is entered by the user in the box provided on the front end page. When doing pattern matching if a pattern is found then a match is said to have occurred. When the match occurs, the matched pattern is looked for its respective family. To whichever pattern of the consensus of the family it matches, classification is done in that family of the query sequence. The name of the family is displayed on the front end page. Identification of consensus is crucial in classifying a protein sequence into different micronutrient transporter families for an unknown protein sequence. It is expected that the tool would prove useful to biologists of different streams.

Keywords

Bioinformatics, Visual Basic, Fasta, NCBI, C#

CHAPTER 1

INTRODUCTION

Comparative genomics is the study of relationships between the genomes of different species or strains. Comparative genomics is an attempt to take advantage of the information provided by the signatures of selection to understand the function and evolutionary processes that act on genomes

Transport of metals and alkali cations across plant plasma and organellar membranes is essential for plant growth, development, signal transduction, nutrition, and also for use of plants in toxic metal phytoremediation. Alkali cation and metal transporters have been analyzed traditionally in great depth as models for understanding plant membrane transport.

Plant transporters also play important roles in shuttling potentially toxic cations across plant membranes. The cation selectivity filters of plant transporters often allow toxic cations to be transported, along with cationic nutrients. Powerful genetic approaches have been developed that allow high-throughput selection of point mutations that reduce or block transport of toxic cations, while maintaining nutrient transport.

Plants use these macronutrients and micronutrients in:

1. Structural components in carbohydrates and proteins
2. Organic molecules used in metabolism, such as the Magnesium in chlorophyll and the Phosphorous found in ATP
3. Enzyme activators like potassium, which activates possibly fifty enzymes
4. Maintaining osmotic balance

Our project mainly emphasizes on the comparative genomic analysis of seven co-transporter families which are as follows-:

- ZIP Family
- Ammonium transporter Family.
- Sodium transporter Family.
- Inorganic phosphate transporter Family.
- Nitrate transporter Family.
- Putative sulfate transporter Family
- NRAMP Family

As in all organisms, metal cations are crucial for nutrition in plants. Several metals, such as copper, iron, zinc, and manganese, act as important cofactors for many enzymes and are essential for both mitochondrial and chloroplast functions. However, when supplied in excess, these essential cations can become toxic, like heavy metals with no generally established function, such as cadmium, lead, or mercury. To maintain micronutrient metal homeostasis and to cope with the deleterious effects of nonessential heavy metals, plants have developed a complex network of metal uptake, chelation, trafficking, and storage processes. Metal transporters are required to maintain metal homeostasis and thus constitute important components of this network

INORGANIC COTRANSPORTER FAMILIES

AMMONIUM TRANSPORTER FAMILY-:

The ammonium transporters are designated as Amt proteins. The proteins of the Amt family vary in size from 391 to 622 amino acyl residues and possess 11 (most members) or 12 (the E. coli AmtB protein) transmembrane a-helical spanners. They occur in Gram-negative and Gram-positive bacteria, archaea, yeast, plants and animals. The eukaryotic proteins are, in general, larger than the prokaryotic proteins. All functionally characterized members of the family are ammonia or ammonium uptake transporters. Some, but not others, also transport methyl ammonium. Proteins of the Amt family are probably ubiquitous. Homologues have been found in archaea and animals, and many organisms from all major kingdoms of living organisms possess multiple homologues. One of these proteins, Mep2 of Saccharomyces cerevisiae, has been shown to function both as a transporter and as a sensor, generating a signal that regulates filamentous growth (pseudohyphal differentiation) in response to ammonium starvation. This protein has an N-terminal, asparaginyl-linked glycosylated domain where only Asn-4 is glycosylated.

The transport of ammonium across cell membranes is a process of fundamental importance in bacteria and plants as in both cases ammonium is a primary source of inorganic nitrogen. However the means by which ammonium enters cells was the subject of debate for many decades until genes encoding high-affinity ammonium transporters (Amt) were isolated in 1994. The Amt family of proteins is unique and is ubiquitous, being found in eubacteria, archaebacteria, fungi, plants, nematode worms and insects.

INORGANIC PHOSPHATE TRANSPORTER FAMILY

Transport of inorganic phosphate (Pi) through plant membranes is mediated by a number of families of transporter proteins. Studies on the topology, function, regulation and sites of expression of the genes that encode the members of these transporter families are enabling roles to be ascribed to each of them. The Pht1 family, of which there are nine members in the Arabidopsis genome, includes proteins involved in the uptake of Pi from the soil solution and the redistribution of P i within the plant. The efficient acquisition of Pi by cluster roots arises primarily from their capacity to increase the availability of soil Pi immediately adjacent to the rootlets by excretion of carboxylates, protons and phosphates within the cluster.

ZIP TRANSPORTER FAMILY

Members of the ZIP family consist of 220-430 amino acyl residues with eight putative transmembrane spanners. They are derived from animals, plants and yeast, bacteria and archaea. They comprise a diverse family, with several paralogues in any one organism (e.g., twelve in humans, at least five in Caenorhabditis elegans and Arabidopsis thaliana, and two in Saccharomyces cervisiae). The two S. cerevisiae proteins, Zrt1 and Zrt2, both probably transport Zn2+ with high specificity, but Zrt1 transports Zn2+ with ten-fold higher affinity than Zrt2. Some members of the ZIP family have been shown to transport Zn2+ while others transport Fe2+, and at least one transports a range of metal ions. One human protein member of the ZIP family is designated "growth arrest inducible gene product," but its presumed transport activity has not been identified. The energy source for transport has not been characterized, but these systems probably function as secondary carriers.

Despite the importance of Zn as an essential micronutrient for plant growth, relatively few studies have examined the mechanisms and regulation of Zn absorption by roots. Zn is taken up from the soil solution as a divalent cation .Currently, there is little agreement as to whether uptake is via ion channels or via a divalent cation carrier protein and whether there is a link between uptake and metabolic energy transduction. Studies of Zn uptake in plants mainly have been focused on hyper accumulators, i.e., plants that can grow in soils containing high levels of Zn and accumulate high concentrations of Zn in their shoots. Zn is believed to be transported into the xylem, taken up by leaf cells, and then stored in the vacuoles of leaf cells thus preventing the buildup of toxic levels in the cytoplasm. These observations indicate that there are also Zn transporters that transport the metal between cells and into sub cellular compartments within the plant.

SODIUM TRANSPORTER FAMILY

Sodium transporter proteins, which plays a central role in plant tolerance to salt. Upon prolongated exposure to high concentrations, Na (+) translocates from the roots to the transpiring leaves where it can increase to toxic level. Involved in Na(+) recirculation from shoots to roots, probably by mediating Na(+) loading into the phloem sap in shoots and unloading in roots, thereby removing large amounts of Na(+) from the shoot. Does not transport K(+) but regulates K(+) nutrient status via its ability to facilitate Na(+) homeostasis. Probably not involved in root uptake of Na (+).It is subcellularly in Cell membrane; multi-pass membrane protein. Highly expressed in roots. Expressed in flowers, leaves and stems. Expressed in the vascular tissues of every organs. In roots, leaves and flower peduncles, it is only expressed in the phloem tissues. Not expressed in root peripheral cells.

PUTATIVE SULFATE TRANSPORTER FAMILY

Putative sulfate transporter family is a large and ubiquitous family with over 30 sequenced members derived from bacteria, fungi, plants and animals. Many organisms including Bacillus subtilis, Synechocystis sp, Saccharomyces cerevisiae, Arabidopsis thaliana and Caenorhabditis elegans possess multiple SulP family paralogues. Many of these proteins are functionally characterized, and all are sulfate uptake transporters. Some transport their substrate with high affinities, while others transport it with relatively low affinities. Sulfur is available to plants primarily in the form of anionic sulfate (SO_4^{2-}) present in soil. It is actively transported into roots and then distributed, mostly unmetabolized, throughout the plant. SO_4^{2-} is a major anionic component of vacuolar sap; therefore, it does not necessarily enter the assimilation stream. Gaseous sulfur dioxide (SO_2) is readily absorbed and assimilated by leaves, but it is significant as a nutrient source only in industrial areas with air pollution. Sulfur is assimilated in one of two oxidation states. SO_4^{2-} .Proton/sulphate co-transport in the plasma membrane of root cells is the first step for the uptake of sulphate from the environment by plants. Further intracellular, cell-to-cell and long-distance transport must fulfill the requirements for sulphate assimilation and source/sink demands within the plant.

A gene family of sulphate transporters, which may be subdivided into five groups, has been identified with examples from many different plant species. For at least two groups, proton/sulphate co-transport activity has been confirmed. It appears that each group represents sulphate transporters with distinct kinetic properties, patterns of expression, and cell/tissue specificity related to specific roles in the uptake and allocation of sulphate. High-affinity sulphate uptake and low-affinity vascular transport, as well as vacuolar efflux, are controlled by the nutritional status of the plant. Most notably there is an apparent increase in capacity for cellular sulphate uptake and vacuolar efflux when sulphur supply is limiting. Within the groups, the individual sulphate transporters may be further subdivided by differences in temporal, cellular and tissue expression. Many of the transporters are regulated by the nutritional status of the individual tissues, to optimize sulphate movement within and between sink and source organs.

Sulphur is an essential nutrient required for plant growth. Sulphur is mainly taken up by the plant as inorganic sulphate from the soil, and the assimilation into cysteine is considered to be the key entry point of the natural sulphur cycle. The acquisition of sulphur by plants has become an increasingly important concern for the agricultural industry due to the decreasing trends of S-emissions from industrial sources and the consequent limitation of inputs from deposition (McGrath et al., 1996). The recognition of the importance of sulphate for plant growth and vigour and hence crop yield, as well as the nutritional importance of sulphur for human and animal diets, has led to an increased emphasis on research on the processes of sulphate uptake, transport and assimilation.

NITRATE TRANSPORTER FAMILY

Nitrate is an essential element for plant growth, both as a primary nutrient in the nitrogen assimilation pathway and as an important signal for plant development. The uptake of nitrate from the soil and its translocation throughout the plant has been the subject of intensive physiological and molecular studies. Using a reverse genetic approach, the AtNRT2.1 gene has been shown to be involved in the inducible component of the high-affinity nitrate transport system in Arabidopsis. The Arabidopsis Genome Initiative has released nearly the whole genome sequence of Arabidopsis, allowing the identification of a small NRT2 multigene family in this species. Thus, we investigated the phylogenetic relationship between NRT2 proteins belonging to several kingdoms and compared the structure of the different members of the Arabidopsis family. We analyzed, by semi quantitative reverse transcriptase-polymerase chain reaction, the expression pattern of each gene depending on plant organ and development or nutritional status, and compared the relative level of each gene by real-time polymerase chain reaction. We also evaluated the significance of each paralog on the basis of the relative levels of gene expression.

Nitrate uptake by root cells is the first step of the nitrate assimilation pathway in higher plants. To cope with large variations in nitrate concentrations in cultivated soils, plants have developed two different uptake systems (for review, see Glass and Siddiqi, 1995; Forde and Clarkson, 1999). The low-affinity nitrate transport system is used preferentially when external nitrate concentration is high (above 1 mm; Siddiqi et al., 1990), whereas the high-affinity transport system (HATS) takes place at very low external concentrations (between 1 µm and 1 mm; Behl et al., 1988). After its entry in the cytoplasm of root epidermal cells, nitrate is either translocated and stored in the vacuole or excreted back into the apoplasm. It also can be reduced into nitrite in the cytoplasm by nitrate reductase. The nitrite is then translocated to the chloroplast where it is further reduced into ammonium by nitrite reductase (Faure et al., 2001). Finally, nitrate can be excreted from the root cytoplasm in the xylem vessels and unloaded in aerial organs, where it can follow the same fates as in roots. Although it is known that nitrate transfer across the root occurs by diffusion along a radial concentration gradient in the symplasm, many of these nitrate or nitrite fluxes entail the passage across the plasmalemma or the tonoplast. The passage through one or the other of these membranes requires active processes that involve transporters anchored in these structures.

