NAIST-IS-DD1361202

Doctoral Dissertation

Systematization of Genome and Transcriptome Data based on BLSOM

Yu Bai

August 16, 2016

Department of Applied Informatics
Graduate School of Information Science
Nara Institute of Science and Technology
Japan

A Doctoral Dissertation
submitted to the Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of Science.

Yu Bai

Thesis Committee:

Professor Shigehiko Kanaya	(Supervisor)
Professor Shoji Kasahara	(Co-supervisor)
Associate Professor MD. Altafl-Ul-Amin	(Co-supervisor)
Assistant Professor Naoaki Ono	(Co-supervisor)

Systematization of Genome and Transcriptome Data based on BLSOM*

Yu Bai

Abstract

Kohonen's Self-Organizing Map (SOM), is a powerful tool for clustering and visualizing high-dimensional complex data on a two-dimensional map. An alternative version of SOM is known as batch learning SOM (BLSOM) which can produce results unaffected by the order of the input data. In this thesis we applied BLSOM for genome sequence analyses and RNA-seq data analyses. We characterized vertebrate genomes using BLSOM. We first analyzed pentanucleotide compositions in 100 kb long segments corresponding to a wide range of vertebrate genomes and then as a separate study the compositions in the human and mouse genomes in order to investigate a method for detecting differences between the closely related genomes. BLSOM successfully recognized the species specific key combination of oligonucleotide frequencies in each genome, which is called a "genome signature," and the specific regions specifically enriched by transcription-factor binding sequences. We further analyzed RNA-seq data of wild type and bleached type of *Euglena gracilis* using BLSOM. Our analyses characterized the regulation changes for transcription factors under aerobic and anaerobic conditions and linked transcription factors with metabolic pathways. Transcription factors and other genes placed at the same area on the map follows the same patterns of regulation changes. Collectively, the results of our analyses can provide a valuable resource for *Euglena gracilis* research and specially novel insights into *Euglena gracilis* responses to anaerobic stress and offer candidate genes or markers that can be used to guide future efforts attempting to develop anaerobic tolerant *Euglena gracilis* cultivars.

Keywords:

BLSOM, Genome, Genome signatures, Transcriptome, Transcription factors

---

*Doctoral Dissertation, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD1361202, August 16, 2016.

Acknowledgement

It is a very important stage of my life to study in Nara Institute of Science as a doctoral student. In the period of my Ph. D study here, I met a lot of people, and learned a lot of knowledge.

I want to thank to my supervisor, Professor Shigehiko Kanaya, who gave me the opportunity to study in the Computational Systems Biology Laboratory. I greatly appreciate his assistance and support during my studies. Professor Kanaya not only gives me the knowledge and the skill during the doctoral study, but also shows me the example of a real scientist.

I want to thank to my supervisor, Professor Toshimichi Ikemura, who led me into bioinformatics. I greatly appreciate his assistance and support during my studies.

I would like to thank my thesis committee, Professor Shigehiko Kanaya, Professor Shoji Kasahara, Associate Professor MD. Altafl-Ul-Amin, Associate Professor Tadao Sugihura, Associate Professor Naoaki Ono. Thank for their valuable comments and supervision of my work. Their valuable advice, guidance, critics greatly helped to improve this dissertation.

I would like to thank all the other staff in Computational Systems Biology Laboratory: Associate Professor Tetsuo Sato, Associate Professor Alex Ming Huang, Aki H. Morita, Minako Ohashi. It is very kind of you to assist me in daily work and life. I want to give my appreciation to my friend Katsuragi Tetsuo, Donghan Li, Nelson Kipchirchir Kibinge, Shun Ikeda, Sony Hartono Wijaya, Kang Liu and Hiroki Okuda, who helped me a lot in my daily research and life.

Last but not least, my appreciation also extends to my parents, my girlfriend and my friends. With your understanding, patience, critics, companion, support, and encouragement, I could achieve my doctoral degree.

Thank you very much

List of abbreviations

ARF	Auxin response factors
bHLH	basic/helix-loop-helix
BLSOM	batch learning Self-Organizing Map
CGS	sets of correlated genes
C3H	Cys3His zinc finger
DEGs	differentially expressed genes
E. gracilis	*Euglena gracilis*
FDR	false discovery rate
FPKM	fragments per kilobase of exon model per million mapped reads
GO	Gene Ontology
HC	hierarchical clustering
KEGG	Kyoto Encyclopedia of Genes and Genomes
mRNA	messenger RNA
NAC	NAM, ATAF, and CUC
ncRNA	nonconding RNA
NMF	non-negative matrix factorization
PCA	principal component analysis
SL	spliced leader-sequence

SOM	Self-Organizing Map
Sz	Specific Zones
TALE	three-amino-acid-loop-extension
TF	Transcription factor
TFB	transcription factor-binding
TFs	Transcription factors

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Preliminaries

With the increasing amount of available genomic sequences and RNA sequences, novel tools are needed for comprehensive analysis of species-specific sequence characteristics for a wide variety of genomes and co-regulation (or co-expression) networks analysis of species under different conditions.

Kohonen's Self-Organizing Map (SOM), is a powerful tool for clustering and visualizing high-dimensional complex data on a two-dimensional map (Kohonen et al, 1982; Kohonen et al, 1990; Kohonen et al, 1996). In this thesis study, on the basis of batch learning SOM (BLSOM), we have previously developed a modification of the conventional SOM for genome and gene sequence analyses and RNA-seq analyses.

1.2 Current progress of Bioinformatics

Since the completion of the human genome project in 2003, extraordinary progress has been made in genome sequencing technologies, which has led to a decreased cost per megabase and an increase in the number and diversity of sequenced genomes and transcriptomes (Goodwin et al, 2016). An astonishing complexity of genome architecture has been revealed, bringing these sequencing technologies to even greater advancements (Goodwin et al, 2016). Some approaches maximize the number of bases sequenced in the least amount of time, generating a wealth of data that can be used to understand increasingly complex phenotypes. Alternatively, other approaches now aim to sequence longer contiguous pieces of DNA, which are essential for resolving structurally complex regions (Goodwin et al, 2016). These and other strategies are providing researchers and clinicians a variety of tools to probe genomes in greater depth, leading to an enhanced understanding of how genome sequence variants underlie phenotype and disease.

1.2.1 Genome information

In early 2001, a rough draft of the human genome was completed by the Human Genome Project. This project, completed in 2003, sequenced the entire genome for one specific person, and by 2007 this sequence was declared "finished" (less than one error in 20,000 bases and all chromosomes assembled) (Goodwin et al, 2016). In the years since then, the genomes of many other individuals have been sequenced, partly under the auspices of the 1000 Genomes Project, which announced the sequencing of 1,092 genomes in October 2012 (Goodwin et al, 2016). Completion of this project was made possible by the development of dramatically more efficient sequencing technologies and required the commitment of significant bioinformatics resources from a large international collaboration (Nielsen et al, 2010). The continued analysis of human genomic data has profound political and social repercussions for human societies (Barnes et al, 2009).

The genomic signature refers to the characteristic frequency of oligonucleotides in a genome or sequence. It has been observed that the genomic signature of phylogenetically related genomes is similar.

Genomes are not a random sequence of bases and neither it is their oligonucleotide composition (Bikandi et al, 2004). Oligonucleotides may be present in the genome more (over-representation) or less (under-representation) often than expected from its nucleotide composition due to the evolutionary pressure within the cell internal environment. As the internal biochemical environment of phylogenetically related organisms is similar, oligonucleotide composition of phylogenetically related organisms is similar.

BLSOM constructed for oligonucleotide compositions of fragmental sequences (e.g. 100 kb) from genomes of species showed that the sequences were clustered (self-organized) primarily according to species without information regarding the species, demonstrating a powerful ability to unveil genome signatures.

1.2.2 Transcriptome information

Transcriptome is the complete set of messenger RNA (mRNA) and nonconding RNA (ncRNA) transcripts produced by a particular cell, cell type, or organism. An intriguing problem in molecular biology is how the identical genetic make-up of cells can give rise to different cell types, each of which plays a defined role in the functioning of multicellular organism. This phenotypic diversity has been linked to the fact that different cell types within the organism activate (or express) different sets of genes (transcriptomes) that lead to different cell fates and functions (Morozova et al, 2009).

BLSOM constructed for expression change of statistically significant differentially expressed genes (DEGs), demonstrating a powerful ability to unveil co-expression between Transcription factors (TFs) and other genes and predict the target genes of TFs.

Chapter 2: Systematization of Genome data based on BLSOM

2.1 Introduction

Genome sequences, both protein coding and non-coding parts of the sequences, contain a wealth of information. The G + C content (G + C%) is a fundamental characteristic of individual genomes and used for a long period as a basic phylogenetic parameter to characterize inter- and intragenomic differences. The G + C%, however, is too simple to differentiate wide varieties of genomes. Many groups have reported that the oligonucleotide composition, which is an example of high-dimensional data, varies significantly among genomes and can be used to study genome diversity (Nussinov R et al, 1984; Phillips GJ et al, 1987; Karlin S. et al, 1998; Karlin S. et al, 1998; Rocha EP et al, 1998; Gentles AJ et al, 2001; Bernardi G. et al, 2004; Vinga S. et al, 2003; Bolshoy A. et al, 2003), and the oligonucleotide compositions, including dinucleotide composition, are called the "genome signature" of each species. Various linguistic tools for analyzing DNA sequence have been developed (Vinga S. et al, 2003; Bolshoy A. et al, 2003). Unsupervised neural network algorithm, Kohonen's Self-Organizing Map (SOM), is a powerful tool for clustering and visualizing high-dimensional complex data on a two-dimensional map (Kohonen T. et al, 1982; Kohonen T. et al, 1990; Kohonen T. et al, 1996).

SOM machine learning is chosen as classifiers because it combines strong clustering, dimension reduction, multidimensional scaling and visualization capabilities which have been shown to be advantageous compared to alternative methods such as

K-means hierarchical clustering (HC), non-negative matrix factorization (NMF), and PCA-component plots when applied to molecular high-throughput data (Wirth et al., 2011). I also explain detail algorithm of HC, NMF, and PCA as Appendix hierarchical clustering (HC), non-negative matrix factorization (NMF), and principal component analysis (PCA).

The hierarchical clustering algorithm used is based closely on the average-linkage method of Sokal and Michener, which was developed for clustering correlation matrixes (Eisen, M. B et al, 1998). Detail algorithm is represented in Appendix hierarchical clustering (HC). To compare HC with SOM, SOM is more accuracy in classifying the objects to their clusters than HC (Abbas, O. A., 2008). Notably, HC cannot readily recognize the most important genes defining the branches of the clustering tree, and can be lost important patterns due to the deterministic nature of clustering or the high dimensionality of data (Tamayo, P. et al, 1999; Wang, J. et al, 1999). For comparing large genome segments and expression patterns of transcriptome from NGS, SOM is easier than HC for visualization.

Non-negative matrix factorization (NMF) is a technique for finding parts-based, linear representations of non-negative data (Hoyer et al, 2004). Detail algorithm is represented in Appendix non-negative matrix factorization (NMF). Although NMF is more stable than SOM, the convergence of NMF is slower than SOM because method is more complex than SOM. NMF does not incorporate statistical dependencies between the metagenes or metagene expression patterns, and not identify any structural relationships between them (Brunet et al, 2004; Devarajan, K. et al, 2008). So SOM clusters easier than NMF for comparing genomes and expression patterns.

