

Utilization of Computer Programs and Molecular Biology in Documentation of Egyptian clover

Abd alaziz.T. Bondok¹, Shereen M. ElNahrawy¹, and Ahamed Foud²

ABSTRACT

Egyptian clover varieties are part of a very important agricultural family, second only to cereals. In this paper, the materials under the present investigation consisted of three biological datasets, containing different protein characteristics annotated from interpro, prints and quick Godatabasesthey were selected and have been applied to data from XML flat files into a local database of size 10MB in oracle tool which is used throughout the experiments Shima A. Badawy(2013).Moreover, in the present study, a new program was designed and applied which showed high flexibility and efficient of mining the hierarchy (Molecular Biology data of *T. alexandrinum* L.). In addition describe protein characteristics including family of protein, finger print of protein of *Trifoliumalexandrinum* L.

Keywords: Molecular biology, computer programs, documentation, *Trifoliumalexandrinum* L., hierarchy databases, sequences, framgments, protein.

INTRODUCTION

The genus *Trifolium* comprises approximately 290 species. While, *T. repens*, *T. Pratense*, *T. nigrescens*, etc., are important constituents of temperate pastures, *T. alexandrinum* and *T. resupinatum* are cultivated as winter annual fodder in the tropical and sub-tropical belt. *Trifoliumalexandrinum*, commonly known as berseem or Egyptian clover is an important winter annual fodder legume cultivated in Egypt Khaled Y. Abdel-Halim (2014).

In the present study an attempt was carried out have tried to develop methods of the Egyptian clover documentation through working on two levels:

- Moleclar biology data.
- Databases of computer programs.

Databases are the heart of computer programs essentially they are electronic filing cabinets that offer a convenient and efficient method of storing vast amounts of information. There are many different databases types, depending on the nature of the information being stored (e.g, sequences, structures, etc.).The number of different databases is growing very rapidly. During the year 2000, 55 new databases were created, bringing the total at the end of the year to 2811.

Documentation contains data relatedto molecular biology characteristics and data bases of computer programs characteristics related to Egyptian Clover.

The Reasons for doing documentation on the Egyptian cloverare to:

- Avoiding genetic erosion which has also been caused by the replacement of domestic cultivars of Egyptian clover by improved cultivars with a narrow genetic base (Hawkes, 1983).

Moreover, Gene Banks will have the fact that genetic variability allows populations to adapt to environmental changes is valied for all organisms including Egyptian clover,etc and microflora where the evaluation course is quite variable.The evolutionary process that maintained this diversity in the past is unable to surve in the present technological era. The present abundance of genetic diversity which still survives is being threatened by a combination of population pressures, adverse economic conditions and the interaction among these factors and subsequently further deterioration of genetic resources (Egyptian Clovercultivars).

The science of genetics in general and precisely the conservation genetics should play a sizable part to minimize the effect of these risks (Abdel Salam et al., 1994).

In Egypt, The population is rapidly increasing and at the same time these is a big gap between the food and feed needs and the available production of the crops. This due to the limited natural recourses especially the water resources. The visible way to narrow this gap is through vertical development of our agricultural system. High yielding varieties of various crops are playing very efficient roles in increasing the agricultural production vertically.Our objective in current research paper was developing high yielding cultivars of Egyptian clover

To develop varieties so, this is a need to use of huge information in molecular biology and computer science area to make databases on DNA hierarchy and amino acids sequences levels in Egyptian clover varieties to create a new varieties in a short time.

¹Forage crops Res. Department, field Crops Res. institute, ARC,Giza, Egypt.

²Faculty of computer and information, Ain Shams univ., Cairo, Egypt.

Received February 8, 2016, Accepted June 23, 2016

The aims of the present study were:

- Designed and applied new database for mining the hierarchy related to Egyptian Clover. This database will enable to describe protein attributes including family of protein and its fingerprint related to Egyptian clover.
- Determination of best programs which might be used efficiently in documentation of Egyptian clover germplasm.
- In future goal, update of the breeding methods of Egyptian clover species employing computer programs.

MATERIALS AND METHODS

Plant material:

Plant material of Egyptian clover was obtained from the Forage Crops Res. Depart. Field Crops Res. Institute, ARC, Giza Egypt.