NRAMP FAMILY

Genes encoding members of the NRAMP family of integral membrane proteins have been identified in bacteria, fungi, plants and animals. The proteins encoded by AtNRAMP genes cluster in two subfamilies: one including AtNRAMP1 and 6 and the other including AtNRAMP2 to 5. In addition, the ethylene insensitivity gene EIN2 which functions in transduction of multiple stress signals contains a NRAMP homologous domain but its homology with other members of the NRAMP family is much lower (Alonso et al., 1999). The functions of AtNRAMP proteins in metal transport have been demonstrated both in the heterologous yeast expression system and in planta (Alonso et al., 1999; Curie et al., 2000; Thomine et al., 2000).

The yeast proteins exhibit greater than 50% similarity to so-called "natural resistance-associated" macrophage proteins (Nramp) found in mammals, birds, nematodes and insects. Homologues are also found in other yeast, plants, archaea, and Gram-negative and Gram-positive bacteria. The animal proteins play a role in resistance to intracellular bacterial pathogens such as Salmonella typhimurium, Leishmania clonovani and Mycobacterium bovis. It is hypothesized that a deficiency for Mn2+ or some other metal prevents the generation of reactive oxygenic and nitrogenic compounds that are used by macrophage to combat pathogens. Nramp family members are found in many animal tissues besides macrophages where they are expressed at low levels. One mammalian mutant species, Nramp2 of rat, has been shown to exhibit defective endosomal iron export within the ferritin cycle and plays roles in intestinal iron absorption. It has been reported to transport a number of different transition metals with similar affinities It is found in apical membranes of intestinal epithelial cells, but also in late endosomes and lysosomes.

As their animal and yeast counterparts, plant NRAMP genes encode transition metal transporters with a broad selectivity. In the future, it will be important to determine whether they actually transport several transition metals in vivo or if physiological conditions or cofactors present in plant cells determine a narrower selectivity.

plant NRAMP encode intracellular metal transporters with putative subcellular localization as diverse as the plastid envelope or the vacuolar membrane. In the future, it will be important to determine systematically the subcellular localization of all plant NRAMP proteins. With such localizations, plant NRAMP proteins are expected to play important functions in intracellular metal homeostasis.

NEED FOR PATTERN MATCHING TOOL

Pattern matching is needed to check for the presence of the constituents of a given pattern. In contrast to pattern recognition, the pattern is rigidly specified. The pattern concerns conventionally either sequences or tree structures. Pattern matching is used to test whether things have a desired structure, to find relevant structure, to retrieve the aligning parts, and to substitute the matching part with something else. Sequence (or specifically text string) patterns are often described using regular expressions and matched using respective algorithms. Sequences can also be seen as trees branching for each element into the respective element and the rest of the sequence, or as trees that immediately branch into all elements.

In our project we have used the consensus sequences to classify the query protein sequence entered by the user into respective inorganic cotransporter families
by generating a "PATTERN MATCHING TOOL" .

CHAPTER 2

MATERIALS AND METHODS

Materials used:

Hardware:

Pentium IV

256 MB RAM

Software:

Visual Studio 2005

Framework 2

WINDOWS XP SP2 or above.

Language:

C#

Methodology-:

Online tools used
PRATT (Prosite Scan)

PROSITE SCAN-:

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them. PROSITE is complemented by Pro Rule , a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids.

PRATT

Interactively generates conserved patterns from a series of unaligned proteins

An important problem in sequence analysis is to find a pattern matching sets or subsets of sequences. This tool allows the user to search for patterns conserved in sets of unaligned protein sequences. The user can specify what kind of patterns should be searched for, and how many sequences should match a pattern to be reported.

The ScanProsite tool allows to scan protein sequence(s) (either from UniProt Knowledgebase (Swiss-Prot/TrEMBL) or PDB or provided by the user) for the occurrence of patterns, profiles and rules (motifs) stored in the PROSITE database, or to search protein database(s) for hits by specific motif(s) The program PRATT can be used to generate your own patterns.

PRATT can be used for finding the consensus sequences from the input protein sequences that are given in FASTA format for the Transporter families.

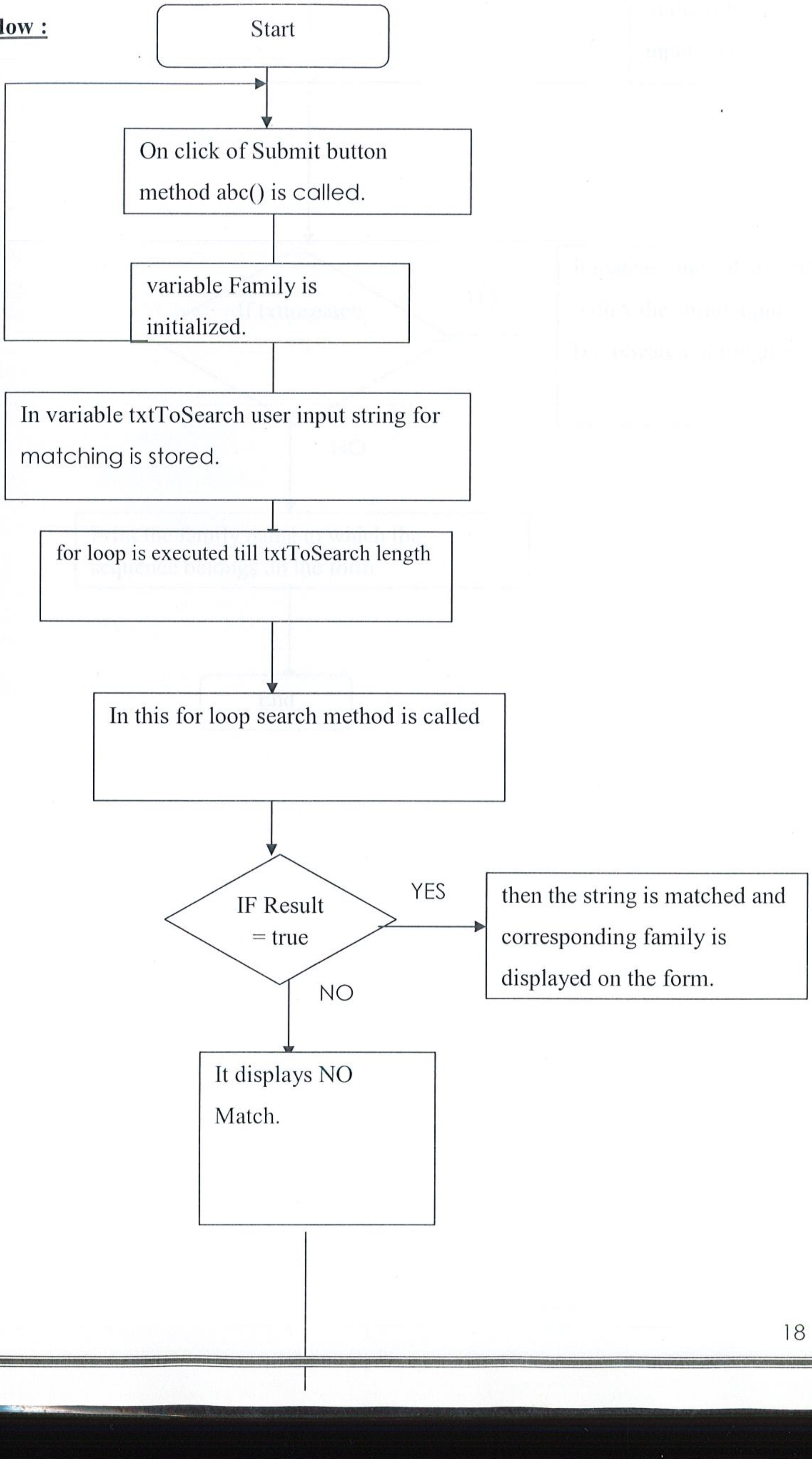
A consensus sequence is a way of representing the results of a multiple sequence alignment, where related sequences are compared to each other, and similar functional sequence motifs are found. The consensus sequence shows which residues are conserved (are always the same), and which residues are variable. We used consensus sequences as a basic approach for generating this pattern matching tool.

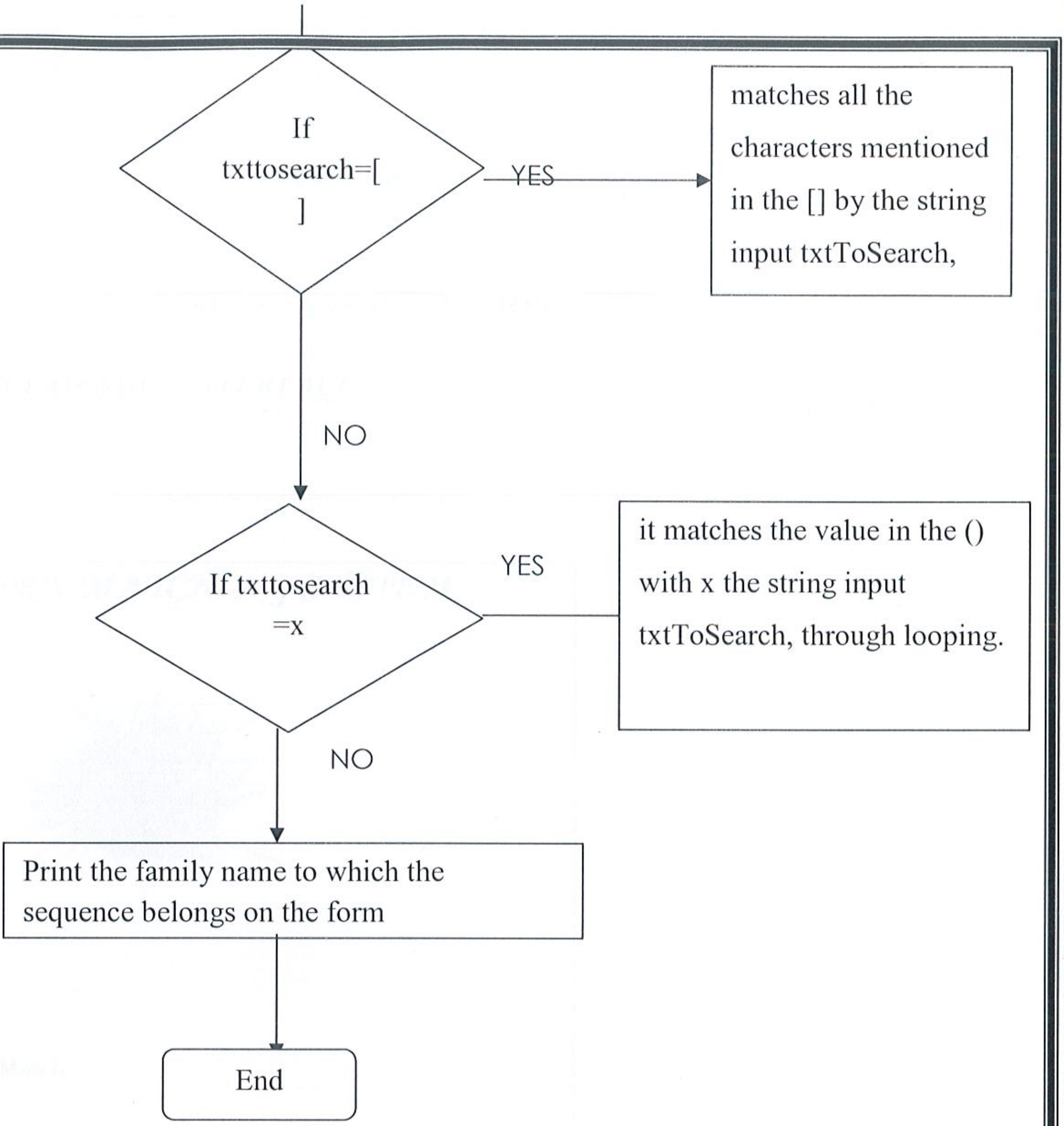
PROTOCOL FOR GENERATING PATTERN MATCHING TOOL

- We generated a pattern matching tool which helps the user to classify the query protein sequence into its respective families.
- The sequences were downloaded from NCBI which belongs to particular family in FASTA format.
- We then used prosite (pratt) to find the consensus sequence of each family.
- These consensuses were further used to generate a pattern matching tool which classifies the user query sequence into respective families.
- Further these consensus sequences are used to design a pattern matching tool which is completely user friendly.
- The source code is written in C# and the front end is made using dot net.