Principal component analysis of a data matrix extracts the dominant patterns in the matrix in terms of a complementary set of score and loading plots (Wold S. et al, 1987). Detail algorithm is represented in Appendix principal component analysis (PCA). Although PCA has explained most of the cumulative variance of data, when the very large-dimensional sequence-based machine-part incidence matrix is available, obtaining the clustering structure from PCA projection becomes very difficult, because the visual extraction of clustering information has become very difficult from the 2D principal component projection plot. Most importantly, The SOM map is very useful to obtain visual clustering information with the help of color extraction of U-matrix and SOM projection. (Chattopadhyay et al, 2011). When we compare genomes using oligonucleotide compositions which are large-dimensional data, SOM is better than PCA.

There are many machine learning methods but BLSOM has advantage in analyzing genome sequence data and gene expression mainly in forms of visualization. Therefor previously served other studies successfully applied SOM in analyzing genome sequence and gene expression data. (ENCODE Project Consortium, 2012; Zamani, Neda et al, 2013; Guojian, C et al, 2013; Panigrahi et al, 2013)

On the basis of batch learning SOM, we have previously developed a modification of the conventional SOM for genome and gene sequence analyses, which makes the learning process and resulting map independent of the order of data input: BLSOM (Kanaya S. et al, 1998; Abe T. et al, 2002; Abe T. et al, 2003). Importantly, BLSOM is solely depending on the oligonucleotide composition, is suitable for actualizing high performance parallel-computing, allows classification of sequence fragments according to species and, therefore, can analyze big sequence data such as millions of genomic sequences simultaneously (Abe T. et al, 2006). It can be used for classification of genomic data (Iwasaki Y. et al, 2013), additionally, it has been implemented to analyze gene expression data with success (Hirai et al. 2004; Hirai et al. 2005).

Consequently, the advantages of BLSOM can be summarized as follows:

(a) It can cluster high dimensionality of data using dimension reduction; (b) it has strong visualization capabilities for obtaining clustering structural relationships between data; (c) it is independent of the order of data input.

When we constructed BLSOMs for di-, tri-, and tetranucleotide composition in 10 kb genomic sequences derived from a wide range of prokaryotic and eukaryotic genomes, the sequences were clustered (i.e., self-organized) according to species without any information regarding the species during the BLSOM calculation, and increasing the length of the oligonucleotides from di- to tetranucleotides increased the clustering power (Abe T. et al, 2003).

An apparent causative factor for the genome signature is the context-dependent DNA mutation and repair mechanisms. It should also be noted that oligonucleotides especially longer than trinucleotides often represent motif sequences responsible for sequence-specific protein binding (e.g., transcription factor binding). The occurrence of such motif oligonucleotides in the genome should differ from the level expected from the mononucleotide composition in the respective genome and may differ among genomic portions of one genome. We have recently found that DegPenta and DegHexa for the human genome can effectively detect characteristic occurrence patterns of many transcription factor-binding motifs in pericentromeric heterochromatin regions (Iwasaki Y. et al, 2013).

In the present study, in order to clarify vertebrates' genome signatures, we first analyzed pentanucleotide compositions in 100 kb genomic sequences derived from a wide range of vertebrates and then those from human and mouse genomes in order to investigate the power to detect differences between the closely related genomes.

2.2 Materials and Methods

2.2.1 BLSOM

BLSOM is an unsupervised neural network algorithm that implements a characteristic non-linear projection from the high-dimensional space of input data onto a two-dimensional array of weight vectors (Kohonen T et al, 1982; Kohonen T. et al, 1996). We have previously modified the conventional SOM for genome informatics to make the learning process and resulting map independent of the order of data input, and established a BLSOM (Kanaya S. et al, 1998; Abe T. et al, 2002; Abe T. et al, 2003). Here, we explain the BLSOM method developed by Kanaya et al (Kanaya S. et al, 1998).

In the original Kohonen's SOM, the initial vectorial data were set by random values, but in the BLSOM the initial vectors are set based on the widest scale of the sequence distribution in the oligonucleotide frequency space with the principal component analysis (PCA) (Kanaya S. et al, 1998). Weights in the first dimension (I) was arranged into lattices corresponding to a width of five times the standard deviation ($5\sigma_1$) of the first principal component: the second dimension (J) was defined by the nearest integer greater than $\sigma_2/\sigma_1 \times I$; and I was set in the present study as the average number of sequence data per neuron becomes approximately four. σ_1 and σ_2 were the standard deviations of the first and second principal components, respectively. The weight vector on the *ij*th lattice ($\mathbf{w}_{ij}$) was represented as follows:

$$\boldsymbol{w}_{ij} = x_{av} + \frac{5\sigma_1}{I}\left[b_1\left(i - \frac{I}{2}\right) + b_2\left(j - \frac{J}{2}\right)\right] \quad (2.1)$$

where $\mathbf{x}_{av}$ is the average vector for oligonucleotide frequencies of all input vectors, and $\mathbf{b}_1$ and $\mathbf{b}_2$ are eigenvectors for the first and second principal components. In Step 2, the Euclidean distances between the input vector $\mathbf{x}_k$ and all weight vectors $\mathbf{w}_{ij}$ were calculated; then $\mathbf{x}_k$ was associated with the weight vector (called $\mathbf{w}_{i'j'}$) with minimal distance. After associating all input vectors with weight vectors, updating was done according to Step 3.

In Step 3, the *ij*th weight vector was updated (Kanaya S. et al, 1998) by

$$\boldsymbol{w}_{ij}^{(new)} = \boldsymbol{w}_{ij} + \alpha(r)\left[\frac{\sum_{x_k \in S_{ij}} \boldsymbol{x}_k}{N_{ij}} - \boldsymbol{w}_{ij}\right] \tag{2.2}$$

where components of set S_{ij} are input vectors associated with $\mathbf{w}_{i'j'}$ satisfying $i-\alpha(r) \le i' \le i+\alpha(r)$ and $j-\beta(r) \le j' \le j+\beta(r)$.

The two parameters α(r) and β(r) are learning coefficients for the *r*th cycle, and N_{ij} is the number of components of S_{ij}. α(r) and β(r) are set by

$$\alpha(r) = \max\ \{0.01,\ \alpha(1)(1 - r/T)\} \tag{2.3}$$

$$\beta(r) = \max\ \{1,\ \beta(1) - r\} \tag{2.4}$$

where, α(1) and β(1) re the initial values for the T-cycle of the learning process. In the present study, we selected 60 - 100 for T, 0.6 for α(1), and 40 ~ 80 for β(1) depending on the map size (approximately a fourth of I). The learning process is monitored by the total distance between x_k and the nearest weight vector $w_{i'j'}$, represented as

$$Q(r) = \sum_{k=1}^{N}\left\{\left\|\boldsymbol{x}_k - \boldsymbol{w}_{ij}\right\|^2\right\} \tag{2.5}$$

where N is the total number of sequences analyzed.

BLSOM learning for oligonucleotide composition was conducted as described previously (Abe T. et al, 2003). BLSOM program was obtained from Niigata Univ. (mailto:takaabe@ie.niigata-u.ac.jp) or UNTROD, Inc. (y_wada@nagahama-i-bio.ac.jp).

2.2.2 U-matrix

Distances of weight vectors between neighboring lattice points on BLSOM can be visualized as black levels with a U-matrix method (Ultsch A. et al, 1993), and this provides information regarding similarity of oligonucleotide composition in local areas on BLSOM; the areas composed of lattice points with similar or distinct oligonucleotide composition can be recognized as low or high black level, respectively.

2.2.3 Genome Sequences

Genome DNA sequences were obtained from UCSC ftp site (http://www.ncbi.nlm.nih.gov/genomes/). When the number of undetermined nucleotides (Ns) in a fragment sequence (e.g. 100 kb) exceeded 20% of the sequence, the sequence was omitted from the analysis. In the case where the number of Ns was less than 20%, the oligonucleotide frequencies were normalized to the length without Ns and included in the analysis.

2.3 Results

2.3.1 Characteristics of BLSOM Clustering

In the era of extensive genome sequencing, it is important to develop novel bioinformatics tools to support an efficient knowledge discovery from massive amounts of genomic sequences. Analyses on the species-specific oligonucleotide composition “genome signature” (e.g., penta- and hexanucleotide compositions) may provide in silico information concerning important signal sequences such as transcription factor binding sequences (Iwasaki Y. et al, 2013). To show the clustering ability of BLSOM for vertebrate genome sequences and to explain the basal features of BLSOM clustering patterns, we first analyze pentanucleotide compositions in 100 kb sequence fragments derived from 10 vertebrate genomes.

In DNA databases, only one strand of each pair of complementary sequences is registered. Previous analysis of prokaryotic species that was done by Abe et al. (Abe T. et al, 2003) revealed that sequences (e.g., 10 kb sequences) from a single prokaryotic genome were often split vertically into two territories according to the transcriptional direction of the genes present in the fragment. However, to study general characteristics of genomic sequences such as the genome signature, differences in the oligonucleotide composition between two complementary strands are not necessarily important. Therefore, we construct a BLSOM in which the frequencies of a pair of complementary pentanucleotides (e.g., AAAAC and GTTTT) in each fragment are summed (Iwasaki Y. et al, 2013). The BLSOM for this degenerate set of a pair of complementary pentanucleotides is designated as DegPenta.

On the BLSOM, lattice points containing sequences from a single species are indicated in a color specifying the species; those containing sequences from multiple species are indicated in black. Because most lattice points are colored, a high separation power is apparent for DegPenta (Figure 2.1(a)), with no information concerning species during the BLSOM calculation. We next explain the basal characteristics of BLSOM separation observed for the vertebrate sequences. G + C% has long been used as a fundamental value that characterizes both inter- and intragenomic differences. For example, on a warm-blooded vertebrate genome, there exists a long-range segmental G + C% distribution "isochores," which have been connected with chromosomal bands (Bernardi G. et al, 2004; Bernardi G. et al, 1985; Ikemura T. et al, 1985; Ikemura T. et al, 1988; Ikemura T. et al, 1991). Figure 2.1(b) presents the G + C% that is calculated from pentanucleotide composition at each lattice point in the DegPenta. Sequences with high and low G + C% (wine red or green in Figure 2.1(b)) are located on the left and right side of the map, respectively, showing that the G + C% level is reflected primarily in the horizontal direction. The territory of each species is often split into several subterritories, which should relate at least in part to isochore structures because the G + C% level differs between subterritories of a single species, for example, chicken (🟪) and human (🟫) territories.

BLSOMs can visualize diagnostic oligonucleotides responsible for species-specific clustering (self-organization). We first calculate the pentanucleotide frequencies expected from the mononucleotide composition that is obtained from the vectorial data (i.e., pentanucleotide composition) at each lattice point and indicate the observed/expected ratio as follows: red (overrepresented), blue (underrepresented), and white (moderately represented) (Figure 2.1(c)). This observed/expected ratio is useful in unveiling genome signatures, since it allows us to examine the oligonucleotide composition at each lattice point, independently of a simple effect derived from its mononucleotide composition (Iwasaki Y. et al, 2013). For various pentanucleotides, transitions between red and blue often coincide exactly with species-specific territory borders. AACAT + ATGTT, ACAAC + GTTGT, ATTTA + TAAAT, and CAGCG + CGCTG are overrepresented in fishes (Fugu, Medaka, Stickleback, Tetraodon, and Zebrafish) but not in almost all tetrapods (Human, Lizard, Mouse, Chicken, and Xenopus). ACCCT + AGGGT and CCAAG + CTTGG are overrepresented in tetrapods but not in fishes. AACCC + GGGTT is underrepresented in chicken and a part of fish (Fugu, Stickleback, and Tetraodon). GAAGA + TCTTC is underrepresented in Xenopus and Zebrafish. These findings show that BLSOM can recognize the species-specific oligonucleotide composition and identify the combinatorial diagnostic oligonucleotides responsible for species-specific clustering; that is, a combination of not a few but many pentanucleotides contributes to the accurate clustering (self-organization) of genomic sequences according to species.