SDS – PAGE electrophoresis:

Total protein content was determined in grounded fine powder seeds of each sample by the method described by Bradford (1976) using bovine serum albumin (96%, Sigma Chemical Co, St. Louis, MO, USA), as standard, then total soluble proteins were extracted with extraction buffer. Fifty μ l of the extract were mixed with 50 of SDS, 5% v/v β -mecaptothanol, 7% v/v glycerol and 0.03% bromphenol blue and boiled for 7min in a boiling water bath. 14 μ l of the sample was loaded on to each well.

Electrophoresis SDS-PAGE was carried out according to the procedures of Laemmli (Laemmli, 1970) in 1.5 mm thick gels with 14% (w/v) separating gel and 4% (w/v) stacking gel in a vertical

electrophoresis unit (Cleaver Scientific, England). SDS – PAGE was carried out at 75 volt for 3 hours. After electrophoresis, the gels were over night stained using 0.1% (w/v) Coomassie Brilliant Blue R-250. Then, destained using a 10%(v/v) acetic acid solution until a clear background was achieved. A page ruler pertained protein ladder (thermo – Fisher Scientific) was used as protein molecular weight marker.

Gel documentation system (Geldoc – It^e imaging system, uvp, England), was applied for data scoring and documentation. Total lab analysis software (Total Lab TL120,V2008)was employed for constructingbinary matrix for SDS PAGE data according to presence or absence of a band of each sample which remarked as I or O.

DATA BASE LOADER

The main issue now is how to get available biological data on to a local database in order to perform different kinds of computation.

In fact, the three discussed databases, above, are available in XML flat files to down load via the web,interpro, prints and quick go. The whole process is illustrated in Fig (1). First, XML flat files are parsed by a Java or C++.

Parser program to extract table definitions tages and data. Second extracted data is loaded into oracle database, using a DBLoader<http://www.ailab25.engr.uconn.edu>: in order to produce a local biological database, called BioDB. Note that one has to determine if he wants to process all data or part of it, database loader application in the present study is performed to the molecular biology data.

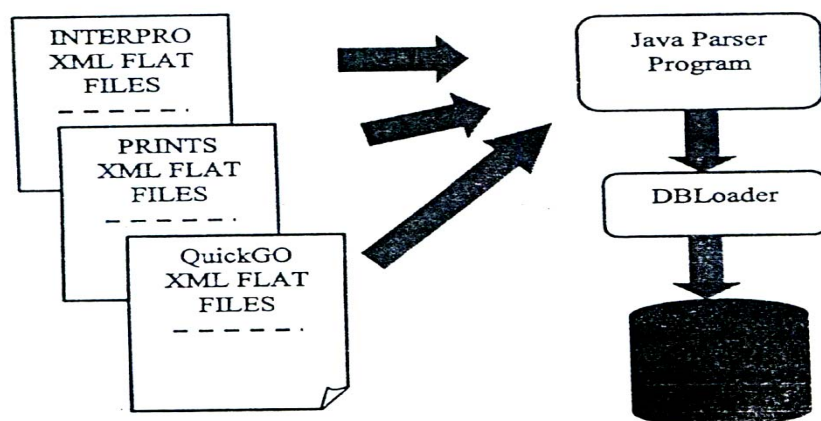


Fig. 1. Preprocessing to create local database (BioDB)

Inter pro database:

Inter pro is one of the database with signature diagnostic for protein families, domains, repeats, or functional sites, which amalgamates the efforts of prosite, prints, Pfam (Batman et al, 2000), and pro dom (Corpet et al., 2000) database projects.

Prints database:

Prints database has been the international project to cooperate with prosite, pfam, and pro Dom databases. This database houses a collection, of fingerprint information for protein families, fingerprints are groups of motifs (i.e., specific protein sequences fragments) that could be inferred by aligning similar sequences (Attwood et al., 2002).

Quick Godatabase:

Quick go contains information about the gene ontology produced by the gene ontology(Go) consortium (Ashburner, et al, 2000). The gene ontology components of IPR000276. Gene ontology part of biological process of IPR 000276 is shown. Each gene ontology components has its own go entryid, in the form of go xxxx with x as digits.

RESULTS AND DISCUSSIONS**Molecular Biology Data****Biological data:**

Biological data of *Trifoliumalexandrinum* is an informative -rich domain, where data needs to be analyzed. Biological data of *Trifoliumalexandrinum* has been chosen for two reasons: the first one is its built-in hierarchy in most of the databases available as will be illustrated in details in the next section: The other reason is the use of biological data is timely as the information community is in great need for using data mining techniques to predict, for example gene, functions by analyzing data cumulated from diversified sources and protein functions as well in the present study used three available databases.