Our tool provides the user with the easiest way of classifying any of the given protein sequence into the respective families. The "PATTERN MATCHING TOOL" helps in classifying the protein sequences.

Main Window :

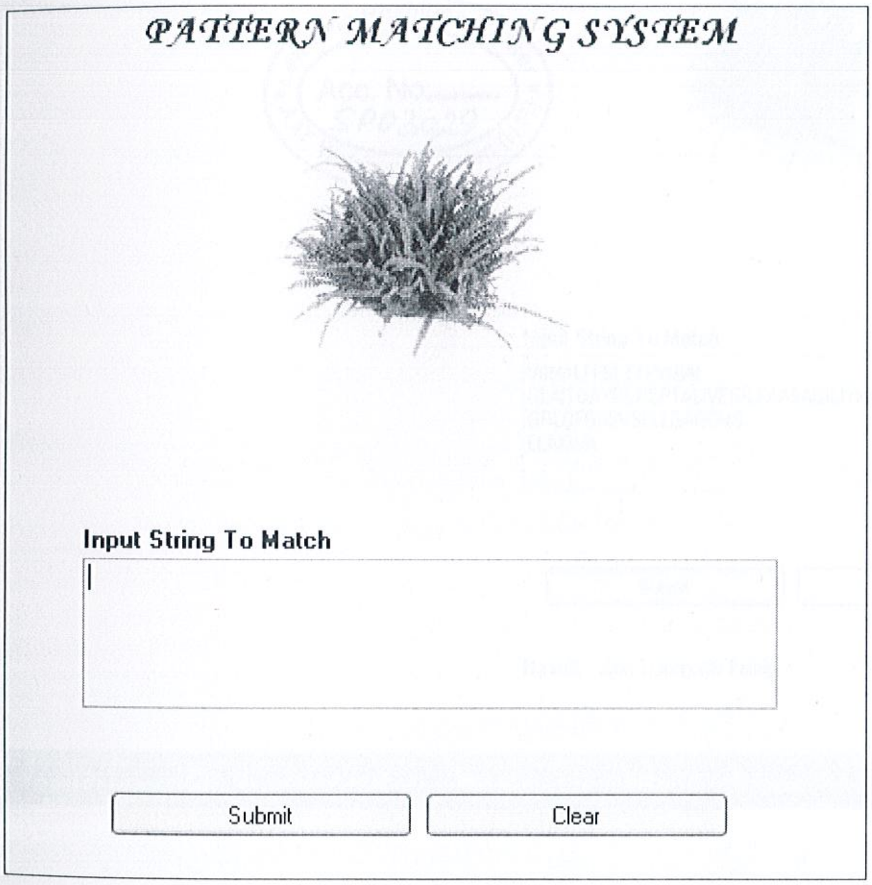

Start
On click of Submit button method abc() is called.
variable Family is initialized.
In variable txtToSearch user input string for matching is stored.
for loop is executed till txtToSearch length
In this for loop search method is called
IF Result = true
YES
then the string is matched and corresponding family is displayed on the form.
NO
It displays NO Match.


If txttosearch=[ ]
YES
matches all the characters mentioned in the [] by the string input txtToSearch,
NO
If txttosearch =x
YES
it matches the value in the () with x the string input txtToSearch, through looping.
NO
Print the family name to which the sequence belongs on the form
End

CHAPTER 3

RESULTS AND DISCUSSION

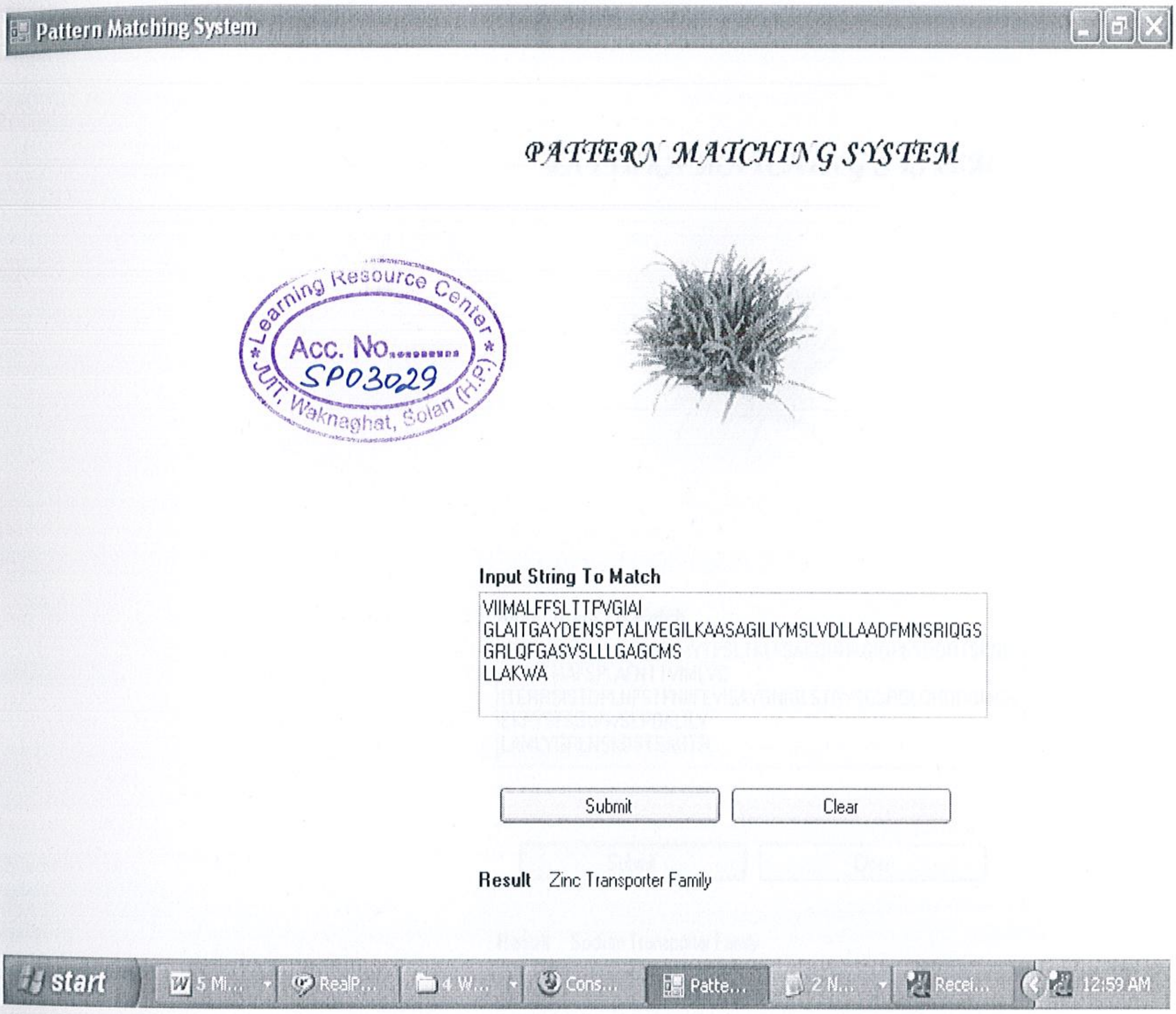
USER FRIENDLY INTERFACE



This is the user friendly interface which has been generated to get the input by the user.

The box which has been provided at the front end asks for the query sequence which is to be classified into one of the inorganic cotransporter families.The output will be displayed on the same page at the left hand side corner i.e the name of the family to which the sequence to which the query sequence belongs.

>gi|110832251|gb|ABH01187.1| zinc transporter protein [Ammopiptanthus mongolicus]



PATTERN MATCHING SYSTEM

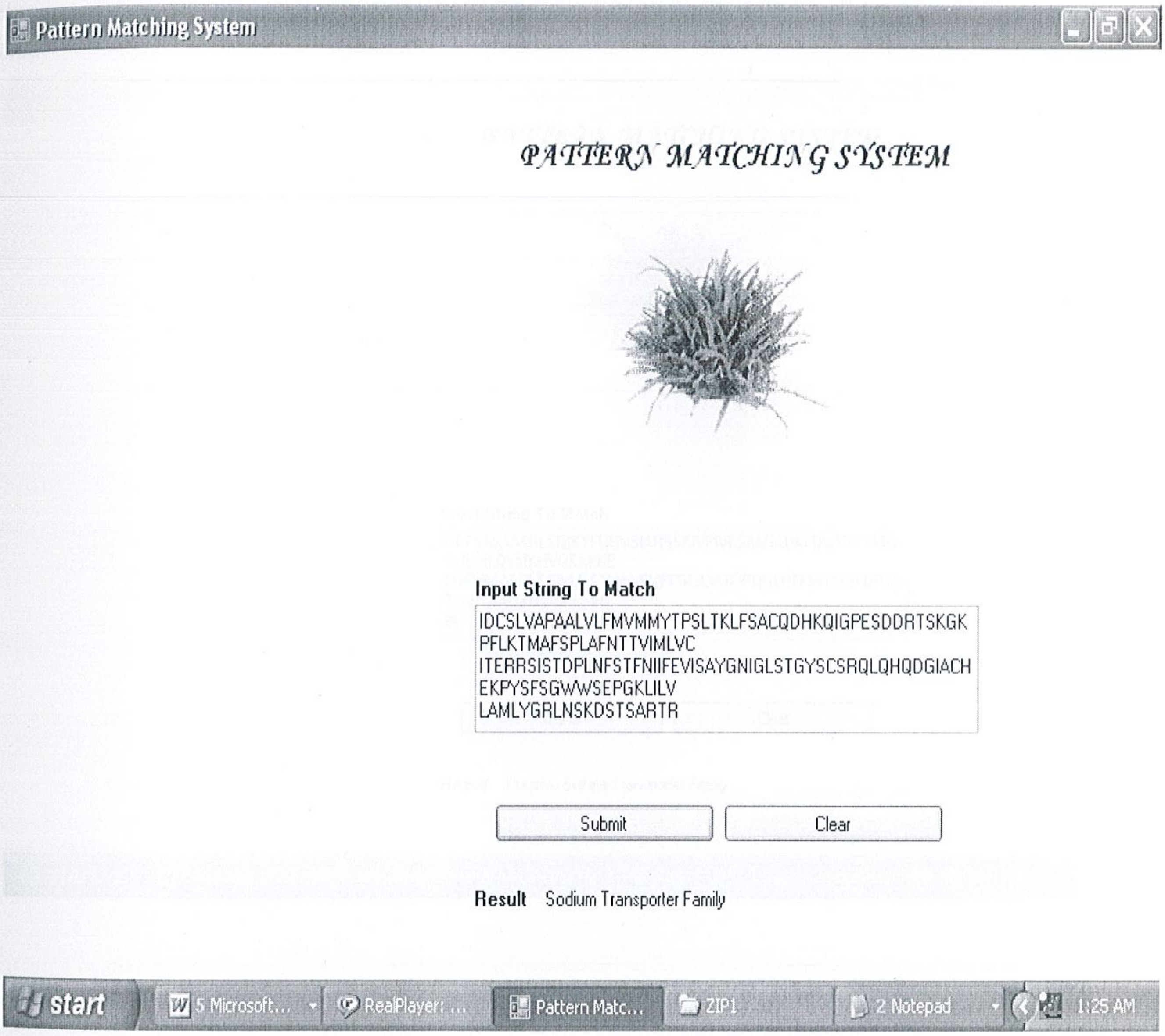
Input String To Match

```
VIIMALFFSLTTPVGIAI
GLAITGAYDENSPTALIVEGILKAASAGILIYMSLVDLLAADFMNSRIQGS
GRLQFGASVSLLLGAGCMS
LLAKWA
```