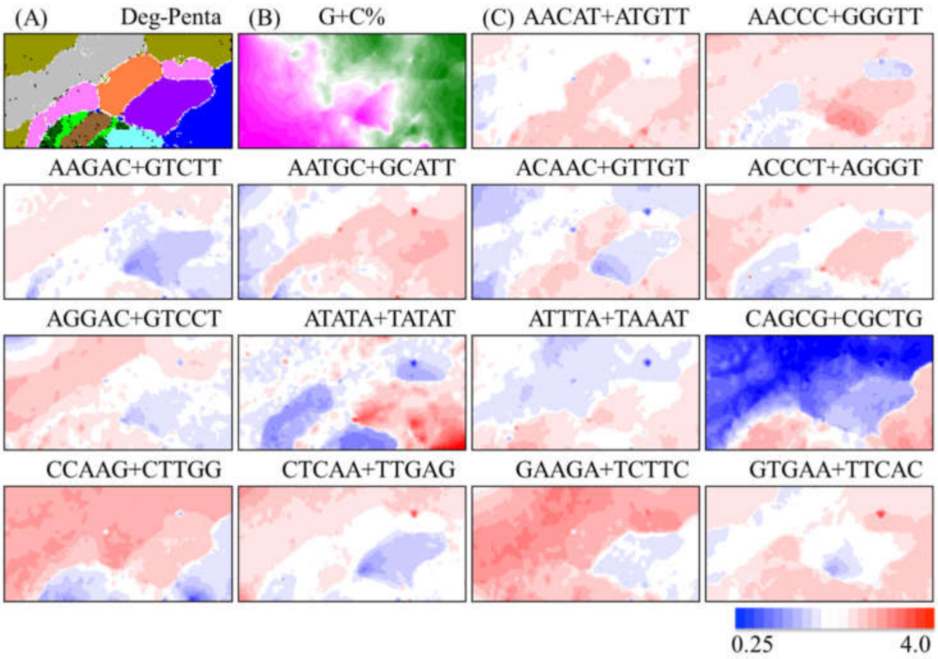


Fig. 2.1. BLSOMs for 100-kb sequences derived from 10 vertebrata genomes.

(A) DegPenta. Lattice points containing sequences from multiple species are indicated in black and those containing sequences from a single species are indicated in color as follows: Chicken (■), Fugu (■), Human (■), Lizard (■), Medaka (■), Mouse (■), Sticleback (■), Tetraodon (■), Xenopus (■), Zebrafish (■). (B) G+C%. For each lattice point in the DegPenta, G+C% was calculated and divided into 21 categories with an equal number of lattices. The lattice points belonging to the categories of the highest, middle and lowest G+C% are shown in wine red, white and green, respectively. (C) Diagnostic pentanucleotides responsible for species-specific clustering. Occurrence of each pentanucleotide for each lattice point was calculated and normalized with occurrence expected from the mononucleotide composition for the respective lattice point (Abe T. et al, 2006; Iwasaki Y. et al, 2013). This observed/expected ratio is indicated in color presented under the panel. This ratio has been shown to be useful in unveiling genome signatures because the oligonucleotide composition can be analyzed independently of a simplex effect reflecting the mononucleotide composition of genomic sequences (Abe T. et al, 2006; Iwasaki Y. et al, 2013; Abe T. et al, 2009).

2.3.2 BLSOMs for Human and Mouse Genomes

We have next constructed DegPenta with 100 kb sequences derived from the human and mouse genomes (Figure 2.2). This enables us to examine a BLSOM power for separating the species with a relatively close phylogenetic relationship and to clarify difference in the genome signatures of the closely related species. Lattice points that contain sequences derived from human and mouse are indicated in red and blue, respectively, and those that include sequences from both human and mouse are indicated in black. With no information regarding species during the BLSOM calculation, the species-specific clustering (self-organization) of the 100 kb sequences is clear.

In Figure 2.2(b), the observed/expected ratios of individual pentanucleotides calculated as explained in Figure 2.1(c) are illustrated in red (overrepresented), blue (underrepresented), and white (moderately represented). Transitions between red (overrepresentation) and blue (underrepresentation) for various pentanucleotides often coincide exactly with species territory borders, showing that BLSOM recognizes the speciesspecific combination of oligonucleotide frequencies that is the representative signature of one genome and enables us to identify the frequency patterns that are characteristics of individual genomes.

Seven examples of the pentanucleotides diagnostic for the species territory formation are presented (Figure 2.2(b)). AAATT + AATTT, ATCAC + GTGAT, and TTCAA + TTGAA are preferred in the human genome but not in the mouse genome. On the other hand, AACAC + GTGTT, ACAAC + GTTGT, ACAAG + CTTGT, and ACACT + AGTGT are preferred in the mouse genome but not in the human genome. It should be stressed that a complex combination of many pentanucleotides contributes to the species specific clustering (i.e., self-organization) of sequence fragments. Some of these diagnostic pentanucleotides, if not all, may have important biological significances, which should be related to functions.

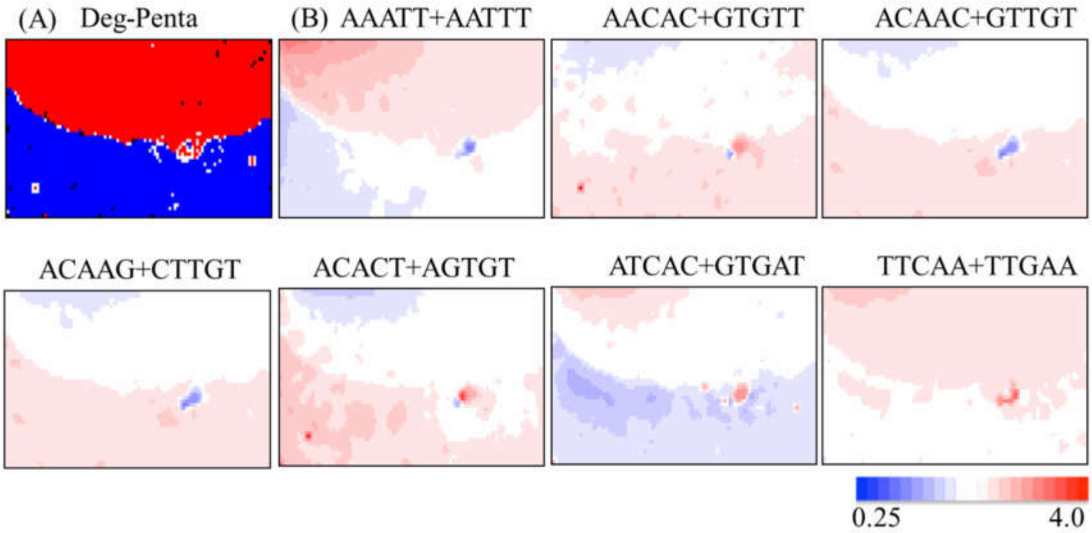


Fig. 2.2. BLSOMs for 100-kb sequences derived from the human and mouse genomes.

(A) DegPenta. Lattice points containing sequences from human and mouse are indicated in black and those containing sequences from a single species are indicated in color as follows: human (🟥) and mouse (🟦). (B) Occurrence level of each pair of complimentary pentanucleotides in the DegPenta. Level of a complimentary pentanucleotide pair for each lattice point is calculated and normalized with the level expected from the mononucleotide composition for the lattice point. The observed/expected ratio is indicated in colors presented at the bottom of the figure. Seven examples of the pentanucleotides diagnostic for species-specific separations are presented.

2.3.3 Characteristics of Sequences Belonging to Specific Zones

While most 100 kb sequences are classified primarily into species-specific territories, there are a few minor human zones (red) that are located within the mouse territory (blue) and are surrounded with white lattice points. In addition, there is a nub-type human zone that is located in the border region between human and mouse territories and also is surrounded by white lattice points. In Figure 2.2(a), lattice points with no genomic sequence assigned after the BLSOM calculation are left white. It should be mentioned that Abe et al. (Abe T. et al, 2003) and Iwasaki et al (Iwasaki Y. et al, 2013). have previously shown that lattice points containing genomic sequences whose oligonucleotide composition is very distinct from other genomic sequences tend to be surrounded by lattice points containing no genomic sequence.

Similarity in oligonucleotide composition between neighboring lattice points in BLSOM (and thus between sequences belonging to neighboring lattice points) can be visualized using a U-matrix (Ultsch A. et al, 1993) with a level of blackness (Figure 2.3(a)), as described in Section 2.2.2.

On the U-matrix, borders between human and mouse territories are visualized as black lines, which represent distinct pentanucleotide compositions between human and mouse sequences. Furthermore, there are small dark black zones and gray zones surrounded by a black circle, which should contain sequences with peculiar oligonucleotide composition distinct from the compositions from other genomic sequences; the respective zones composed of human sequences are numbered as SzH1 and Sz-H2 and that of mouse sequences is specified as Sz-M (Figure 2.3(b)).

Importantly, these numbered zones primarily correspond to zones surrounded by white lattice points in Figure 2.2(a), confirming that the sequences in these specific zones have peculiar oligonucleotide compositions very distinct from a major portion of the respective genome. Actually, occurrence levels of individual pentanucleotides in the specific zones are clearly different from those in the major portion of the respective genome (Figure 2.2(b)). AATCT + AGATT and AGATA + TATCT are preferred in Sz-H2 but not in Sz-H1 and Sz-M. ATTGA + TCAAT is preferred in Sz-H1 and Sz-H2 but not in Sz-M. ATTGG + CCAAT is preferred in Sz-H1 but not in Sz-H2 and Sz-M. The pentanucleotides listed in Figure 2.2(b) correspond to human transcription factor-binding (TFB) motifs and the reason why these motif pentanucleotides are chosen is explained below.

The oligonucleotides such as penta- and hexanucleotides often provide the binding sites of proteins such as transcription factors. When we consider the oligonucleotides that can function as important signal sequences such as TFB motifs, their occurrence levels in genomic sequences should be biased significantly from the levels expected from random sequences. Therefore, the overrepresentation of a certain oligonucleotide only in a restricted portion of the BLSOM (and thus of the genomic sequences) is thought to provide useful information for understating the biological significance of the respective sequence, especially when a biological function of the oligonucleotide of interest is known.

In our previous study (Iwasaki Y. et al, 2013), we have shown that oligonucleotide BLSOM such as DegPenta can be used for studying sequences derived even from one genome. In that study, addition of computer-generated random sequences to real human sequences can successfully separate the specific sequences with distinct oligonucleotide composition from a major portion of the human genome; that is, these specific sequences are displaced well from the major portion of human sequences and surrounded by the random sequences.

Interestingly, the specific human sequences thus found are derived mainly from pericentromeric regions and enriched by TFB motif sequences (Iwasaki Y. et al, 2013). Instead of the human plus random sequences used in the previous study, human plus mouse sequences are analyzed in the present study, and the addition of the closely related species appears to effectively assign the 100 kb sequences with peculiar oligonucleotide compositions very distinct from those in the major portion of the respective genome (Figure 2(a)).

In order to clarify the characteristics of the specific sequences found in this study and to compare with those found previously, we have analyzed the occurrence of the pentanucleotides corresponding to human TFB motifs analyzed in the previous paper. All of the TFB pentanucleotides are overrepresented (red) in a certain specific zone but underrepresented (blue) in almost all other human 100 kb sequences, confirming the previous result. When we examine their occurrences in the mouse territory, AATCT + AGATT, ATTGA + TCAAT, and TATCA + TGATA are underrepresented in a major portion of the mouse genome.