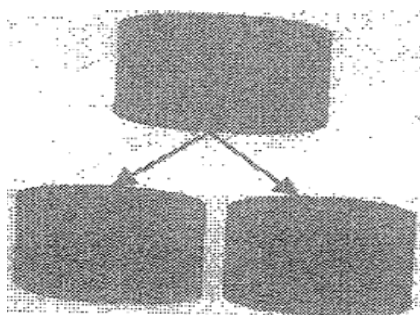


Fig. 2. Three used biological databases

- Interpro database:

Interproof *Trifoliumalexandrinum* is one of the databases with signatures diagnostic for protein families, domains, repeats, or functional sites (Apweiler et al 2000) which amalgamates the efforts of PROSITE (Hoffman et al, 1999), PRINTS' (Attwood et al, 2002), Pfam(Batman et al, 2000), and ProDom(Corpet et al, 2000) database projects. It is a vital tool for the computation of functional classification of newly determined sequences of *Trifoliumalexandrinum* that lack biochemical characterization. Interproof *Trifoliumalexandrinum* has been developed to rationalize protein family characterization and inherit functional insights in order to discover new functionalities. Interproof *T.alexandrinum* provides interface for both text-and sequence-based searches. Overlapping domains, signatures or profiles describing common domains or protein families were merged into a single Interpro entry with a unique accession number (which takes the form IPRxxxx, where x is a digit). An example of an Interpro of *T.alexandrinum* search results is shown in Fig. (3) Other links of protein family signatures are also provided, such as Pfam database (PF00001), PRINTS database (PR00237), and PROSITE (PS00237). The type of ID is specified whether it is a protein family domain repeat, or functional site, where it is a family type in the current case.

Figure (3) also illustrates family tree if there exists one, whether it is a-family or a domain type. Hierarchical information in the interproof *Trifoliumalexandrinum* database includes both 'is-a' (parent -child) and contains found -in information for example, in 'is-a' relationship if there exists an edge between two nodes we can say that one of them is a parent of the other.

- PRINTS database:

PRINTS of *Trifoliumalexandrinum* database contains information about the Pfam, and ProDom databases. This database houses a collection of fingerprint information for protein families (Attwood et al, 2002). Fingerprints of *Trifoliumalexandrinum* are groups of motifs (i.e., specific protein sequences fragments) that could be inferred by aligning similar sequences. Those motifs characterize aligned family and provide specific diagnostic signature. Fingerprints are more powerful than single motif approaches. The technique used to collect fingerprints is to discriminate patterns in a hierarchical form, i.e., protein sequences. Such a hierarchical approach has been used to, resolve G-protein-Coupled -Receptors (GPCRs) super families into their constituent families and receptors subtypes and to classify a variety of channel proteins,

transporters, and enzymes. Fig. (4) shows the results for family ID IPR000276 fingerprint at PRINTS. Each fingerprint of *Trifolumalexandrinum* entry has its own PRINTs entry ID< called accession number, which is PR00237 in this cases.

- QuickGO Database:

QuickGO of *Trifolumalexandrinum* contains information about the Gene ontology, produced by the gene Ontology (GO) Consortium (Ashburner et al 2000)/ The Gene ontology components of IPR000276 are shown in Fig. (6) where they are linked to interpro database. In Fig.(5) (a), Gene Ontology part of Biological process. IPR000276 is shown. Each Gene

Ontology of *Trifolumalexandrinum* component has its own Go-entry id, in the form of GO-XXXXX with x as digits. Representation of hierarchical knowledge has widely seen as an important aspect in the design of a formal ontology. Gene Ontology structure is represented as a directed-Acyclic Graph (DAG) that represents a network rather than a tree. Each node can be a child or a parent; a child may have more than one parent. There are two types of parents in a hierarchical structure that a node can have; is -relationship or 'part-of relationship. A node can have more than one parent of one type 'is-a/'part-of relationship or a mixed of both and part-of relationships,