Result Zinc Transporter Family

Learning Resource Center
JUIT, Waknaghat, Solan (H.P.)
Acc. No. SP03029

>gi|75153723|sp|Q8L4K5|HKT9_ORYSA Probable cation transporter HKT9 (OsHKT9)

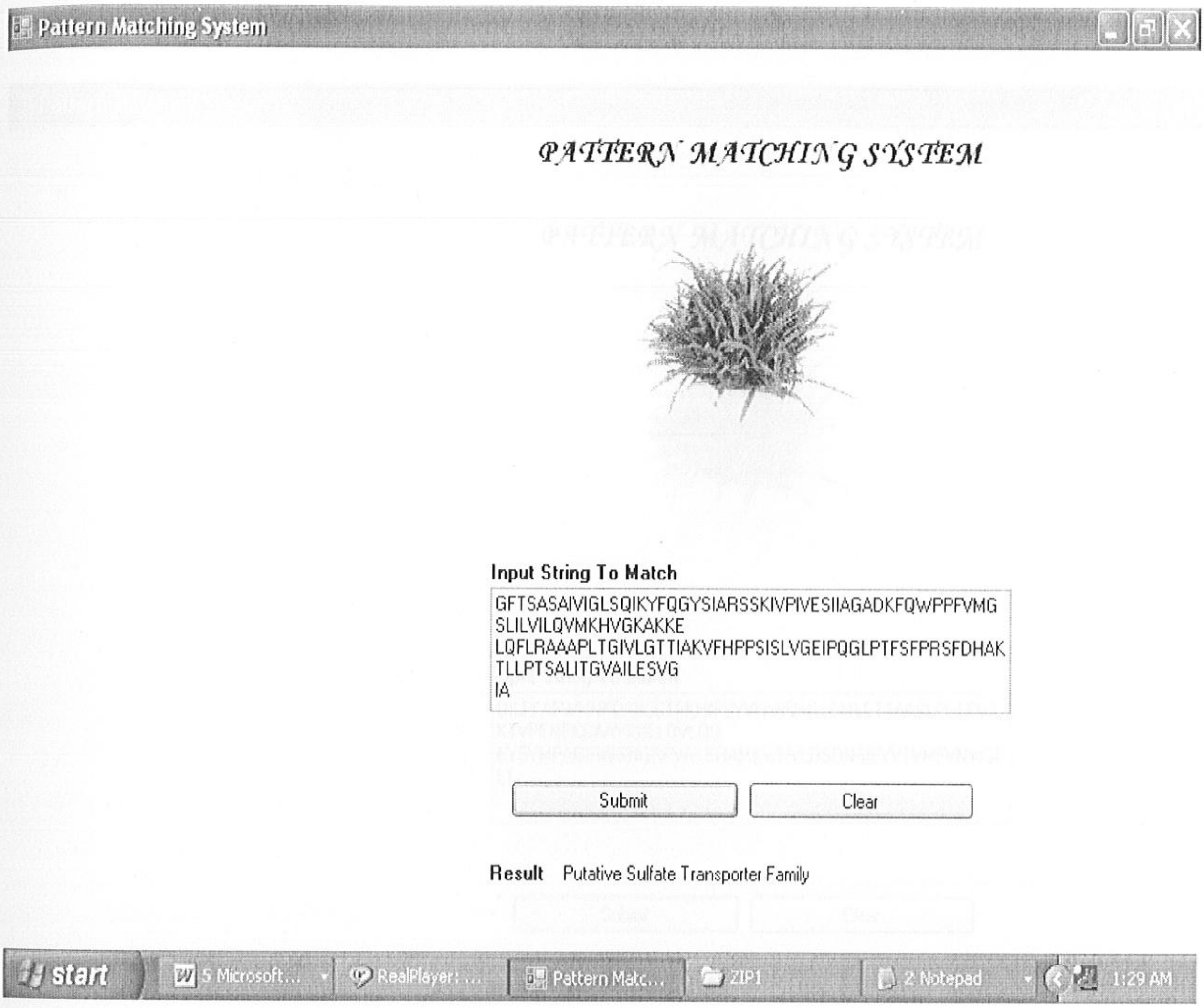


Input String To Match

```
IDCSLVAPAALVLFMVMMYTPSLTKLFSACQDHKQIGPESDDRTSKGK
PFLKTMAFSPLAFNTTVIMLVC
ITERRSISTDPLNFSTFNIIFEVISAYGNIGLSTGYSCSRQLQHQDGIACH
EKPYSFSGWWSEPGKLILV
LAMLYGRLNSKDSTSARTR
```

Result Sodium Transporter Family

>gi|26453094|dbj|BAC43623.1| putative sulfate transporter [Arabidopsis thaliana]

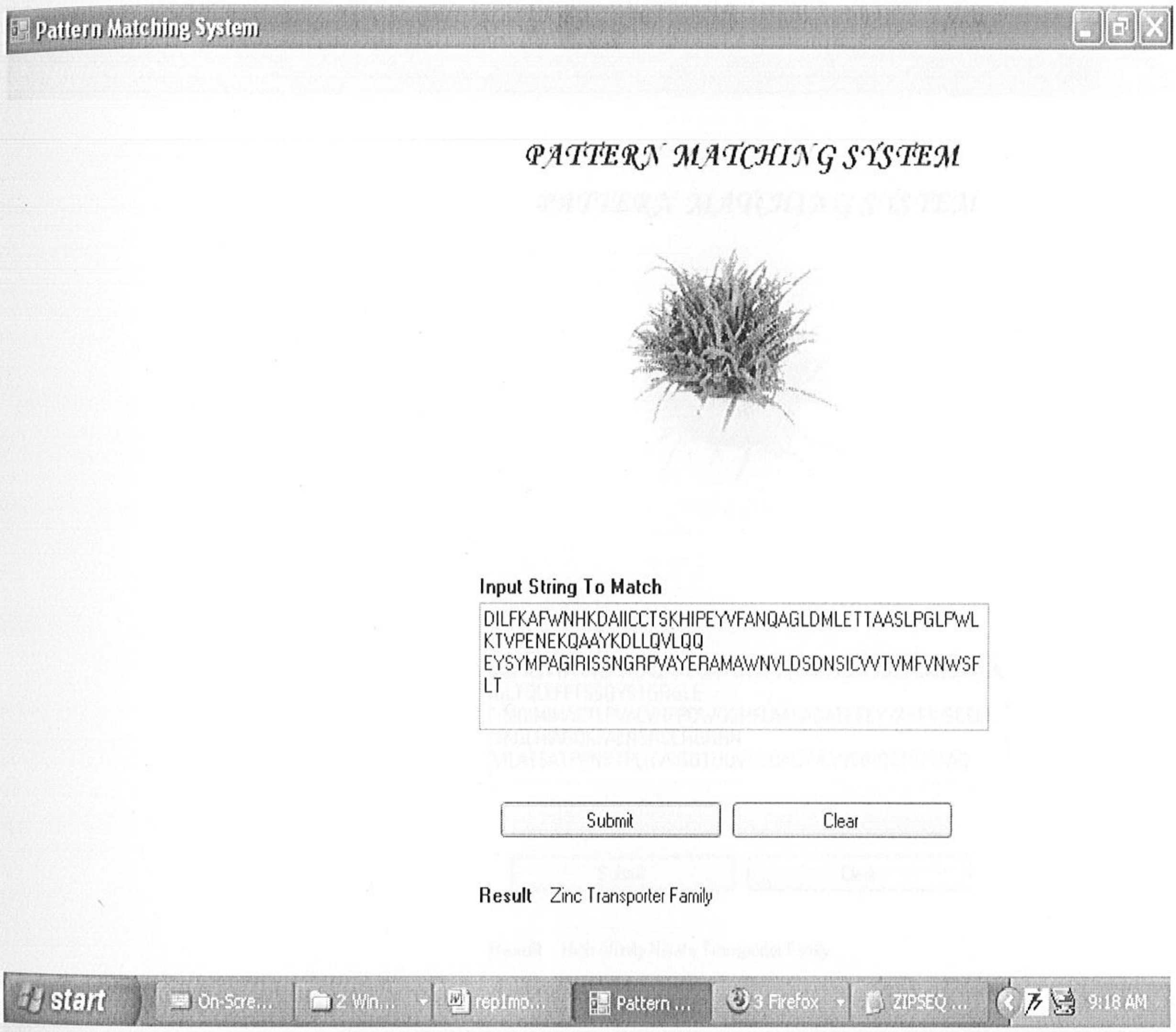


Input String To Match

```
GFTSASAIVIGLSQIKYFQGYSIARSSKIVPIVESIIAGADKFQWPPFVMG
SLILVILQVMKHVGKAKKE
LQFLRAAAPLTGIVLGTTIAKVFHPPSISLVGEIPQGLPTFSFPRSFDHAK
TLLPTSALITGVAILESVG
IA
```

Result Putative Sulfate Transporter Family

>gi|110349548|gb|ABG73249.1| class III HD-Zip protein HDZ35 [Pinus taeda]