However, ATTGG + CCAAT and AGATA + TATCT are underrepresented only in a half portion of the mouse genome, indicating that the biological function of these two pentanucleotides may differ from that for human. Comparative analyses of the closely related species can provide this type of information concerning a possible evolutionary change in functional signal sequences such as TFB motif sequences, but the addition of the computergenerated ransom sequences cannot provide the information concerning the molecular evolutionary change.

The reason why the specific zones of mouse on DegPenta are less evident than human will be discussed below.

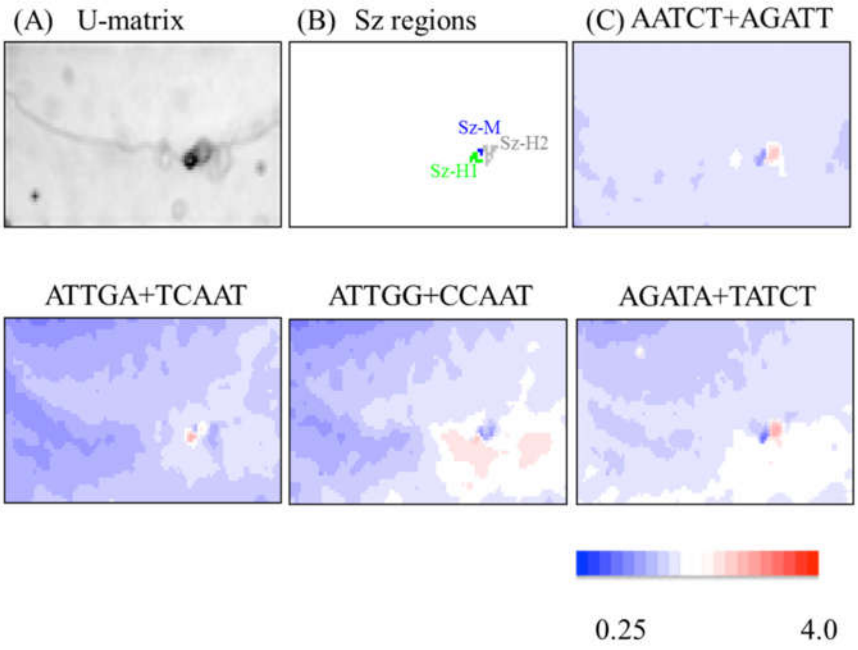


Fig. 2.3. Pentanucleotides specifically enriched in specific region.

(A) U-matrix for the BLSOM listed in Fig. 2.2A. (B) Sz regions. Sz-H1 and Sz-H2 regions of human sequences are indicated in green and gray letters, respectively. A very tiny Sz-M region of mouse sequences is indicated in blue. (C) Pentanucleotides specifically enriched in Sz. The observed/expected ratio for each pair of complimentary pentanucleotides is calculated as described in Fig. 1C and indicated in colors presented under the panel.

2.4 Discussion

2.4.1 Repeat and Unique Sequences

Vertebrate genomes are composed of repeat and nonrepeat, unique sequences, which have distinct biological functions. Since repeat sequences usually have peculiar oligonucleotide composition, there exists a possibility that the specific zones' sequences with peculiar oligonucleotide compositions distinct from a major portion of the genome are repeat sequences, and this possibility is examined as follows. In the UCSC database, repeat sequences identified by Repeat Masker and Tandem Repeats Finder are specified in lower-case letters for distinguishing from unique sequences specified in upper-case letters. We first concatenated unique or repeat sequences separately, divided these concatenated sequences into 100 kb sequences, and counted pentanucleotide composition in each 100 kb sequence.

Clear separation between species and between repeat and unique sequences is observed on DegPenta (Figure 2.4(a)). Interestingly, human repeat sequences (pink) forms one satellite-type minor territory located at the lowest part of the map and the mapping of the specific zones' sequences marked in Figure 2.3(b) shows that these specific sequences are mainly located in the minor territory of human repeat sequences (Figure 2.4(b)). Therefore, the specific sequences actually belong to the repeat category. However, it has been separately shown that these specific sequences are different from the ubiquitously distributed human repeat sequences such as Alu and LI. As another separate analysis, we have found that these sequences are also different from alphoid sequences, which are a major component of human centromeric regions. Core parts of human centromeric regions mainly composed of alphoid sequences have not been included in the standard human genome sequences currently available, because of the difficulty to get contiguous sequences. The minor human territory of interest is colored in black on U-matrix (Figure 2.4(c)) and appears to be split into two parts: a very dark small part and its adjacent gray part.

In Figure 2.4(d), we list six examples of pentanucleotides (including a TFB motif) diagnostic for separation between species and/or between repeat and unique sequences. Interestingly, all pentanucleotides show a very high (dark red) or very low (dark blue) occurrence in the minor human repeat territory, again showing its very peculiar oligonucleotide composition. The specific characteristics in this minor repeat territory are further confirmed when we examine the dinucleotide CG-containing pentanucleotides (Figure 2.5). All these CG-containing pentanucleotides listed are specifically overrepresented in the very dark, small region visualized in U-matrix (Figure 4(c)) but evidently underrepresented in all other regions. When we examine the occurrence of all CGcontaining pentanucleotides in detail, the CGA-containing pentanucleotides are particularly enriched in the very dark, small region in U-matrix, and almost all examples listed in Figure 2.5 correspond to the CGA-containing pentanucleotides.

The evident underrepresentation of the CG dinucleotide (i.e., CG suppression) is well known in vertebrate genomes and the CG suppression is believed to relate to methylation at CG dinucleotide, which is a well-characterized epigenetic marker. Concerning the CG occurrence level, CpG islands, in which the CG occurrence is clearly higher than in other genomic regions, are well known to have important roles in transcriptional regulation. The sizes of the CpG islands are known to be a few or several hundred bp and, therefore, are clearly different from the size of specific sequences found in the present study (a 100 kb level). Furthermore, CpG islands belong primarily to the unique sequence regions. Therefore, the 100 kb level sequences enriched with the CG-containing pentanucleotides are not the CpG island sequences. As noted above, the CG dinucleotide is a target of methylation and this C methylation is known to have important roles in epigenetic systems. The 100 kb level specific sequences may have important roles that are different from but possibly related to the function of CpG islands. The finding that the CGAcontaining pentanucleotides are more preferred in the specific sequences than other CG-containing pentanucleotides may give information for clarifying biological functions of the 100 kb level sequences of interest.

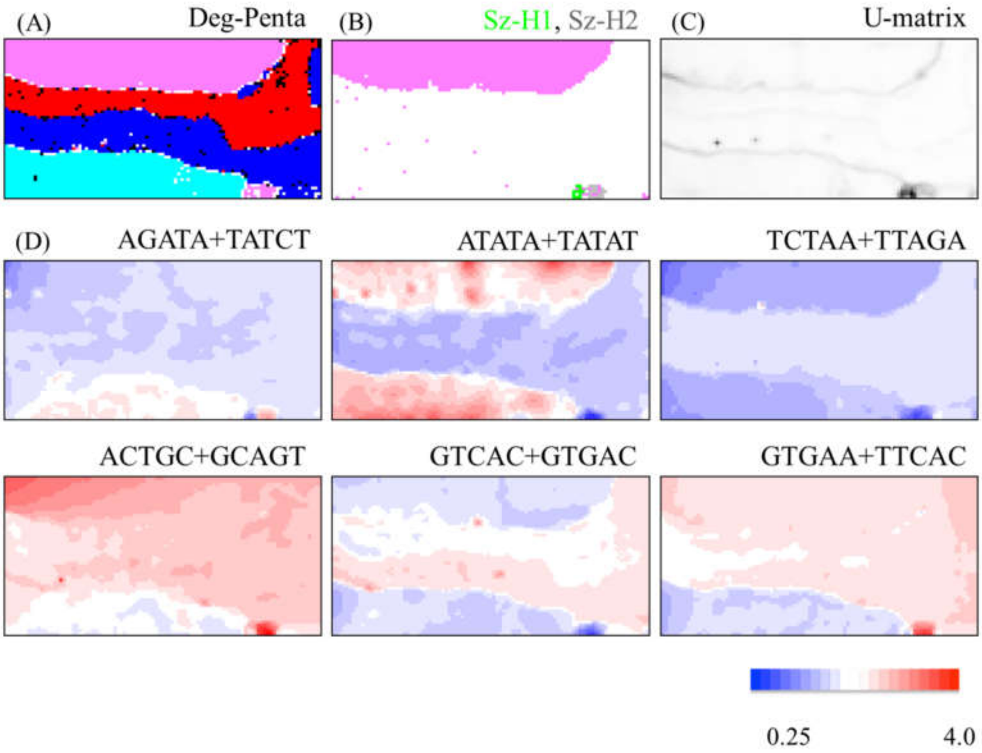


Fig. 2.4. BLSOMs for 100-kb repeat and unique sequences derived from the human and mouse genomes.

(A) DegPenta. Lattice points containing sequences from more than one category are indicated in black and those containing sequences from a single category are indicated in color as follows: human repeat (■), human unique (■), mouse repeat (■) and mouse unique (■). (B) Human repeat sequences. Human Sz-H1 and Sz-H2 sequences defined in Fig.2.3B are mapped and indicated in green and gray, respectively. (C) U-matrix for the DegPenta listed in A. (D) Diagnostic pentanucleotides responsible for species-specific clustering. The observed/expected ratio for each pair of complimentary pentanucleotides is calculated as described in Fig. 2.1C and indicated in color presented under the panel.

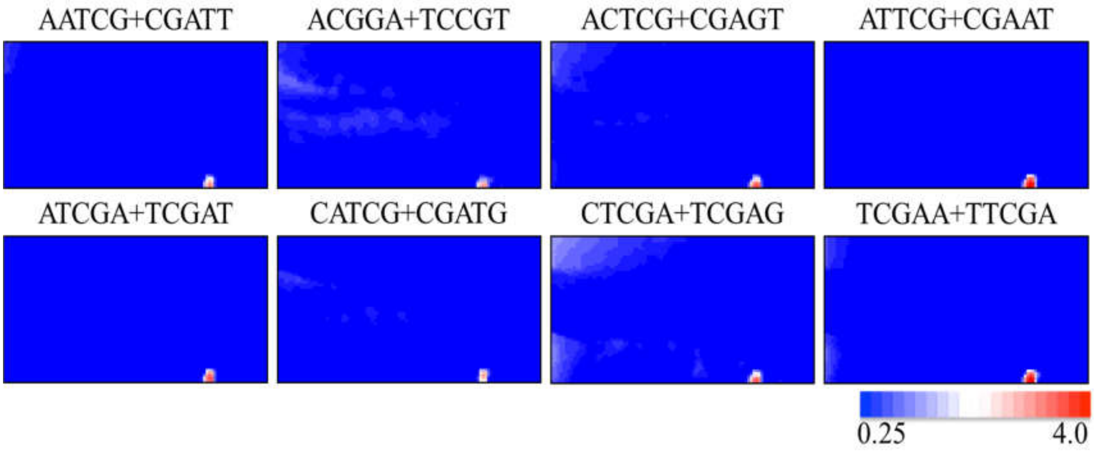


Fig. 2.5. CG-containing pentanucleotides enriched in Sz region.