Access. Number	Number	Matthes: 6914 proteins
Signatures of protein families	Signatures	Rhodopsin-like GPCR superfamily PF0001:7tm -1 PR00237:GPCRRHODOPSN PS'00237:GproteinJR Recep-FI-1 PS50262:Gprotein recap FI-2.
Part of children tree of GPCR superfamily	Type	Family
Gene ontology production	Children	IPR00025;Melatonin Receptors IPR00174;interleuk in-8 Receptor
	Erocess	G;Protein coupled receptor protein signaling pathway (GO:0007186)
	Function	Rhodopsin-like receptor activity (GO.005164)
	Component	Integral to membrane (GO:0016021)
	Database Links	Blocks:BL00.237

Fig. 3. Interporof *T. alexandrinum* Examples of IPR000276 Search results

No. of motifs	Accession	Gpcrrhodopsn (view relation] [view alignment] [view structure]
included in the fingerprint,	No of moieties	PR00237
Literature references	Creation Date	7
where discussion about motifs appears	Title	12-JUL-2010
Part of IDs of protein Matches for	Database References	Rhodopsin-lik'e superfamily signature (GPCR)
	'Literature References	Interpro:IPR000276,BLOCKS:BL002 37:Pfam:PF00001
		1.ATTWOOD,T.K. AND FINDLAYJ.B.C. Fingerprinting G protein-coupled receptors PROTEIN ENG.7(2) 195 -203 (1994)
		G protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide range of functions.
		G protein-coupled receptors (GPCRs) constitute a vast protein family that encompasses a wide range of functions.

Fig. 4. Shows the result for family IDIPR 000276 finger print at PRINTS of *T.alexandrinum*

The is-a' relationship or 'part-of relationship, node can have more than one parent of one type is-a'/part-of relationship refers to 'when a child is an instance of the parent and the 'part of relationship. These types of relationships of *Trifoliumalexandrinum* are available for the three, extensions of gene product; the molecular function, the biological processes and the cellular component. For example, in Fig. (5) a tree term, which has a (P) in front of it, means this term' is a part of the above term. However, a tree term, which has an (I) in front of it, means this term is a child of the term above it. Fig. (6) illustrates a part of the biological process of IPRG000276; a biological process is a part of the Gene ontology, but a cellular process is a biological process

and a cell communication is a cellular process, etc. In Fig. (5), which shows the description of molecular function of IPR000276, molecular function is a part of the Gene ontology, but signal transducer activity is a molecular function. Finally, each item described in the tree of biological process, molecular function, or cellular component, may have other children, which applies to the same rules as their parents.

Tables Created at BioDB:

BioDB of *Trifoliumalexandrinum* contains a database schema from annotating information of the three databases as illustrated in Fig. (7), which consist of 6 Tables as follows:

		GO:0007186
		G-protein coupled receptor protein signaling pathway
Parent tree	Tree	Gene ontology (GO:0003673) (P) Biological process (GO:0008150) (1): Cellular process (1) Cell communication (1) signal transduction
	Child terms	- Signal transduction during conjugaton with cellular fusion. - Signal transduction during conjugation without cellular fusion
Part of child terms	Interpro Mappings	- Psin - Chemokine receptor - G-Protein, Gamma Subunit

Fig. 5. Part of Quick GO of *T.alexandrinum* reference for biological process of IPR000276 (GO:0007186) (HTTP://WWW.geneontology.org)

GO:0001584	
Rhodopsin-like receptor activity	Name
Gene ontology (GO:0003673)	Tree
(P) Molecular function (GO: 0008150)	
(1): signal transducer activity	
(1) Receptor activity	
(1) transmembrane receptor activity	
(I) G-protein coupled receptor Activity	
(I) Rhodopsin-like receptor Activity	
- Nucelotide receptor activity	Child terms
- Viral receptor activity	
- Amine receptor activity	
- Opsin	Interpro Mappings
- Histamine H4 receptor	
- Perojpsin	

Fig. 6. Part of QuickGO reference for molecular function of IPR000276 (GO: 0001584)