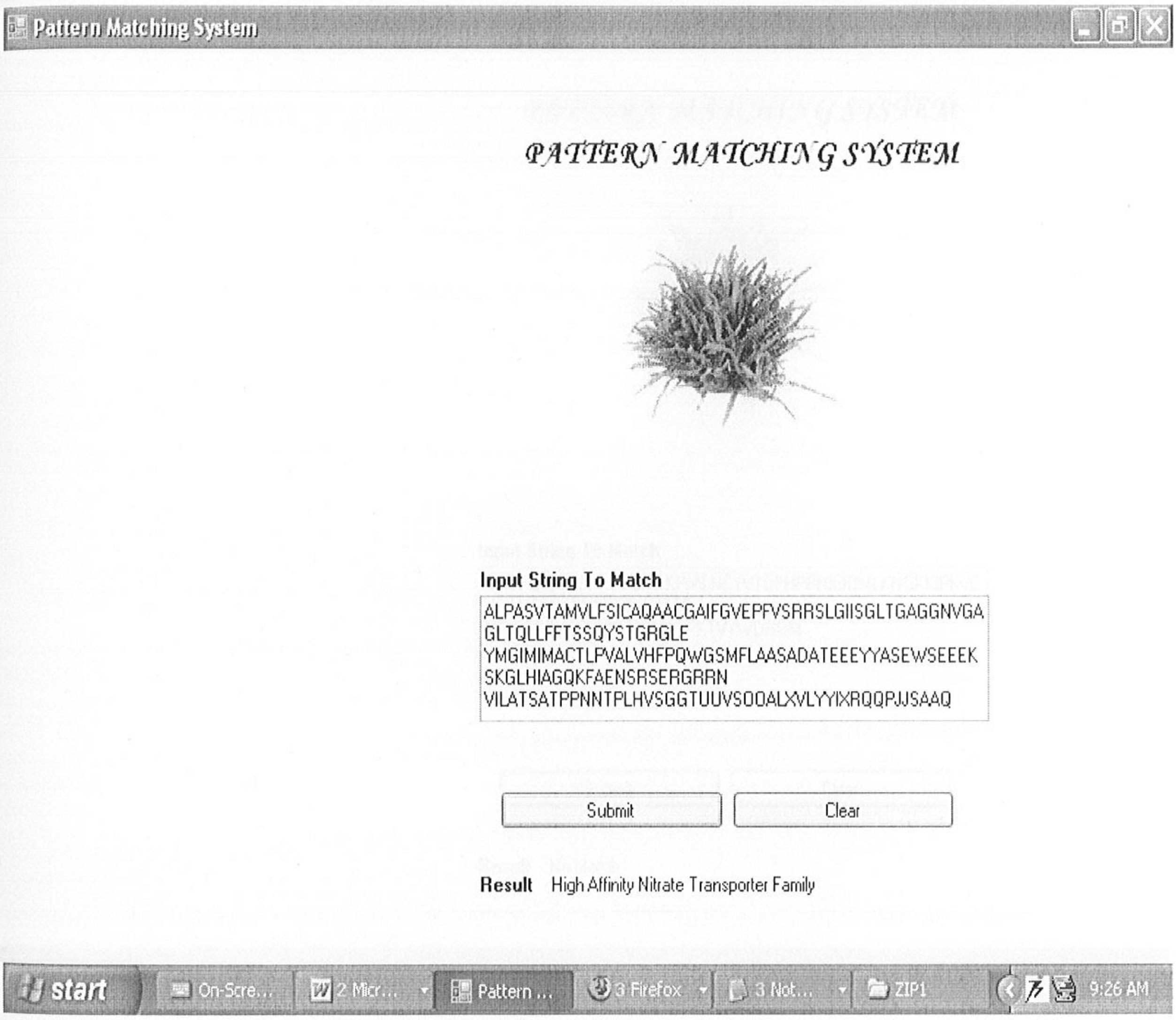


Input String To Match

```
DILFKAFWNHKDAIICCTSKHIPEYVFANQAGLDMLETTAASLPGLPWL
KTVPENEKQAAYKDLLQVLQQ
EYSYMPAGIRISSNGRPVAYERAMAWNVLDSDNSICWTVMFVNWSF
LT
```

Result Zinc Transporter Family

>gi|4731148|gb|AAD28364.1|AF091116_1 putative high affinity nitrate transporter [Hordeum vulgare]



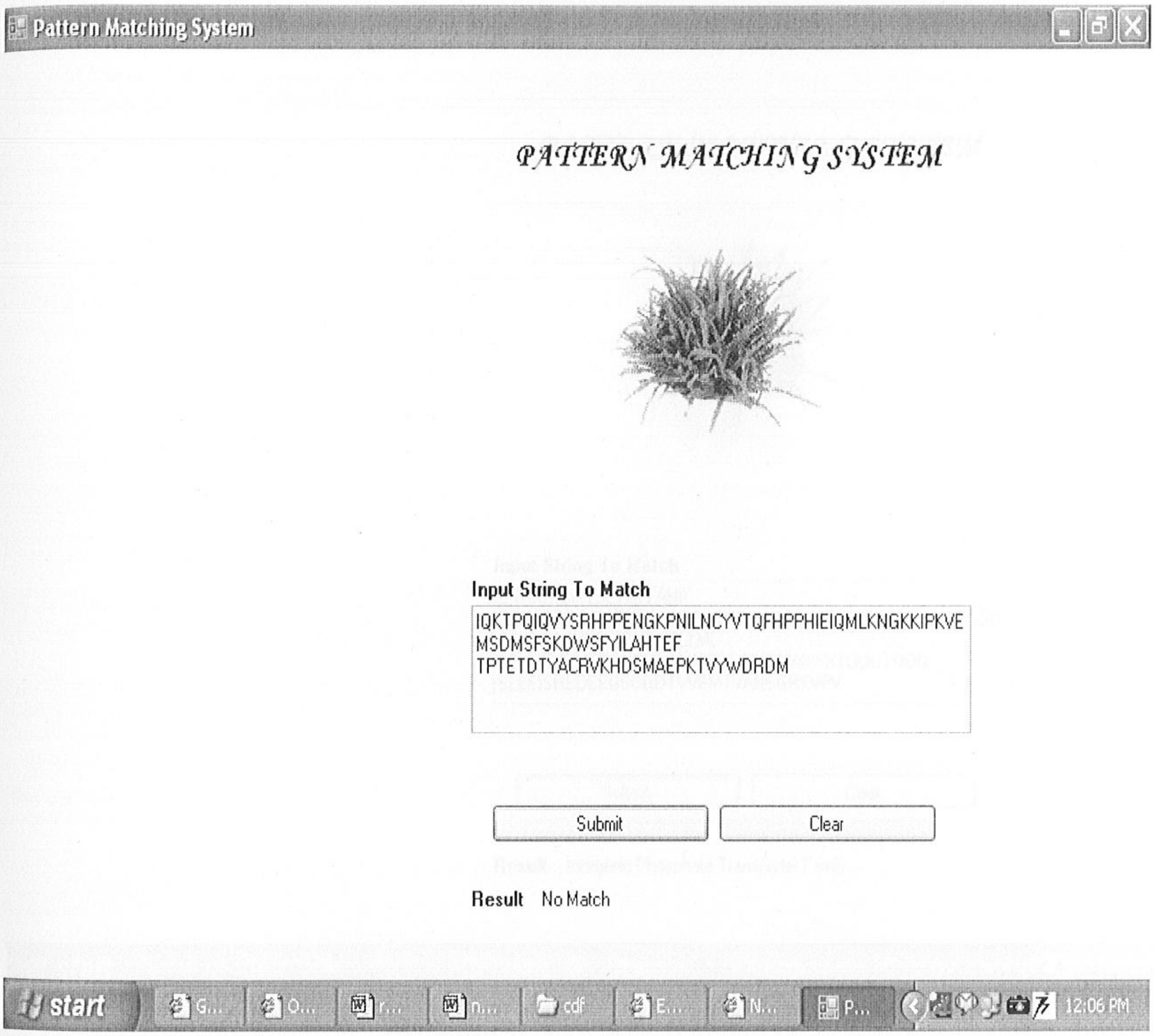
PATTERN MATCHING SYSTEM

Input String To Match

```
ALPASVTAMVLFSICAQAACGAIFGVEPFVSRRSLGIISGLTGAGGNVGA
GLTQLLFFTSSQYSTGRGLE
YMGIMIMACTLPVALVHFPQWGSMFLAASADATEEEYYASEWSEEEK
SKGLHIAGQKFAENSRSERGRRN
VILATSATPPNNTPLHVSGGTUUVSOOALXVLYYIXRQQPJJSAAQ
```

Result High Affinity Nitrate Transporter Family

>gi|28373740|pdb|1ML5|XX Chain x, Structure of the E. coli Ribosomal Termination Complex

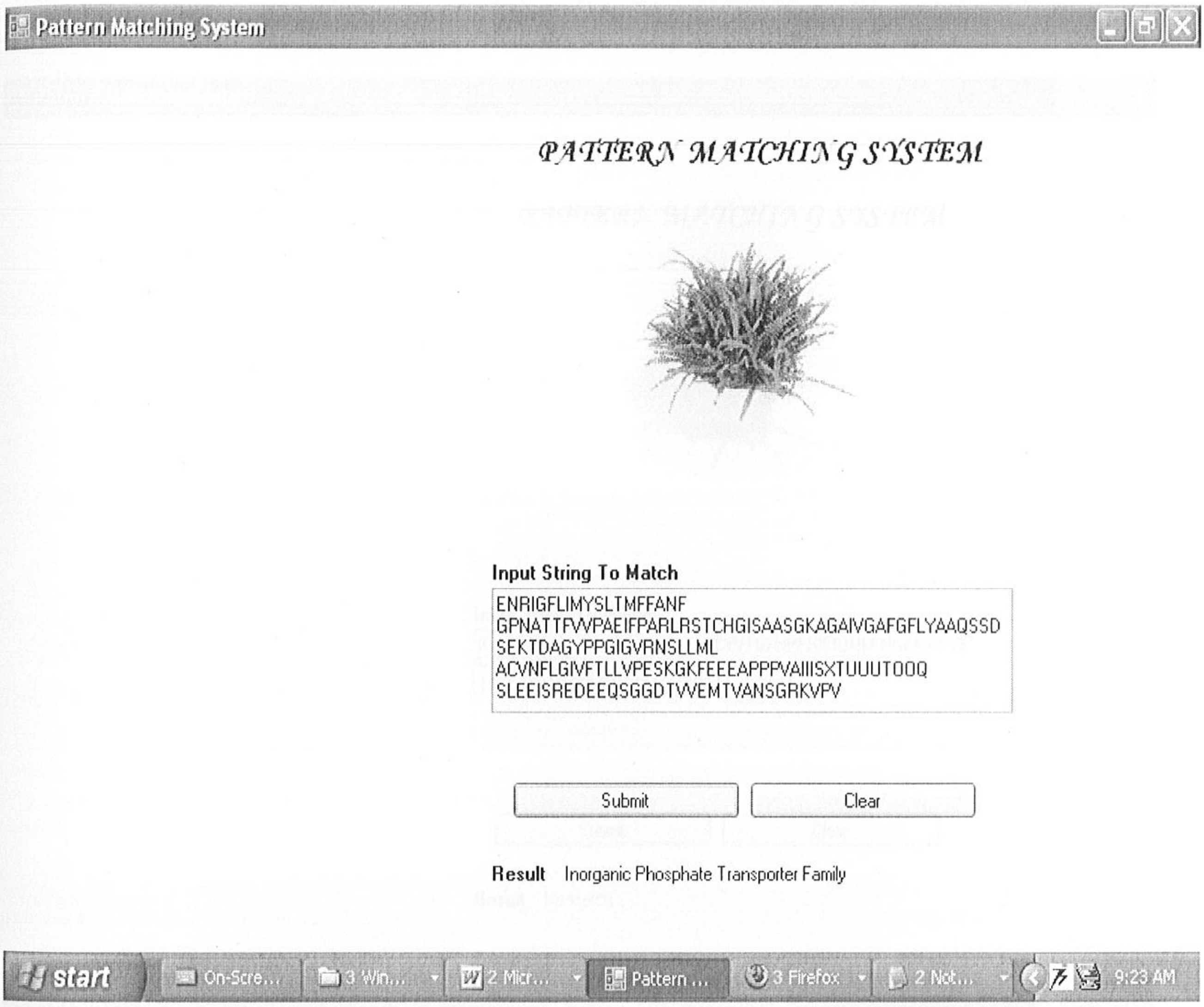


Input String To Match

```
IQKTPQIQVYSRHPPENGKPNILNCYVTQFHPPHIEIQMLKNGKKIPKVE
MSDMSFSKDWSFYILAHTEF
TPTETDTYACRVKHDSMAEPKTVYWDRDM
```

Result No Match

>gi|2780345|dbj|BAA24280.1| inorganic phosphate transporter [Arabidopsis thaliana]