The observed/expected ratio for each pair of complimentary pentanucleotides is calculated as described in Fig. 1C and indicated in color presented under the panel. The small red region for each pair of complimentary pentanucleotides corresponds to the dark black region in the U-matrix

2.4.2 Possible Biological Functions of Sequences with Peculiar Oligonucleotide Composition

As a separate analysis, we have examined the chromosomal locations of sequences belonging to the human specific zones and found a major portion of these 100 kb specific sequences to be derived from pericentromeric heterochromatin regions, as supporting the previous finding (Iwasaki Y. et al, 2013). Pericentromeric regions form the heterochromatin structure “chromocenter” in interphase nuclei. Chromocenter was once thought to be stable in composition and transcriptionally inert but has recently been shown to be surprisingly dynamic (Maison C. et al, 2002; Maison C. et al, 2004; Probst AV et al, 2009; Probst AV et al, 2010; Probst AV et al, 2011).

Mouse centromere-derived double-stranded transcripts appear to be involved in establishing the heterochromatin structure (Maison C. et al, 2002), and Dicer-related RNA interference machinery is involved in the formation of the centromeric heterochromatin structure in higher vertebrate cells (Fukagawa T. et al, 2004). A strand-specific burst in transcription of mouse pericentromeric satellites is required for chromocenter formation during early mouse development (Probst AV et al, 2010), and long nuclear noncoding RNA transcribed from the periphery of pericentromeric heterochromatin has recently been reported (Maison C. et al, 2011). Because the centromere RNA has been shown to be a key component for the assembly of nucleoproteins at the nucleolus and centromere (Du Y et al, 2010; Wong LH et al, 2011), the notable clustering of TFB motifs in the pericentromeric regions should provide novel knowledge about the higher order of nuclear organization.

In Figure 2.2, specific regions are mainly observed for the human genomic sequences. This appears to be related to the finding that the human specific sequences are mainly derived from the pericentromeric heterochromatin regions. In the case of mice, their chromosomes are acrocentric and the highly repetitive sequences in their pericentromeric regions are less represented in the reported genome sequence than for the human genome. When more sequences of the mouse pericentromeric regions will become available, comparative analyses of their sequences should provide novel information concerning biological significance of 100 kb level sequences with the very peculiar oligonucleotide compositions. In the present study, we have analyzed 100 kb sequences, but the analyses of 50 kb sequences give similar results.

2.4.3 Other Applications of BLSOM and Future Prospects

BLSOM can classify genomic sequences according to species with no information other than oligonucleotide frequencies. Because the classification and visualization power is very high, BLSOM is a powerful bioinformatics tool for extracting a wide range of information from a large amount of genomic sequences. A wide variety of oligonucleotide sequences function as genetic signals (e.g., regulatory signals for gene expression).

We have found that occurrence levels of oligonucleotide sequences corresponding to important functional signals (e.g., TFB motif sequences) are often biased significantly from the occurrence levels found in a major portion of the human genome and are diagnostic for the specific zones visualized in Figures 2.2 and 2.3. When we systematically characterize in advance the known signal sequences of various species with enough experimental data with BLSOMs, we may develop an in silicon method of signal prediction, which is most useful for genomes that are sequenced but for which little additional experimental data are available. Because the number of such genomes has increased rapidly, development of the in silicon method has become increasingly important. Functional signals, such as transcription-regulatory signals, are typically longer than pentanucleotides, and therefore analyses of longer oligonucleotides become important.

To conduct BLSOM with longer oligonucleotides such hexa- and heptanucleotides (4,096- and 16,384-dimensional data) for a massive amount of genome sequences currently available, a large-scale computation using a high-performance supercomputer will become essential, and the BLSOM algorithm is suitable for a high level of parallel computing.

One important application of BLSOM to genome informatics is the use for metagenome analyses. Most environmental microorganisms cannot be cultured easily under laboratory conditions. Genomes of uncultured organisms have remained mostly uncharacterized and are thought to contain a wide range of novel genes of scientific and industrial interest (Amann RI et al, 1995; Hugenholtz P. et al, 1996; Rondon MR et al, 2000; Lorenz P. et al, 2002; DeLong EF et al, 2002; Schloss PD et al, 2003).

Metagenomic approaches, which are analyses of mixed populations of uncultured microbes, have been developed to identify novel and industrially useful genes and to study microbial diversity in a wide variety of environments. With the metagenomic approach, genomic DNAs are extracted directly from an environmental sample containing multiple organisms, and the DNA fragments are cloned and sequenced. This is a powerful strategy for comprehensive analysis of biodiversity in an ecosystem.

However, for a simple collection of many sequence fragments, the conventional phylogenetic method cannot predict from what phylotypes individual sequences are derived or the phylogenetic novelty of the individual sequences. Traditional methods of phylogenetic assignment have been based on sequence homology searches and therefore inevitably focused on well-characterized genes, for which orthologous sequences required for constructing a reliable phylogenetic tree are available. However, most of the well-characterized genes are not industrially attractive. BLSOM is an alignmentfree clustering method, and thus is the most suitable method for this metagenomics analysis. For phylogenetic classification of species-unknown sequences obtained from environmental and clinical samples, we have constructed BLSOMs in advance with all available sequences from species-known prokaryotes and eukaryotes, as well as from viruses and organelles, and found that the sequences are clustered (self-organized) according to phylotypes with high accuracy (Abe T. et al, 2006).

By mapping a large number of environmental metagenomic sequences on the large-scale BLSOM, we can predict phylotypes of these environmental sequences (Abe T. et al, 2005). Because BLSOM does not require orthologous sequence sets, this alignment-free method can provide a systematic strategy for revealing microbial diversity and relative abundance of different phylotype members of uncultured microorganisms including viruses in an environmental sample (Abe T. et al, 2005). Actually, as collaborative studies with experimental research groups, we have used the BLSOM for phylogenetic classification of genomic sequence fragments obtained from mixed genomes of uncultured microbes in environmental samples (Hayashi H. et al, 2005; Nakao R. et al, 2013; Abe T. et al, 2009).

We have recently found that the addition of a large number of computer-generating random sequences can classify the metagenomic sequences according to phylotypes (Uehara H. et al, 2011). In addition, BLSOM with oligopeptide composition can classify protein sequences mainly according to function (Abe T. et al, 2009).

2.5 Conclusions

Because of the remarkable progress of various high-throughput measuring instruments, a massive amount of various data other than sequence data has been accumulated. Complex data can be represented by a high-dimensional multivariate data. BLSOM can analyze a massive amount of high dimensional multivariate data because the algorithm is suitable for high-level parallel computing. BLSOM can support efficient knowledge discoveries from such big data, showing that the BLSOM is a timely bioinformatics method in the era of big data studies in bioscience. In the present study, we characterized vertebrate genomes using BLSOM. We first analyzed pentanucleotide compositions in 100 kb sequences derived from a wide range of vertebrate genomes and then the compositions in the human and mouse genomes in order to investigate a method for detecting differences between the closely related genomes. BLSOM can recognize the species specific key combination of oligonucleotide frequencies in each genome, which is called a “genome signature,” and the specific regions specifically enriched by transcription-factorbinding sequences.

Chapter 3: Systematization of Transcriptome data based on BLSOM

3.1 Introduction

Euglena gracilis (*E. gracilis*) is a genus of single-celled flagellate Eukaryotes (Adl et al, 2012). *Euglena gracilis* is the best known and most widely studied member of the class *Euglenoidea* that contain some 54 genera and at least 800 species (Pechenik, J. A., 2014), which has been used extensively in the laboratory as a model organism (Russell, Anthony G., et al, 2005). *Euglenoids* are characterized by chlorophylls a and b, one membrane of chloroplast endoplasmic reticulum, a mesokaryotic nucleus, flagella with fibrillar hairs in one row, no sexual reproduction, and paramylon or chrysolaminarin as the storage product in the cytoplasm (Lee et al, 2008).

Euglena gracilis (wild type) have photosynthesizing chloroplasts within the body of the cell, which enable them to feed by autotrophy, like plants (Margulis et al, 1997). However, they can also take nourishment heterotrophically, like animals (Margulis et al, 1997).

Euglena gracilis (bleached) is decreased in chlorophyll (bleaching) and a gradual transformation of chloroplasts into rudimentary plastids (Ortiz, W. et al, 1988), which can't photosynthesize.

Paramylon is an important starting material for wax ester production under anaerobic conditions (Yoshida et al, 2016). When aerobically grown *E. gracilis* cells are transferred into anaerobic conditions, they will degrade paramylon to actively synthesize and accumulate wax esters, which consist of medium-chain fatty acids and alcohols including 14:0 carbon chains as the major constituent (Inui H. et al, 1982). This phenomenon is designated wax ester fermentation due to the concomitant generation of ATP without any energy loss during wax ester production (Inui H. et al, 1982). However, till now, it is still largely unknown about regulatory mechanism in response to anaerobic conditions in *Euglena gracilis*.

Using high-throughput Illumina RNA-seq technology, a total of 372,233,028 paired-end 100 bp reads were sequenced and *de novo* assembled by Trinity software (Haas, Brian J. et al, 2013; Grabherr et al, 2013) (version: trinity/r2013-02-25p1) to produce 358,420 transcripts. 47,399 transcripts were annotated by their homologous proteins from NCBI NR database, known transcription factors and protein kinases, respectively.

We then compared global expression profiles of *E. gracilis* between aerobic and anaerobic conditions, between wild type and bleached type, and between midlogarithmic growth phase and late logarithmic growth phase. Then we identified a number of statistically significant differentially expressed genes (DEGs) (false discovery rate (FDR) <0.05, |fold changes| > 2) (Reiner et al, 2003) for three compare group (2,406 differentially expressed unique transcripts between aerobic cells and anaerobic cells, among which 1,253 were induced by anaerobic stress and 1,153 repressed; 3,865 differentially expressed unique transcripts between wild type and bleached type of *E. gracilis*, among which 1,683 were induced and 2,182 repressed for different type; 2,764 differentially expressed unique transcripts between midlogarithmic growth phase and late logarithmic growth phase of wild type *E. gracilis*, among which 1,108 were induced by late logarithmic growth phase and 1,656 repressed).

Compared with the functional proteins, the transcription factors always act at the upstream position of the signal transduction and gene regulatory network, controlling a broad range of downstream genes; which makes them efficient in tolerating abiotic stress. To date, several transcription factors belonging to different transcription factor families, such as NAC, C3H, MYB and WRKY, have been implicated in the regulation of stress responses (Schwechheimer et al, 1998; Singh et al, 2002; Xu, Yanjie, et al, 2013).

I identified a total of 5,467 transcripts of 56 Transcription Factors (TFs) families through transcriptome sequencing, 518 transcripts of 39 TFs families that were responsive to anaerobic stress, 906 transcripts of 44 TFs families that were between wild type and bleached type, 792 transcripts of 43 TFs families that were between wild type and bleached type.

I have next constructed BLSOM with expression change of 7,291 statistically significant differentially expressed genes (DEGs) (Reiner et al, 2003) derived from the *Euglena gracilis* transcriptome, include 1,418 transcription factor and 153 KEGG metabolic pathways. BLSOM map is 350 × 192 lattice map. I identified the patterns of regulation changes for TFs and other gene.