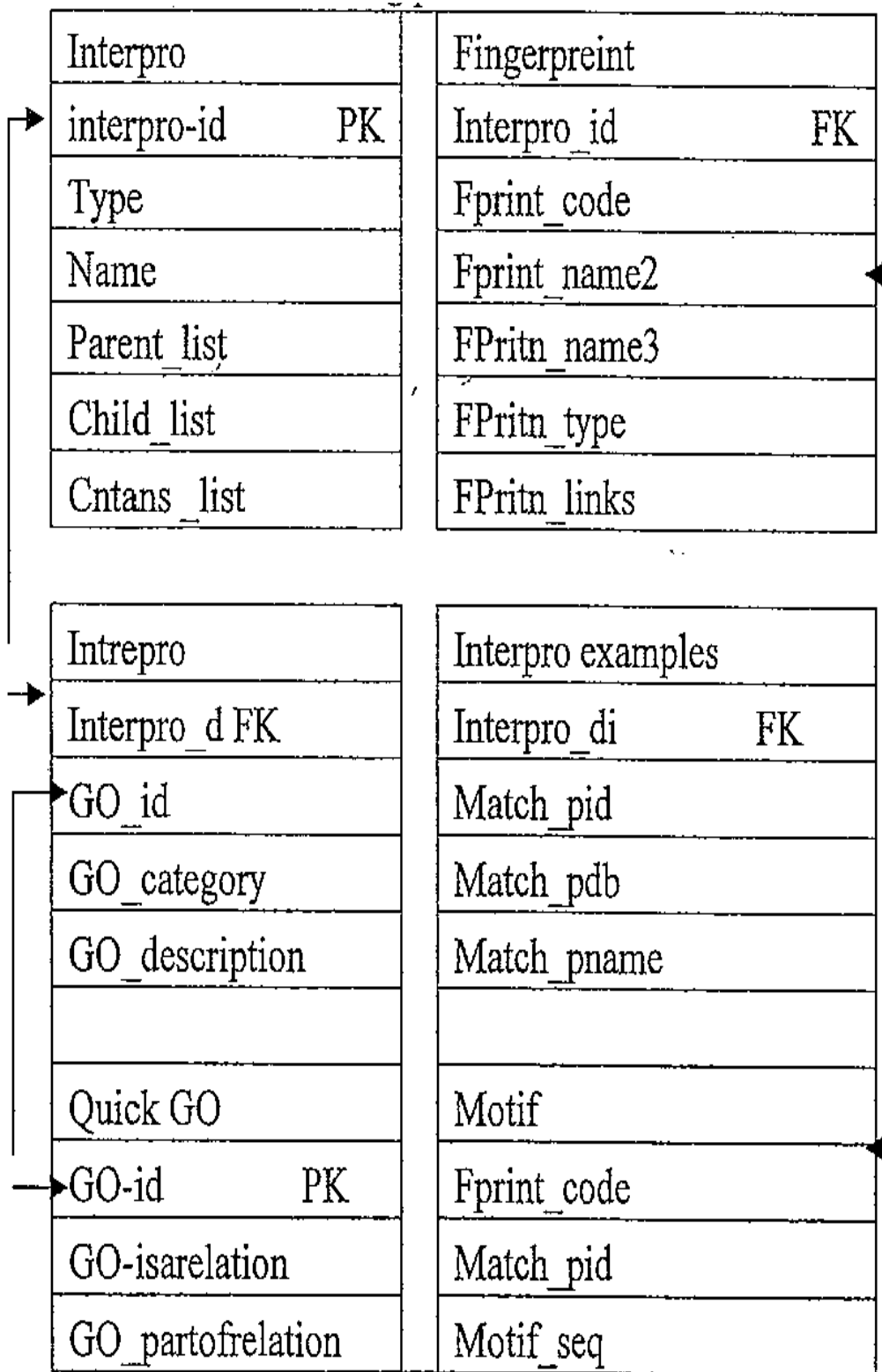


Fig. 7. BioDatabase of *Trifolium alexandrinum* Schema used in experiments

A- Interpro Table of *Trifolumalexandrinum*:

Which includes Interpro entry ID, (primary key), type of the entry (family, domain, etc), parent list/child list (the relationship used to indicate true protein family/subfamily relationships), contains list the relationship used to indicate domain composition. Some domains can be found in more than one type of protein or family of proteins, but is not a SUBTYPE in the family.

B- INTERPRO-GO Table of *Trifolumalexandrinum*:

That includes Interpro entry ID. Interpro entry may have multiple Gene ontology (GO) annotations, hence the possible duplications of Interpro-ID in this table. Not all Interpro entries are annotated by GO. Another attribute in Interpro,-GO also includes GO-ID, which may be associated with multiple Interpro entries as different proteins may share similar functions. For example, in Table (1) Interpro-id IPR000003 has three molecular functions; DNA binding (GO:0003677), Ligand-dependent nuclear receptor and GO-category describes functional classification of the entry, where Gene Ontology (GO) terms are described in' three Molecular biology

categories: Molecular function, Cellular Component, Biological Process. Finally, " GO-description gives Brief annotation about the function of the interpro entry. An example of the outcome of INTER-PRO-GO table is given in Table (1).