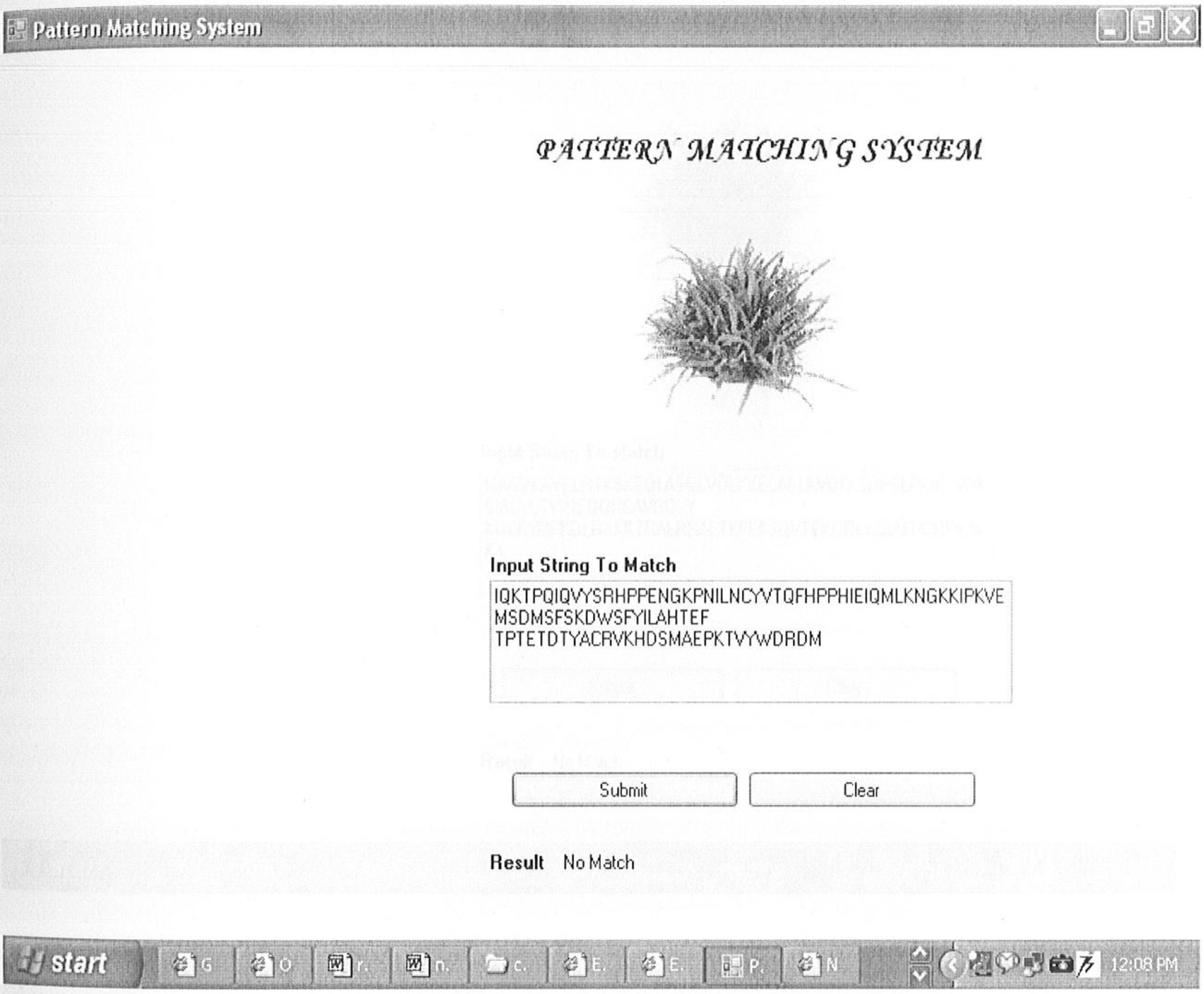


Input String To Match

```
ENRIGFLIMYSLTMFFANF
GPNATTFVVPAEIFPARLRSTCHGISAASGKAGAIVGAFGFLYAAQSSD
SEKTDAGYPPGIGVRNSLLML
ACVNFLGIVFTLLVPESKGKFEEEAPPPVAIIISXTUUUTOOQ
SLEEISREDEEQSGGDTVVEMTVANSGRKVPV
```

Result Inorganic Phosphate Transporter Family

>gi|110591037|pdb|2CLZ|P Chain P, Mhc Class I Natural Mutant H-2kbm8 Heavy Chain Complexed With Beta-2 Microglobulin And Pbm1 Peptide

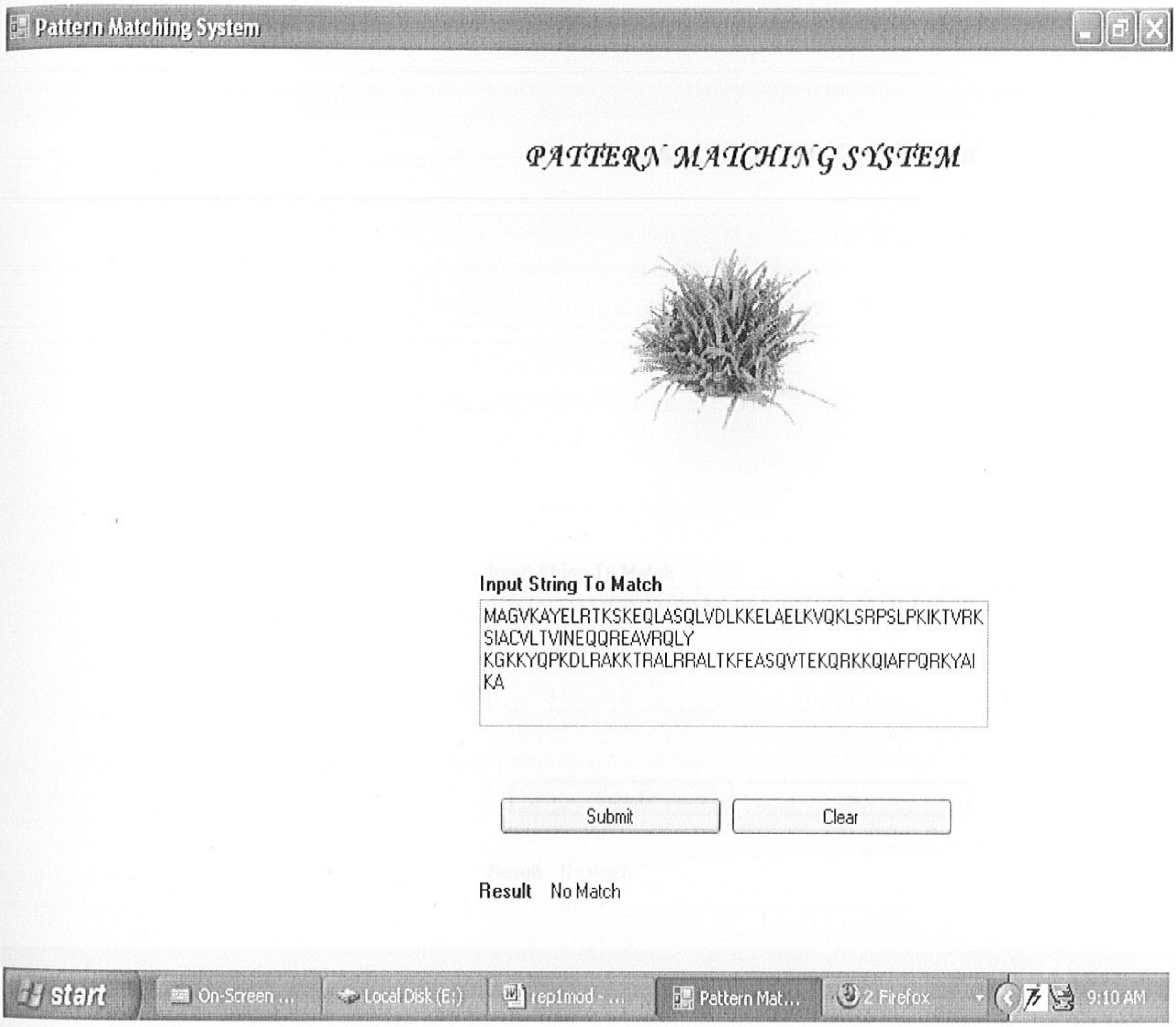


Input String To Match

```
IQKTPQIQVYSRHPPENGKPNILNCYVTQFHPPHIEIQMLKNGKKIPKVE
MSDMSFSKDWSFYILAHTEF
TPTETDTYACRVKHDSMAEPKTVYWDRDM
```

Result No Match

>gi|49258863|pdb|1S1I|X Chain X, Structure of the Ribosomal 80s-Eef2-Sordarin Complex From Yeast

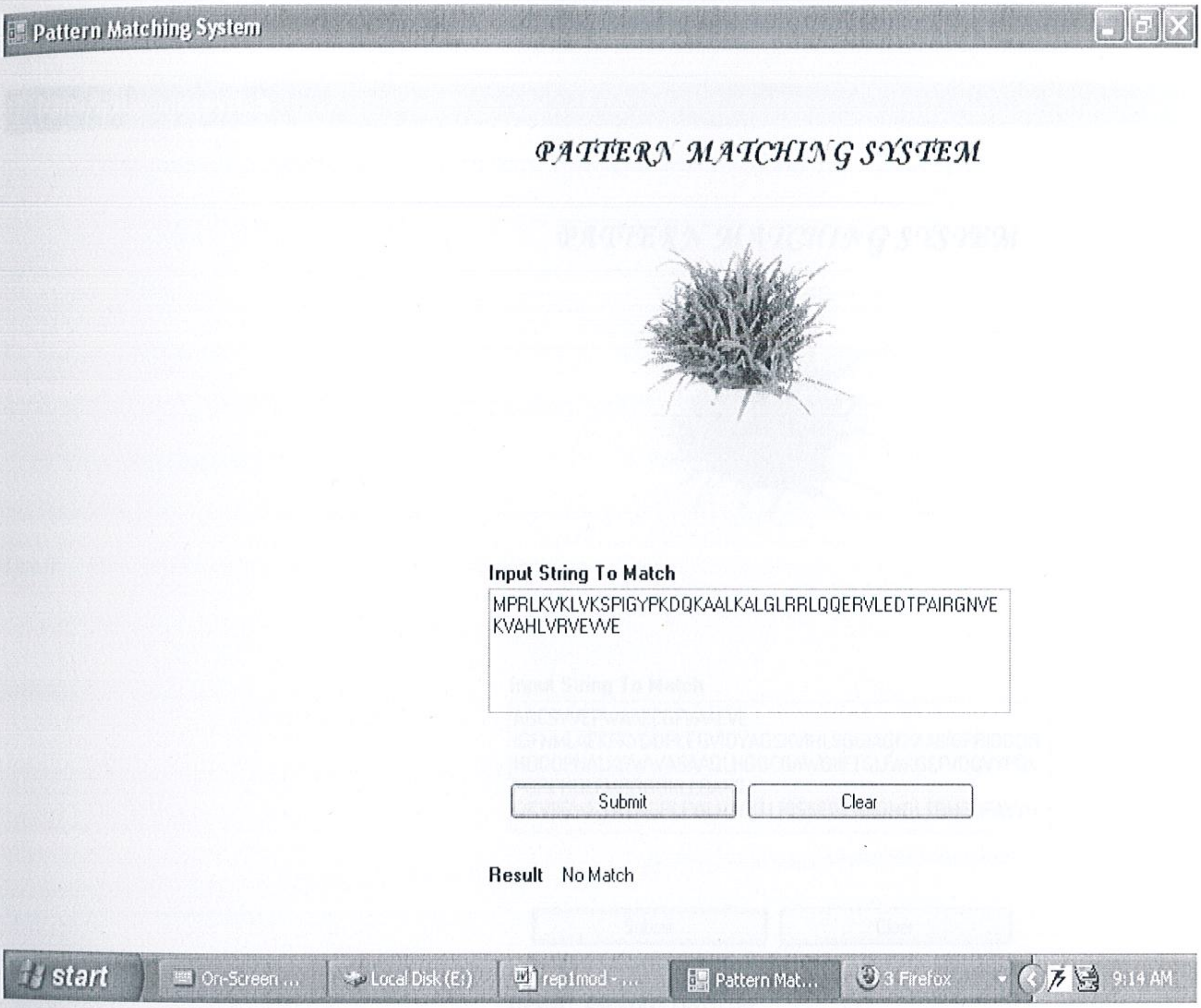


Input String To Match

```
MAGVKAYELRTKSKEQLASQLVDLKKELAELKVQKLSRPSLPKIKTVRK
SIACVLTVINEQQREAVRQLY
KGKKYQPKDLRAKKTRALRRALTKFEASQVTEKQRKKQIAFPQRKYAI
KA
```