3.2 Materials and Methods

3.2.1 Specimens

The wild-type *Euglena gracilis* and bleached type were cultured for 2 daysand 4 daysunder aerobic condition. Then cells were incubated 24 h under anaerobic condition. The second day is midlogarithmic growth phase for aerobic culture; the forth day is late logarithmic growth phase. For each condition, wild type and bleached type were performed. The sum of the sample is 8 samples. (Table 3.1)

Table 3.1. Eight samples original NGS data from wild type and bleached type of *E. gracilis*

sample ID	No. of reads	Size
G_2d	20,281.646×2	4.1GB
G_2d_24h_an	24,762,047×2	5.0GB
G_4d	27,084,024×2	5.5GB
G_4d_24h_an	24,194,393×2	4.9GB
B_2d	22,594,403×2	4.6GB
B_2d_24h_an	24,516,325×2	5.0GB
B_4d	19,683,518×2	4.0GB
B_4d_24h_an	23,000,158×2	4.6GB

G_2d means wild-type *Euglena gracilis* cultured for 2 days under aerobic condition, which is midlogarithmic growth phase; G_2d_24h_an means wild-type *Euglena gracilis* cultured for 24 h under anaerobic condition after 2 days under aerobic condition; G_4d means wild-type *Euglena gracilis* cultured for 4 days under aerobic condition, which is late logarithmic growth phase; G_4d_24h_an means wild-type *Euglena gracilis* cultured for 24 h under anaerobic condition after 4 days under aerobic condition; B_2d means bleached type *Euglena gracilis* cultured for 2 days under aerobic condition, which is midlogarithmic growth phase; B_2d_24h_an means bleached type *Euglena gracilis* cultured for 24 h under anaerobic condition after 2 days under aerobic condition; B_4d means bleached type *Euglena gracilis* cultured for 4 days under aerobic condition, which is late logarithmic growth phase; B_4d_24h_an means bleached type *Euglena gracilis* cultured for 24 h under anaerobic condition after 4 days under aerobic condition;

3.2.2 De novo transcriptome assembly

To obtain high-quality clean reads, the raw reads were filtered to remove reads with adaptor sequences, low quality reads (Phred quality score <20 bp), and reads with a high percentage of unidentified nucleotides, using Biopython (Cock, Peter JA et al, 2009).

De novo assembly of the clean reads was carried out using Trinity software (Haas, Brian J. et al, 2013; Grabherr et al, 2013) (version: trinity/r2013-02-25p1) with default parameters and no reference sequence. The sequences resulting from the *de novo* Trinity assembly were called unique transcripts (or unigenes) (Xu, Yanjie et al, 2013).

3.2.3 Annotation of Euglena gracilis unique transcript sequences

In order to annotate unigenes, a BLASTX (Altschul, Stephen F. et al, 1990) search against the the GenBank non-redundant (nr) database was conducted with an E-value cut-off of 1e−6. The Blast2GO program (Conesa, Ana et al, 2005) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa, Minoru et al, 2014) (http://www.genome.jp/kegg/) were used to identify the Gene ontology (GO) (Ashburner, Michael et al, 2000) annotation and biological pathways in *E. gracilis*, respectively. The annotated unique transcripts enriched in Gene ontology terms were calculated and the figures of Gene ontology terms were drawn by WEGO (Ye, Jia et al, 2006) (http://wego.genomics.org.cn/cgi-bin/wego/index.pl). The Plant Transcription Factor Database (Zhang, He et al, 2011) (http://planttfdb.cbi.pku.edu.cn/) was used to identify the transcription factors family of *Euglena gracilis*.

3.2.4 Gene expression quantification and differential expression analysis

I aligned the high-quality cleaned RNA-seq reads to the assembled *E. gracilis* transcripts with the Bowtie program (Langmead, Ben et al, 2009) allowing one mismatch. Following alignments, raw counts for each transcript and in each sample were derived and normalized to fragments per kilobase of exon model per million mapped reads (FPKM) (Wagner et al, 2012).

FPKM is given by:

$$\text{FPKM} = \frac{Counts\ of\ mapped\ fragments}{Lenth\ of\ transcript\ (kb) \times Total\ Counts\ of\ mapped\ fragments\ (million)} \tag{3.1}$$

If a fragment is mapped to multiple locations, it is only counted once for "Total Count" in the formula above. If uniquely mapped reads in a pair are aligned within an existing transcript, it is counted as 1; otherwise 0. Existing transcripts include annotated exons and exon junctions. In other words, partially aligned reads which indicates potential fusion events or novel exon junctions won't be counted.

We calculated FPKM by Trinity software.

Differentially expressed genes (|fold changes| > 2 and false discovery rate (FDR) < 0.005) were identified with the edgeR package (Robinson et al, 2010).

Fold changes given by:

$$\text{fold changes} = \log_2 \frac{Counts\ of\ gene\ a\ of\ sample\ A}{Counts\ of\ gene\ a\ of\ sample\ B} \tag{3.2}$$

The false discovery rate (FDR) is one way of conceptualizing the rate of type I errors in null hypothesis testing when conducting multiple comparisons (Benjamini et al, 1995).

The false discovery rate (FDR) is given by (Benjamini et al, 1995):

$$\text{FDR} = Q_e = E(Q) = E\left(\frac{V}{V+S}\right) = E\left(\frac{V}{R}\right) \tag{3.3}$$

V, S are unobservable random variables; R is the number of hypotheses rejected, which is an observable random variables; Q_e is the expectation of Q; Q is the proportion of the rejected null hypotheses which are erroneously rejected, which is an observable random variables. Q=0, when V+S=0.

3.2.5 BLSOM

TF and metabolic pathways co-expression networks was drawn by BLSOM. Here, we explain the BLSOM method developed by section 2.2.

BLSOM map was analyzed and displayed by R language ggplot2 package (H. Wickham, 2009).

3.3 Results

3.3.1 RNA-Seq and de novo assembly

We sequenced and *de novo* assemble *euglena gracilis* (wild type and bleached type) transcriptomes under aerobic and anaerobic condition, respectively. After 2 days and 4 days culture under aerobic condition, cells were incubated 24 h under anaerobic condition. The second day is midlogarithmic growth phase for aerobic culture; The forth day is late logarithmic growth phase. For each condition, wild type and bleached type were performed. The sum of the sample is 8 samples.

The *Euglena gracilis* spliced leader-sequence (SL) was shown to consist of 26 nucleotides after processing of precursor 5S rRNA (Tessier et al, 1991; Frantz et al, 2000). These SLs had slight variations, especially in the leader sequence (Tessier et al, 1991; Frantz et al, 2000). Therefore, 11 conserved nucleotides ("ATTTTTTTTCG") and 9 conserved nucleotides ("TTTTTTTCG") are rich in 5′-terminus short reads.

To obtain high-quality clean reads, the raw reads were filtered to remove reads with adaptor sequences, low quality reads (Phred quality score <20 bp), and reads with a high percentage of unidentified nucleotides, using Biopython (Cock, Peter JA et al, 2009).

De novo assembly of these clean short reads for 8 sample generated 358,420 unique transcripts with an average length of 750.9 bp and the longest transcript of 13,933 bp. The length distribution of the assembled *Euglena gracilis* unique transcripts is shown in Figure 3.1.

"ATTTTTTTTCG" and "TTTTTTTCG" were present at the 5' UTR of 30.53% (14,471 transcripts) of the transcripts.

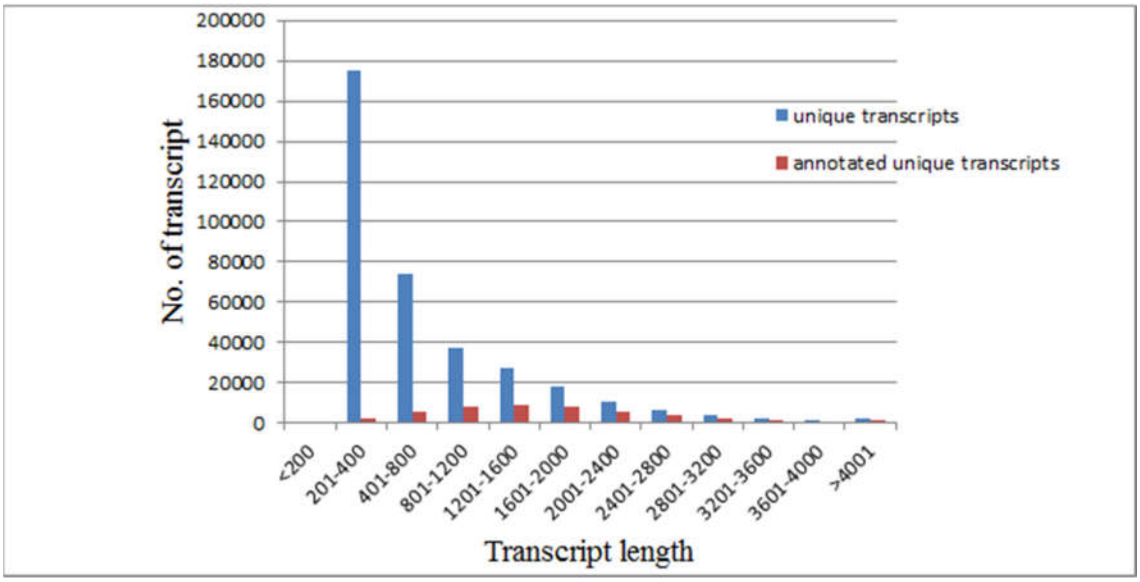


Fig. 3.1. Length distribution of *E. gracilis* unique transcripts.

Unique transcripts and annotated unique transcripts are indicated in red colour and green colour, respectively.

3.3.2 Annotation of transcripts

Using blastx (Altschul, Stephen F. et al, 1990) with an E-value cut-off of 1e−6, shown in Figure 3.1 and Figure 3.2 (A), 47,399 transcripts were annotated by known sequences in the NR database. The distribution of the top blastx hits shows a distribution across the kingdoms of life, shown in Figure 3.2 (B), highlighting the diversity of possible origins of genetic material present in the Euglena genome, which have been integrated through its complex phenotypic and genotypic history (O'Neill et al, 2015).

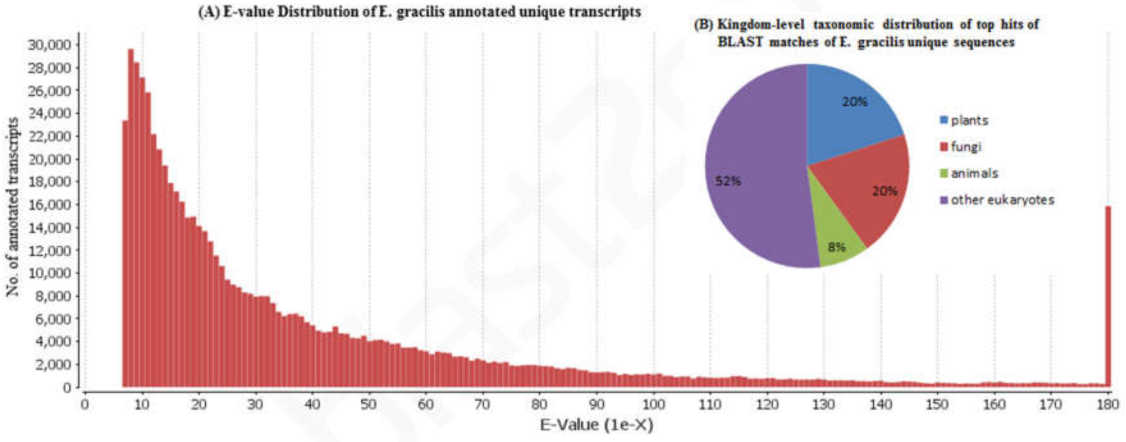


Fig. 3.2. E-value Distribution and Kingdom-level taxonomic distribution of *E. gracilis* unique sequences.

(A) E-value Distribution of E. gracilis annotated unique transcripts. (B) Kingdom-level taxonomic distribution of top hits of BLAST matches of E. gracilis unique sequences.