C- FINGERPINT table of *Trifolumalexandrinum*:

In table (2) lists protein fingerprint fields from PROSITE, PFAM, PRODOM, and PRINTS to each Interpro entry. It includes Inte-rpro-id, Fprint-code, Fprint -name2, F print-name3, Fprint-type and print-links. Fprint' code resembles the entries in each fingerprint database. F print -name2 defines the accession number at the PRINTS database. Fprint-name3 is the given name for this particular fingerprint at PRINTS; Fprint-type is the type of the composition of the fingerprint (the number of motifs in the fingerprint): Fprintlinks contains 8 different databases of fingerprint. Table (2), illustrates an example of *Trifolumalexandrinum* of Attributes of FINGERPINT table, such as Inte-rpro-ID is IPR000482 its Fprint-code is BRECEPTR Fprint -name is PR00651, Fprint-type is compound (8) and Fprint-like is PR00251.

Table 1. Example of Attributes of INTERPRO-GO table

ROW	INTERPROID	GO ID	GO CATEGORY	GO DESCRIPTION
1	IPR000001			
2	IPR000002			
3	IPR000003	0003677	Molecular function	DNA binding
4	IPR000005	0003700	Molecular function	Transcription factor activity
5	IPR000006	0046872	Molecular function	Metal ion binding
6	IPR000007			
7	IPR000008			
8	IPR000009	0000159	Cellular component	Protein phosphatase type 2A complex
9	IPR000010	0004869	Molecular function	Cysteine protease inhibitor activity
10	IPR000011	0006464	Biological process	Protein modification
11	IPR000012	0005554	Molecular function	Molecular function unknown
12	IPR000013	0005576	Cellular component	Extracellular region
13	IPR000014	0004871	Molecular function	Single transducer activity
14	IPR000015	0005215	Molecular function	Transporter activity
15	IPR000018	0007186	Biological process	G- protein coupled receptor protein signalling pathway
16	IPR000020	0005576	Cellular component	Extracellular region
17	IPR000021	0016021	Cellular component	Membrane
18	IPR000022	0016874	Molecular function	Ligase activity
19	IPR000023	0003872	Molecular function	6- photosphofructokinase activity

Table 2. Example of Attributes of Fingerprint table

ROW	Finger pint code	Finger print	Finger print name	Finger print type	interpro	Print links
1	11SGLOBULIN	PRO0439	11-S seed storage protein family signature	COMPOUND(7)	IPR000459	
2	1433ZETA	PRO0305	14-3-3 protein zeta signature	COMPOUND(6)	IPR000308	
3	2CENDOPTASE	PRO0916	2C endopeptidase (C24) cysteine protease family signature	COMPOUND(4)	IPROOO317	PR0017 SRSVCYSP TASE
4	2FENDOPTASE	PRO0159	2Fe-4S ferredoxin signature	COMPOUND(2)	IPR000564	
5	2SGLOBULIN	PRO0551	2-Sglobulin family signature	COMPOUND(9)	IPR000677	PRO0439 11SGLOBULIN
6	3FE4SFRD OXINN	PRO0352	3Fe-4S ferredoxin signature	COMPOUND(3)	IPR001080	
7	4DISULPH CORE	PR00003	4-disulphide core signature	COMPOUND(4)	IPR002221	
8	4FE4SFRD OXIN	PR00353	4fe-4s ferredoxin signature	COMPOUND(2)	IPR001450	
9	5HTIBRECEPTR	PR00512	5-hydroxytryptamine 1A receptor signature	COMPOUND(7)	IPR000610	PR00251 BACTRLOP SIN
10	5HTIDRECEPTR	PR00513	5-hydroxytryptamine 1D receptor signature	COMPOUND(5)	IPR002147	PR00251 BACTRLOP SIN
11	5HTIFRECEPTR	PR00514	5-hydroxytryptamine 1D receptor signature	COMPOUND(5)	IPR000505	PR00251 BACTRLOP SIN
12	5HT2ARE CEPTR	PR00515	5-hydroxytryptamine 1F receptor signature	COMPOUND(7)	IPR000450	PR00251 BACTRLOP SIN
13	5HI2BRECEPIR	PR00516	5-hydroxytryptamine 2A receptor signature	COMPOUND(7)	IPR000455	PR00251 BACTRLOP SIN
14	5HI2BRECEPIR	PR00651	5-hydroxytryptamine 2B receptor signature	COMPOUND(8)	IPR000482	PR00251 BACTRLOP SIN
15	5HT2RECEPTR	PR00517	5-hydroxytryptamine 2C receptor signature	COMPOUND(8)	IPR000377	PR00251 BACTRLOP SIN
16	5HT4RECEPTR	PR01059	COMPOUND(3)	COMPOUND(11)	IPR001520	PR00251 BACTRLOP SIN
17	5HT5ARE CEPTR	PR00518	5-hydroxytryptamine 5A receptor signature	COMPOUND(5)	IPR001397	PR00251 BACTRLOP SIN
18	5HT5BRE CEPTR	PR00519	5-hydroxytryptamine 5B receptor signature	COMPOUND(6)	IPR000431	PR00251 BACTRLOP SIN