Result No Match

>gi|28373740|pdb|1ML5|XX Chain x, Structure of the E. coli Ribosomal Termination Complex

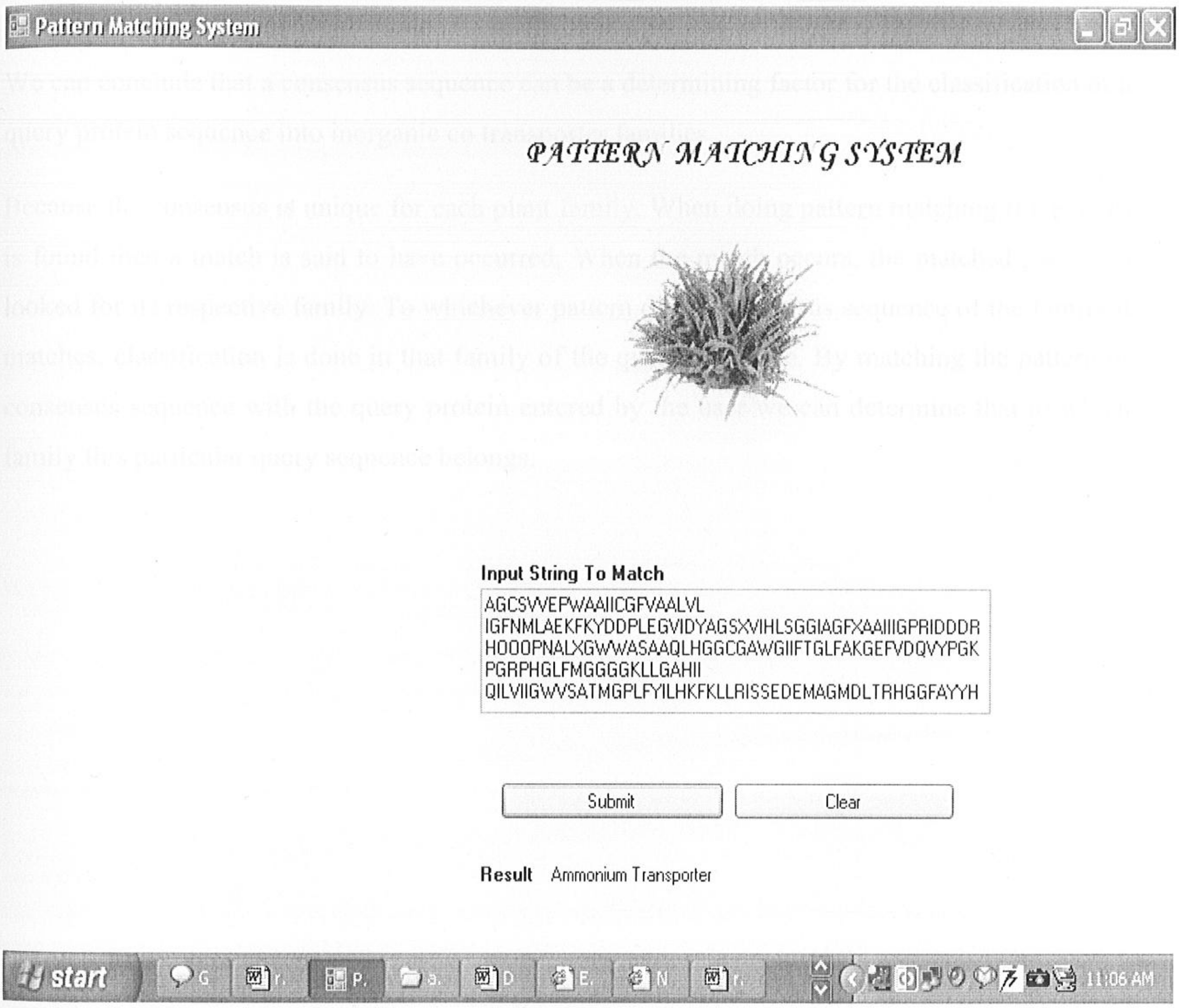


Input String To Match

MPRLKVKLVKSPIGYPKDQKAALKALGLRRLQQERVLEDTPAIRGNVE
KVAHLVRVEVVE

Result No Match

>gi|15217741|ref|NP_176658.1| ATAMT1;2 (AMMONIUM TRANSPORTER 1;2); ammonium transporter



Input String To Match

```
AGCSVVEPWAAIICGFVAALVL
IGFNMLAEKFKYDDPLEGVIDYAGSXVIHLSGGIAGFXAAIIIGPRIDDDR
HOOOPNALXGWWASAAQLHGGCGAWGIIFTGLFAKGEFVDQVYPGK
PGRPHGLFMGGGGKLLGAHII
QILVIIGWVSATMGPLFYILHKFKLLRISSEDEMAGMDLTRHGGFAYYH
```

Result Ammonium Transporter

CHAPTER 4

CONCLUSION.

We can conclude that a consensus sequence can be a determining factor for the classification of a query protein sequence into inorganic co transporter families

Because the consensus is unique for each plant family. When doing pattern matching if a pattern is found then a match is said to have occurred. When the match occurs, the matched pattern is looked for its respective family. To whichever pattern of the consensus sequence of the family it matches, classification is done in that family of the query sequence. By matching the pattern of consensus sequence with the query protein entered by the user we can determine that to which family this particular query sequence belongs.

CHAPTER 5

BIBLIOGRAPHY

Web Pages

http://jxb.oxfordjournals.org/cgi/content/full/53/370/855

http://jxb.oxfordjournals.org/cgi/content/abstract/57/5/1149

http :// www.blackwell-synergy.com/doi/abs/10.1046/j.1365-2958.2001.02264.x

http :// www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22785

http://darwin.informatics.indiana.edu/col/courses/I590-compgen/

http ://www.google.co.in

http://www.ncbi.

http://www.wikipedia.com

Research papers

Tama Christine Fox and Mary Lou Guerinot Molecular biology of cation transport in plants- Vol. 49 : 669-696 (Volume Publication date June 1998).
Department of Biological Sciences.

By R. S. Chauhan Bioinformatics approach toward identification of candidate genes for zinc and iron transporters in maize. Bioinformatics Centre, Advanced Centre of hill Bioresources and Biotechnology, CSK, HP Agricultural University.
current Science, Vol 91, No: 4, 25 Aug 2006.

Frank Delaglio Protein Structure Modeling using Chemical Shifts and Residual Dipolar Coupling Homology, in the Context of an Integrated Software System for Spectral Processing and Analysis 457-461 (1984) Macromolecules.

Darren M Soanes Comparative genomic analysis of phytopathogenic fungi using expressed sequence tag (EST) collections- 2006 BLACKWELL PUBLICATIONS.

CHAPTER 6

GLOSSARY

ALGORITHM-A well defined sequence of steps (procedure) that takes some values as input and guarantees a value as output in some finite no. of steps.

C#- (pronounced "C-sharp") is an object-oriented programming language from Microsoft that aims to combine the computing power of C++ with the programming ease of Visual Basic. C# is based on C++ and contains features similar to those of Java.

.NET platform -C# is designed to work with Microsoft's .NET platform. Microsoft's aim is to facilitate the exchange of information and services over the Web, and to enable developers to build highly portable applications

CODE- A system of symbols and rules used to represent instructions to a computer; a computer program.

DEBUGGING-: The process of eliminating the errors.

GRAPHICAL USER INTERFACE-: Abbreviated GUI (pronounced GOO-ee). A program interface that takes advantage of the computer's graphics capabilities to make the program easier to use. Well-designed graphical user interfaces can free the user from learning complex command languages. On the other hand, many users find that they work more effectively with a command-driven interface, especially if they already know the command language.