Of the sequences that can be annotated, 47,399 were classified into 522 Gene Ontology classifications (Gene Ontology Consortium, 2004) (GO level 2, shown in Fig. 3.3), covering the full spectrum of functions that one might expect in a complex eukaryotic organism.

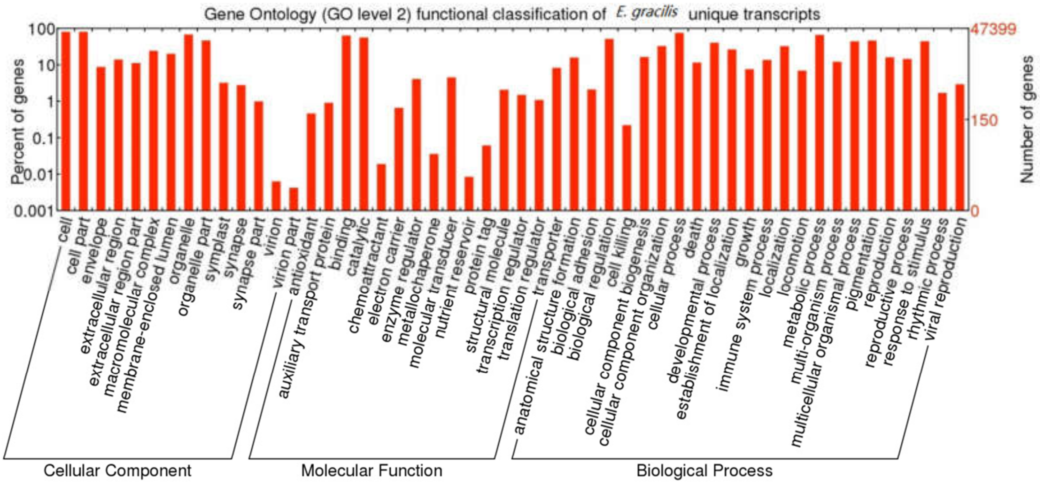


Fig. 3.3. Gene Ontology (GO level 2) functional classification of *E. gracilis* unique transcripts.

3.3.3 Metabolism pathways

Candidate genes analysis based on KEGG (Kanehisa, Minoru et al, 2014) were identified in the *Euglena* transcriptome for all of the metabolic pathways. Highly represented pathways included Biosynthesis of antibiotics, Purine metabolism, Pyruvate metabolism, Amino sugar and nucleotide sugar metabolism, Pyrimidine metabolism, Cysteine and methionine metabolism, Alanine, aspartate and glutamate metabolism, and Glycine, serine and threonine metabolism and so on.(Figure 3.4)

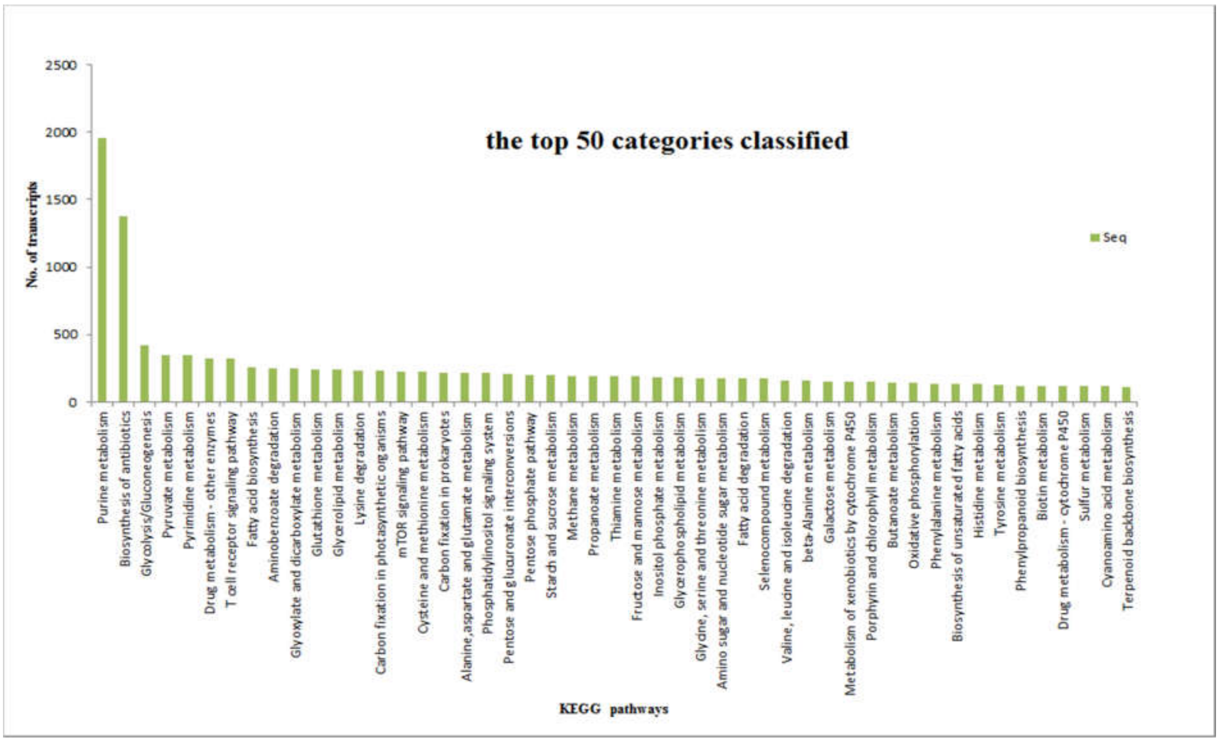


Fig. 3.4. KEGG pathway analysis of assembled components in *E. gracilis*.

The data exhibited are significantly enriched in the top 50 categories classified.

3.3.4 Dynamic transcriptome profiles under anaerobic condition

Following alignments, raw counts for each transcript and in each sample were derived and normalized to fragments per kilobase of exon model per million mapped reads (FPKM) (Wagner et al, 2012).

According to FPKM, I further calculated correlation coefficients of transcriptome profiles among the 8 samples (Table 3.2). The values of correlation coefficients between aerobic condition and anaerobic condition are the highest of all, which include correlations between midlogarithmic growth phase and late logarithmic growth phase and correlations between wild type and bleached type.

Table 3.2. Correlation coefficients of transcriptome profiles among RNA-seq samples

	B_2d	B_2d_24_an	B_4d	B_4d_24h_an	G_2d	G_2d_24h_an	G_4d	G_4d_24h_an
B_2d	1	0.978016	0.93066	0.93345	0.940417	0.946491	0.782343	0.915495
B_2d_24h_an	0.978016	1	0.944712	0.951279	0.951716	0.952516	0.779237	0.92025
B_4d	0.93066	0.944712	1	0.99602	0.954239	0.950117	0.788199	0.944168
B_4d_24h_an	0.93345	0.951279	0.99602	1	0.951123	0.955976	0.789238	0.947841
G_2d	0.940417	0.951716	0.954239	0.951123	1	0.97186	0.823813	0.949188
G_2d_24h_an	0.946491	0.952516	0.950117	0.955976	0.97186	1	0.83886	0.97056
G_4d	0.782343	0.779237	0.788199	0.789238	0.823813	0.83886	1	0.913309
G_4d_24_an	0.915495	0.92025	0.944168	0.947841	0.949188	0.97056	0.913309	1

G_2d means wild-type *Euglena gracilis* cultured for 2 days under aerobic condition, which is midlogarithmic growth phase; G_2d_24h_an means wild-type *Euglena gracilis* cultured for 24 h under anaerobic condition after 2 days under aerobic condition; G_4d means wild-type *Euglena gracilis* cultured for 4 days under aerobic condition, which is late logarithmic growth phase; G_4d_24h_an means wild-type *Euglena gracilis* cultured for 24 h under anaerobic condition after 4 days under aerobic condition; B_2d means bleached type *Euglena gracilis* cultured for 2 days under aerobic condition, which is midlogarithmic growth phase; B_2d_24h_an means bleached type *Euglena gracilis* cultured for 24 h under anaerobic condition after 2 days under aerobic condition; B_4d means bleached type *Euglena gracilis* cultured for 4 days under aerobic condition, which is late logarithmic growth phase; B_4d_24h_an means bleached type *Euglena gracilis* cultured for 24 h under anaerobic condition after 4 days under aerobic condition;

According to a false discovery rate (FDR) (FDR <0.05, |fold changes| > 2), expression profiles were identified as statistically significant differentially expressed genes (DEGs) (Reiner et al, 2003). In total, I identified 2,406 differentially expressed unique transcripts between aerobic cells and anaerobic cells, among which 1,253 were induced by anaerobic stress and 1,153 repressed.

I then identified GO terms (level 2) in the Cellular Component category (13), Molecular Function category (11), and biological process category (22) that were significantly enriched in anaerobic-induced and repressed genes (shown in Figure 3.5).

Synapse (Cellular Component) only has up-regulated transcripts. Auxiliary transport protein (Molecular Function) only has down-regulated transcripts

In the category of Biological Process (level 3), "cell communication", "anatomical structure formation involved in morphogenesis", "multicellular organismal process", "multicellular organismal development", "regulation of body fluid levels", "regulation of metabolic process", "positive regulation of cellular process", "regulation of cellular process", "response to biotic stimulus" and "regulation of biological process" were highly enriched in both induced and repressed genes.

In the category of molecular function (level 3), "structural constituent of ribosome" and "chromatin binding" were highly enriched in both induced and repressed genes.

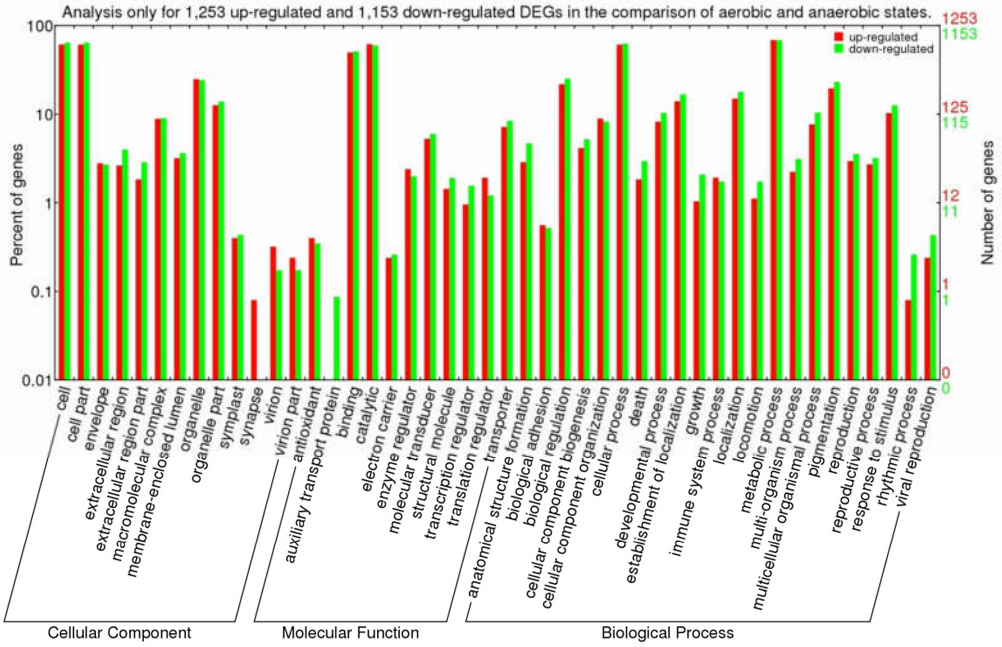


Fig. 3.5. Analysis only for 1,253 up-regulated and 1,153 down-regulated DEGs in the comparison of aerobic and anaerobic states.