Table 3. Part of the Motif Table

Motif seq3	Match protein	Finger print code	ROW
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	GLU2	1ISGLOBULLN	1
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	GU11	1ISGLOBULLN	2
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	GU12	1ISGLOBULLN	3
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	Q40685	1ISGLOBULLN	4
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	Q38780	1ISGLOBULLN	5
RKNIENPQHADTYNPRAGNSKNFPIILNVIQLSAERGYL YNALLAPQWNVNAHSWYATRQ	Q40347	1ISGLOBULLN	6
RVNIENPSRADSYNPRAGNSQKFBILNLIQMSATR VNL YNALLSPFWNNAHSL WMINMQG	GLUB	1ISGLOBULLN	7
RQNDNPNRADTYNPRAGNSQNFPIILNLVQMSAVK VNL YNALLSPFWNINAHSI VYITQG	Q40689	1ISGLOBULLN	8
RVNIENPSRADSYNPRAGNSQKFBILNLIQMDATR VNL YNALLSPFWNNAHSL WMIPNHQG	GLUC	1ISGLOBULLN	9
RVNIENPSRADSYNPRAGNSQKFBILNLIQMSATR VNL YNALLSPFWNNAHSL VYMIINAQG	GLU4	1ISGLOBULLN	10
RLNIENPSHADTYNPRAGNSQKFBILNLVQLSATRVNL YNALLSPFWNNAHSL VYIVQG	GLU5	1ISGLOBULLN	11
RKNIENPQHADTYNPRAGNSKNFPIILNLVQMSATR VNL YNALLSPFWNINAHSI VYIMQG	Q38779	1ISGLOBULLN	12
RQNGNPKRADTHNPRAGHQNFPIILNLVQMSATR VNL YNALLSPFWNINAHSI WYMIQG	Q49258	1ISGLOBULLN	13
HENIARPSRADFYNPKAGNSL TLPALRFQLSAQYWL YNGIYSPHWNLNANSVIWTRG	GLC4	1ISGLOBULLN	14
HENIARPSRADFYNPKAGNSL TLPALRFQLSAQYWL YNGIYSPHWNLNANSVIWTRG	Q43452	1ISGLOBULLN	15
HENIARPSRADFYNPKAGNSL TLPALRFQLSAQYWL YNGIYSPHWNLNANSVIWTRG	Q39921	1ISGLOBULLN	16
RQNDNPNLADTYNPRAGNGQKFBILNLVQKINVQMSAVK VNL YNALLSPFWNIVWIQG	GLU3	1ISGLOBULLN	17
RQNDNPNLADTYNPRAGNGQKFBILNLVQKINVQMSAVK VNL YNALLSPFWNIVWIQG	Q40688	1ISGLOBULLN	18

D-Interpro-examples table of *Trifoliumalexandrinum*:

This table collects a representative list of the kind of proteins matching the entry. The list shows the diversity of the matches in terms of function and/or taxonomic range for each InterPro entry. The proteins are from either SWISSPROT or TREMBL. INTERPRO-EXAMPLES, table contains four attributes of *Trifoliumalexandrinum*: Interpro-ID, Match-pid, match-pd, -and Match-pname; Interpro-ID, is the Interpro entry ID, where Match-pdb contains IDS for matched proteins. Match-pids can be used to protein sequences from protein table. Match-pdb describes two protein databases that contain the matching protein entries. Finally, match-pname is the names of the matching protein entries.