APPENDIX

SOURCE CODE-:

```
using System;

using System.Collections.Generic;

using System.ComponentModel;

using System.Data;

using System.Drawing;

using System.Text;

using System.Windows.Forms;


namespace WindowsApplication1

{

  public partial class Form1 : Form

  {

    public Form1()

    {

      InitializeComponent();

    }

    String Family = "";

    //String Family = "L-x(0,2)-A-x(2,3)-V-[AGSV]";

    //String Family = "S-X(1,2)-L";

    //String Family = "S-x(2,4)-G-x(1,3)-L-[AGILV]";
```

```
//String Family = "F-x(3,5)-A-x(3)-[APV]-A-x(3,4)-S-X-[CST]-x(3)-[AST]-x(2)-[AQST]";


String[] arr;
String txtToSearch ;
int s = 0;


private void abc()
{
    //MessageBox.Show("DataBindings");
    Boolean finalCheck = false;

    Family = "L-x(0,2)-A-x(2,3)-V-[AGSV]";
    arr = Family.Split('-');
    txtToSearch = textBox1.Text.Trim();
    //s = 0;
    //Search();
    for (int k = 0; k < txtToSearch.Length - 1; k++)
    {
        s = k;
        if (Search() == true)
        {
            finalCheck = true;
            // MessageBox.Show("Putative Sulfate Transporter Family");
    label3.Text = "Putative Sulfate Transporter Family";
    break;
  }
  else
  {
    // MessageBox.Show("No Match");
    label3.Text = "No Match";

  }
}
if (finalCheck == false)
{
  Family = "S-x(1,2)-L";
  arr = Family.Split('-');
  for (int k = 0; k < txtToSearch.Length - 1; k++)
  {
    s = k;
    if (Search() == true)
    {
      finalCheck = true;
      //MessageBox.Show("Zinc Transporter Family");
      label3.Text = "Zinc Transporter Family";
      break;
    }
```

```
        else
        {

            // MessageBox.Show("No Match");
            label3.Text = "No Match";

        }
    }

}
if (finalCheck == false)
{
    Family = "S-x(2,4)-G-x(1,3)-L-[AGILV]";
    arr = Family.Split('-');
    for (int k = 0; k < txtToSearch.Length - 1; k++)
    {
        s = k;
        if (Search() == true)
        {
            finalCheck = true;
            // MessageBox.Show("Sodium Transporter Family");
            label3.Text = "Sodium Transporter Family";
            break;
        }
        else

        {



          // MessageBox.Show("No Match");

          label3.Text = "No Match";



        }

      }



    }

  if (finalCheck == false)

    {

      Family = "F-x(3,5)-A-x(3)-[APV]-A-x(3,4)-S-X-[CST]-x(3)-[AST]-x(2)-[AQST]";

      arr = Family.Split('-');

      for (int k = 0; k < txtToSearch.Length - 1; k++)

      {

        s = k;

        if (Search() == true)

        {

          finalCheck = true;

         // MessageBox.Show("Inorganic Phosphate Transporter Family");

          label3.Text = "Inorganic Phosphate Transporter Family";

          break;

        }
```

```
                    else

                    {

                        // MessageBox.Show("No Match");

                        label3.Text = "No Match";

                    }

                }

            }

            if (finalCheck == false)

            {

                Family = "S-x(2,3)-T-x(2)-[AV]-S-x(2)-A-[AL]-X-[APV]-[AL]-x(2)-[IL]-X-[DR]-
x(2)-[ALP]-x(2)-[AST]-x(2)-[DPQS]";

                arr = Family.Split('-');

                for (int k = 0; k < txtToSearch.Length - 1; k++)

                {

                    s = k;

                    if (Search() == true)

                    {

                        finalCheck = true;

                        // MessageBox.Show("High Affinity Nitrate Transporter Family ");

                        label3.Text = "High Affinity Nitrate Transporter Family ";

                        break;
                    }

                    else

                    {

                        // MessageBox.Show("No Match");

                        label3.Text = "No Match";

                    }

                }

            }

            if (finalCheck == false)

            {

                Family = "G-V-I-D-[FY]-[AS]-G-[GS]-X-V-[IV]-H-[LM]-[SV]-[GS]-G-[IV]-A-G-
[FL]-X-[AG]-A-x(3)-G-P-R-[IP]-x(3)-[DR]-[EHR]-x(3)-[PS]-[NV]-[AN]-[LV]-X-[GL]-x(2)-A-
[GS]";

                arr = Family.Split('-');

                for (int k = 0; k < txtToSearch.Length - 1; k++)

                {

                    s = k;

                    if (Search() == true)

                    {

                        finalCheck = true;
                // MessageBox.Show("Ammonium Transporter ");
                label3.Text = "Ammonium Transporter ";
                break;
            }

            else
            {
                // MessageBox.Show("No Match");
                label3.Text = "No Match";
            }
        }
    }
}

private Boolean Search()
{
    Boolean fla = true;
    for (int temp = 0; temp < arr.Length; temp++)
```

```
{
   try
   {
      if (check(arr[temp]))
      {

      }
      else
      {
         fla = false;
         break;
      }
   }
   catch (Exception ex)
   {
      fla = false;
      break;
   }
}
if (fla==true)
{
   //MessageBox.Show("match");
   return true;
}
else
```

```
        return false;
    }

    private Boolean check(String val)
    {
        char c;
        if (!(s >= txtToSearch.Length))
        {
            if (val.Length == 1 && val != "x")
            {
                c = char.Parse(val);
                if (c == txtToSearch[s])
                {
                    s++;
                    return true;
                }
            }
            else if (val[0] == '[')
            {
                Boolean f = false;
                String v = val.Substring(1, val.Length - 2);
    for (int t = 1; t <= val.Length - 2; t++)
    {
        if (txtToSearch[s] == val[t])
        {
            s++;
            f = true;
            break;
        }
    }
    if (f == true)
        return true;
    else
        return false;
}

else if (val[0] == 'x')
{

    String vx = String.Empty;
    int j;
    int i;
    if (val.Length > 4)
    {
        int tak = s;
```

```
String v = val.Substring(2, 1);
int first = int.Parse(v);
int sec = int.Parse(val.Substring(4, 1));
if (first > sec)
{
  j = sec;
  i = first;
}
else
{
  i = sec;
  j = first;
}
Boolean flag = false;
char vc = txtToSearch[s];
int temp = s;
for (int t1 = 1; t1 < i; t1++)
{
  if (vc == txtToSearch[temp + 1])
  {
    temp++;
    flag = true;
  }
  else
  {
        flag = false;
        break;
    }
}
if (flag == true)
{
    s = temp + 1;
    return true;
}
else
{
    flag = false;
    if (j != 0)
    {
        temp = s;

        for (int t1 = 1; t1 < j; t1++)
        {
            if (vc == txtToSearch[temp + 1])
            {
                temp++;
                flag = true;
            }
            else
            {
                flag = false;
                break;
            }
        }
        if (flag == true)
        {
            s = temp + 1;
            return true;
        }
        else
            return false;
    }
    else
        return true;
}
//    if (flag == false)
//    {
//        if (j != 0)
//        {

//            tak = s;
//            for (int t1 = 1; t1 < i; t1++)
//            {
//                if (vc != txtToSearch[s + 1])
```

```
  //        {
  //          s = tak;
  //          return false;

  //        }
  //        else
  //          s++;

  //      }
  //    }
  //    else
  //      return true;
  //  }
  //  else
  //  {
  //    s++;
  //    return true;
  //  }
  //}

}
else
{
  int tak = s;
  String v = val.Substring(2, 1);
```

```
int first = int.Parse(v);
Boolean flag = false;
char vc = txtToSearch[s];
int temp = s;

for (int t1 = 1; t1 <first; t1++)
{
    if (vc == txtToSearch[temp + 1])
    {
        temp++;
        flag = true;
    }
    else
    {
        flag = false;
        break;
    }
}
if (flag == true)
{
    s = temp + 1;
    return true;
}
else
    return false;
```

```
            }

        }
    }
        return false;

    }

private void btnsubmit_Click(object sender, EventArgs e)
{
    abc();
}

private void Form1_Load(object sender, EventArgs e)
{

}

private void button1_Click(object sender, EventArgs e)
{
    textBox1.Text = "";

}
```

```
          //if (char.IsLetter(val) && val != "x")
          //{

          //}

        }

    }
```

CODING FOR THE FRONT END-:

```
namespace WindowsApplication1
{
    partial class Form1
    {
        /// <summary>
        /// Required designer variable.
        /// </summary>
        private System.ComponentModel.IContainer components = null;

        /// <summary>
        /// Clean up any resources being used.
        /// </summary>
        /// <param name="disposing">true if managed resources should be disposed; otherwise, false.</param>
        protected override void Dispose(bool disposing)
        {
            if (disposing && (components != null))
            {
                components.Dispose();
            }
            base.Dispose(disposing);
        }
```

```
#region Windows Form Designer generated code

/// <summary>
/// Required method for Designer support - do not modify
/// the contents of this method with the code editor.
/// </summary>
private void InitializeComponent()
{
    System.ComponentModel.ComponentResourceManager resources = new
System.ComponentModel.ComponentResourceManager(typeof(Form1));
    this.textBox1 = new System.Windows.Forms.TextBox();
    this.btnsubmit = new System.Windows.Forms.Button();
    this.label1 = new System.Windows.Forms.Label();
    this.label2 = new System.Windows.Forms.Label();
    this.label3 = new System.Windows.Forms.Label();
    this.pictureBox1 = new System.Windows.Forms.PictureBox();
    this.label4 = new System.Windows.Forms.Label();
    this.button1 = new System.Windows.Forms.Button();
    ((System.ComponentModel.ISupportInitialize)(this.pictureBox1)).BeginInit();
    this.SuspendLayout();
    //
    // textBox1
    //
    this.textBox1.Location = new System.Drawing.Point(327, 325);
    this.textBox1.Multiline = true;
```

```
        this.textBox1.Name = "textBox1";

        this.textBox1.Size = new System.Drawing.Size(348, 75);

        this.textBox1.TabIndex = 0;

        //

        // btnsubmit

        //

        this.btnsubmit.Location = new System.Drawing.Point(341, 441);

        this.btnsubmit.Name = "btnsubmit";

        this.btnsubmit.Size = new System.Drawing.Size(152, 23);

        this.btnsubmit.TabIndex = 1;

        this.btnsubmit.Text = "Submit";

        this.btnsubmit.UseVisualStyleBackColor = true;

        this.btnsubmit.Click += new System.EventHandler(this.btnsubmit_Click);

        //

        // label1

        //

        this.label1.AutoSize = true;

        this.label1.Font = new System.Drawing.Font("Microsoft Sans Serif", 8.25F,
    System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)(0)));

        this.label1.Location = new System.Drawing.Point(324, 309);

        this.label1.Name = "label1";

        this.label1.Size = new System.Drawing.Size(131, 13);

        this.label1.TabIndex = 2;

        this.label1.Text = "Input String To Match";
        //
        // label2
        //
        this.label2.AutoSize = true;
        this.label2.Font = new System.Drawing.Font("Microsoft Sans Serif", 8.25F,
   System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)(0)));
        this.label2.Location = new System.Drawing.Point(324, 490);
        this.label2.Name = "label2";
        this.label2.Size = new System.Drawing.Size(43, 13);
        this.label2.TabIndex = 3;
        this.label2.Text = "Result";
        //
        // label3
        //
        this.label3.AutoSize = true;
        this.label3.Location = new System.Drawing.Point(373, 490);
        this.label3.Name = "label3";
        this.label3.Size = new System.Drawing.Size(0, 13);
        this.label3.TabIndex = 4;
        //
        // pictureBox1
        //
        this.pictureBox1.Image =
   ((System.Drawing.Image)(resources.GetObject("pictureBox1.Image")));
        this.pictureBox1.Location = new System.Drawing.Point(423, 116);
        this.pictureBox1.Name = "pictureBox1";
        this.pictureBox1.Size = new System.Drawing.Size(153, 120);
        this.pictureBox1.TabIndex = 5;
        this.pictureBox1.TabStop = false;
        //
        // label4
        //
        this.label4.AutoSize = true;
        this.label4.Font = new System.Drawing.Font("Monotype Corsiva", 15.75F,
((System.Drawing.FontStyle)((System.Drawing.FontStyle.Bold |
System.Drawing.FontStyle.Italic))), System.Drawing.GraphicsUnit.Point, ((byte)(0)));
        this.label4.Location = new System.Drawing.Point(351, 49);
        this.label4.Name = "label4";
        this.label4.Size = new System.Drawing.Size(310, 25);
        this.label4.TabIndex = 6;
        this.label4.Text = "PATTERN MATCHING SYSTEM";
        //
        // button1
        //
        this.button1.Location = new System.Drawing.Point(499, 441);
        this.button1.Name = "button1";
        this.button1.Size = new System.Drawing.Size(152, 23);
        this.button1.TabIndex = 7;
        this.button1.Text = "Clear";
        this.button1.UseVisualStyleBackColor = true;
        this.button1.Click += new System.EventHandler(this.button1_Click);
        //
```

```
// Form1
//
this.AutoScaleDimensions = new System.Drawing.SizeF(6F, 13F);
this.AutoScaleMode = System.Windows.Forms.AutoScaleMode.Font;
this.BackColor = System.Drawing.SystemColors.ActiveCaptionText;
this.ClientSize = new System.Drawing.Size(740, 516);
this.Controls.Add(this.button1);
this.Controls.Add(this.label4);
this.Controls.Add(this.pictureBox1);
this.Controls.Add(this.label3);
this.Controls.Add(this.label2);
this.Controls.Add(this.label1);
this.Controls.Add(this.btnsubmit);
this.Controls.Add(this.textBox1);
this.Name = "Form1";
this.StartPosition = System.Windows.Forms.FormStartPosition.CenterScreen;
this.Text = "Pattern Matching System";
this.WindowState = System.Windows.Forms.FormWindowState.Maximized;
this.Load += new System.EventHandler(this.Form1_Load);
((System.ComponentModel.ISupportInitialize)(this.pictureBox1)).EndInit();
this.ResumeLayout(false);
this.PerformLayout();

}
```

```
#endregion

 private System.Windows.Forms.TextBox textBox1;
 private System.Windows.Forms.Button btnsubmit;
 private System.Windows.Forms.Label label1;
 private System.Windows.Forms.Label label2;
 private System.Windows.Forms.Label label3;
 private System.Windows.Forms.PictureBox pictureBox1;
 private System.Windows.Forms.Label label4;
 private System.Windows.Forms.Button button1;
}
}
```