E-Quick Go table:

In table (1) consists of four attributes: Go-id, Go – Isarelation, and go – Partofrelation. Go – id is the Goentry id, where Go – Isarelation shows if there exist a parent – child relationship between this go – id and other Go – ID – finally, Go – Partofrelation shows whether this go – id is Go; a part of other Go-IDS or not for example if Go – ID is 0004872 the Go– isarelation is Go :000481 and Go– Partofrelation would have a null value.

F-Motif Table:

This table contains three attributes print – code, match – proteins, and Motifseq3, where it includes the matched proteins for a specific fingerprint code. For example, in Table (3), Finger – Print code 115Globulin has a number of matching proteins such as Gluz – Trifolium and Gul2 – Trifolium Motifs sequences are shown in the same table too as well.

REFERENCES

- Abd El Salam, A.Z., S.A. Ibrahim, H.Z. Hassan, K.A. Abdelaatyand M.B. Abdelmordity 1994. Study of setting A plan for the Establishment of National Egyptian Gene Bank (NEGB). Genetic Engineering Center-Fac-Agric. Ainshams Univ. 85PP.
- Apweiler, R.,T.K.Attwood, A. Bairoch, A. Bateman, E. Birney, M. Diswas, L. Cerutti; F. Corpet, M.D.R. Croning,R.Durbin, L. Falquet, W. Fleischmann, J. Gouzy, H. Hermajakob, N. Hulo, I. Jonassen, D. Kahn, A. Kanapin, Y. Karavidopoulou, R. Lopez, B. Marx, N.J. Mulder, T.M. Oinn, M. Pagni, F – Servant, C.J.A. Sigrist and E.M. Zdobnov 2000. "Interpro-An Integrated Documentation Resource for Protein Families, Domains, and Functional Sites, Bioinformatics, 16 : 1145 – 1150.
- Ashburner, M., C. A. Ball, J.A. Blake, D. Botstein, H. Butler, J. M. Cherry, A.P.D.K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel – Tarver, A. Kasaraskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G.Sherlock, 2000. "Gene ontology:Tool for unification of Biology" Nature Genetics, Vol., 25, PP.25-29.
- Att wood. T. K., M. J. Blythe, D. Flower, R. Gaulton A. Mabey. J. E. Maudling N., L. McGregor, Mitchell Al, Moulton G, Paine K, and P. Scordis, 2002. Prints and Prints – S shed Light on Protein ancestry, Nucleic acids Research, 30 : 239 – 2451.
- Bateman, A., E. Birney, R. Durbin, S.R. Eddy, K.L. Howe and E.L.L. Sonnhammer 2000. "The Pfam protein families database, Nucleic Acids, Res., Vol., 28, PP.263-266.
- Bradford, M.M. 1976: A rapid sensitive method for the quantification of microgram quantities of protein utilisingthe principle of protein – Dye Binding. Analit.Biochem. 72, 248 – 254.
- Corpet F., F. Servant, J. Gouzy and D. Kahn 2000. "Prodom and Pro-Dom-CG: Tools for Protein Domain Analysis and Whole Genome comparisons, Nucleic Acids Rearch Vol. 28: PP.267-269.
- Hawkes, J.G. 1983. The Diversity of Crop Plants – Harvard Univ. Combridge Massachusehs184P.
- Hoffman, K., P. Bucher, L-Falquet and A. Bairoch, 1999."The Prosite database, Its Status in 1999", Nucleic Acids Research, 27, 215-219.
- Khaled Y.Abdel-Halim 2014. Biomarkers of Freshwater Algae Lemna minor as a Model for Urban pollution with pesticides and heavy Metales. Sci.Exc.Jour.95:105.
- Laemmli, U.K. 1970.Cleavage of Structural Proteins during the assembly of the head of bacteriophage T4. Nature, 227, (S259), 680-685.
- Shimaa A. Badawy 2013. Evaluation of WARM model for simulating some Rice varieties in Northern Delta.Alexandria Sci.Exc.Jour.439:450.

Databases

: (Databases)

- 1- Interpro Databases.
- 2- Prints Databases.
- 3- Quick Go Databases.

Genetic Erosion

(Oracle or SQL

XML Flat Files

Server Software 10MB)