Up-regulated genes and down-regulated genes are indicated in red colour and green colour, respectively. GO terms (level 2) include cellular component, molecular function and biological process.

3.3.5 Dynamic transcriptome profiles between wild type and bleached type under aerobic condition

According to a false discovery rate (FDR) (FDR <0.05, |fold changes| > 2), expression profiles were identified as statistically significant differentially expressed genes (DEGs). In total, I identified 3,865 differentially expressed unique transcripts between wild type and bleached type of *E. gracilis*, among which 1,683 were induced and 2,182 repressed for different typ.

I then identified GO terms (level 2) in the Cellular Component category (14), Molecular Function category (15), and biological process category (23) that were significantly enriched in anaerobic-induced and repressed genes (shown in Figure 3.6).

Virion part (Cellular Component), chemoattractant activity (Molecular Function), nutrient reservoir activity (Molecular Function), and protein tag (Molecular Function) only have up-regulated transcripts. Metallochaperone activity (Molecular Function) only has down-regulated transcripts

In the category of Cellular Component (level 3), “protein complex”, and “cell leading edge” were highly enriched in both induced and repressed genes.

In the category of Biological Process (level 3), “sexual reproduction”, “catabolic process”, “organic ether metabolic process”, “pigment metabolic process”, “macromolecule metabolic process”, “multicellular organismal metabolic process”, “primary metabolic process”, “oxidation reduction”, “organelle organization”, “microtubule-based process”, “cell communication”, “pollen germination”, “cell projection organization”, and “cell junction organization” were highly enriched in both induced and repressed genes.

In the category of molecular function (level 3), “transferase activity”, “oxidoreductase activity”, “hydrolase activity”, “protein binding”, and “nucleotide binding” were highly enriched in both induced and repressed genes.

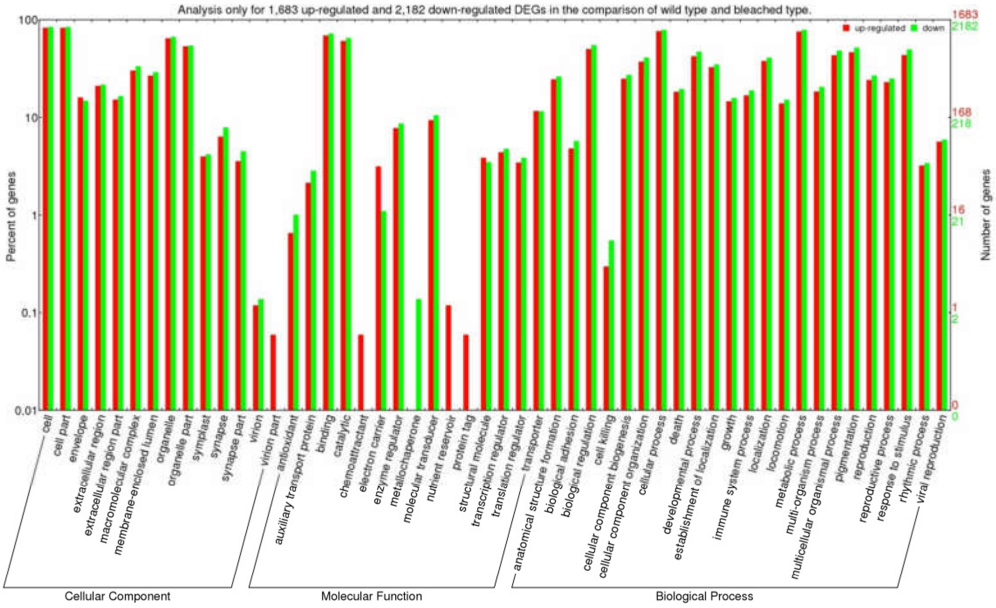


Fig. 3.6. Analysis only for 1,683 up-regulated and 2,182 down-regulated DEGs in the comparison of wild type and bleached type.

Up-regulated genes and down-regulated genes are indicated in red colour and green colour, respectively. GO terms (level 2) include cellular component, molecular function and biological process.

3.3.6 Dynamic transcriptome profiles between midlogarithmic growth phase and late logarithmic growth phase under aerobic condition

According to a false discovery rate (FDR) (FDR <0.05, |fold changes| > 2), expression profiles were identified as statistically significant differentially expressed genes (DEGs). In total, I identified 3,636 differentially expressed unique transcripts between midlogarithmic growth phase and late logarithmic growth phase, among which 1,561 were induced and 2,075 repressed.

I then identified GO terms (level 2) in the Cellular Component category (13), Molecular Function category (15), and biological process category (23) that were significantly enriched in anaerobic-induced and repressed genes (shown in Figure 3.7).

Nutrient reservoir activity (Molecular Function) only has up-regulated transcripts.

Virion (Cellular Component) only has down-regulated transcripts

In the category of Cellular Component (level 3), “Cellular Component”, “organelle envelope”, “protein complex”, “membrane-bounded organelle”, “non-membrane-bounded organelle”, “intracellular organelle”, “organelle part”, and “extracellular space” were highly enriched in both induced and repressed genes.

In the category of Biological Process (level 3), “sexual reproduction”, “reproductive process”, “immune system development”, “antigen processing and presentation”, “leukocyte migration”, “nitrogen compound metabolic process”, “catabolic process”, “pigment metabolic process”, “macromolecule metabolic process” and “biosynthetic process” were highly enriched in both induced and repressed genes.

In the category of molecular function (level 3), “cyclase activity”, “hydrolase activity”, “vitamin transporter activity”, “nucleotide binding”, “pattern binding” and “nucleic acid binding” were highly enriched in both induced and repressed genes.

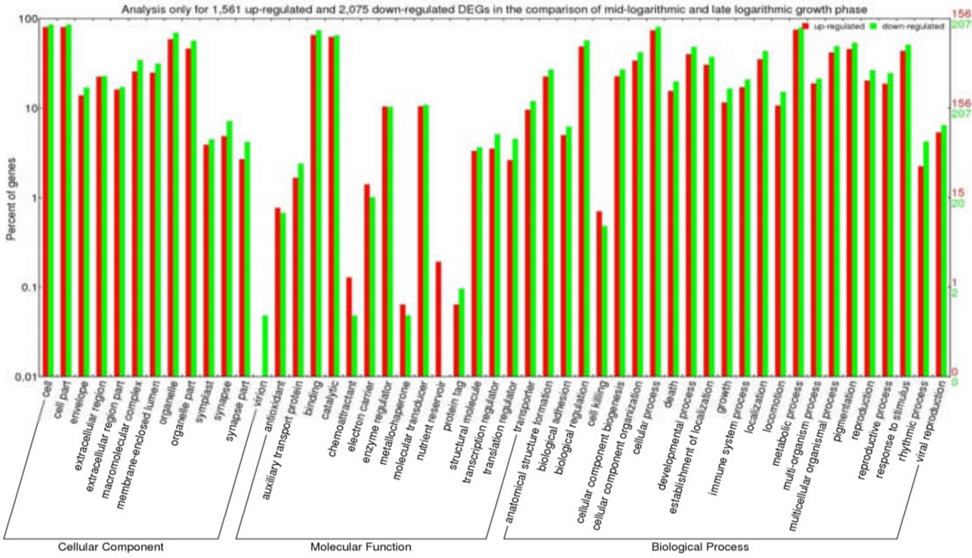


Fig. 3.7. Analysis only for 1,561 up-regulated and 2,075 down-regulated DEGs in the comparison of mid-logarithmic and late logarithmic growth phase.

Up-regulated genes and down-regulated genes are indicated in red colour and green colour, respectively. GO terms (level 2) include cellular component, molecular function and biological process.

3.4 Discussion

3.4.1 Identification of TF-encoding transcripts through transcriptome sequencing

Transcription factors and protein kinases are important upstream regulators of plant responses to various biotic and abiotic stresses. In this study, I identified a total of 5,467 transcripts of 56 Transcription Factors (TFs) families through transcriptome sequencing using the Plant Transcription Factor Database, shown on Figure 3.8.

The E2F/ DP family (Wu et al, 1995) was the largest group. The E2F transcription factor are a family of proteins that share a related DNA-binding domain and that bind to overlapping sets of target promoters, which are conserved in many species of plants and animals (Van Den Heuvel et al, 2008). Most E2F proteins associate with a DP protein and form heterodimeric complexes that bind to DNA in a sequence-specific manner, which control the temporal expression of genes that are needed for multiple processes during the cell cycle (Van Den Heuvel et al, 2008). Consequently, the level of E2F-dependent transcription is important for cell proliferation (Van Den Heuvel et al, 2008). Different types of E2F complexes either activate or repress transcription (Van Den Heuvel et al, 2008). E2F repressor complexes suppress the transcription of their targets in quiescent cells, in differentiated cells and during the G1 phase of the cell cycle (Van Den Heuvel et al, 2008).

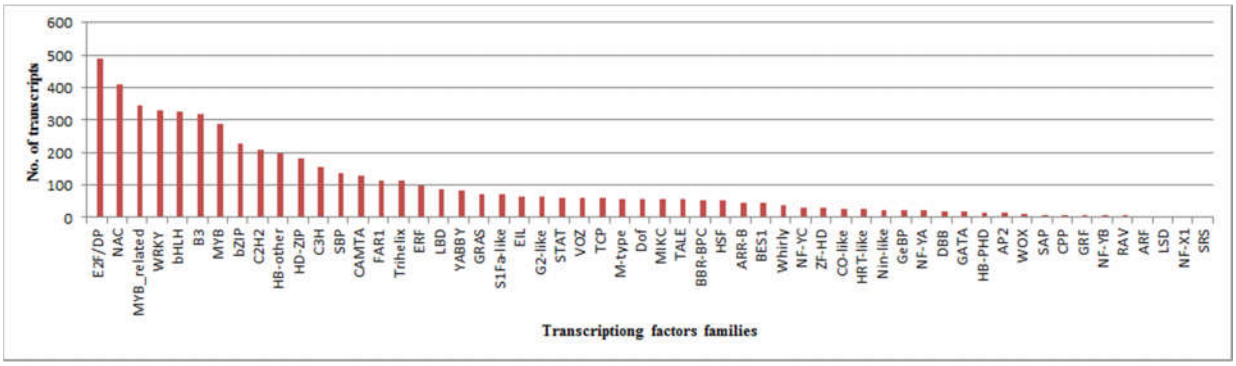


Fig. 3.8. Identification of TF-encoding 5,467 transcripts of 56 TFs families through transcriptome sequencing.

3.4.2 Transcription factors profiles under anaerobic condition

In this study, we identified a total of 518 transcripts of 39 TFs families that were responsive to anaerobic stress (Shown in Figure 3.9). The WRKY family (Rushton et al, 2010) was the largest group.

The WRKY transcription factors are one of the largest families of transcriptional regulators in plants and form integral parts of signaling webs that modulate many plant processes, which often act as repressors as well as activators, and those members of the family play roles in both the repression and derepression of important plant processes (Rushton et al, 2010).

No. of down-regulated transcripts from YABBY transcription factors (Chrispeels et al, 2000) and NAC transcription factors (Fang et al, 2008) are more than up-regulated. Members of YABBY transcription family are responsible for the specification of abaxial cell fate in lateral organs of *Arabidopsis*, which this family may be plant-specific, as genes with a similar arrangement of zinc finger and YABBY domains do not occur in either *Caenorhabditis elegans* or *Saccharomyces cerevisiae* (Chrispeels et al, 2000). NAM, ATAF, and CUC (NAC) transcription factors comprise a large plant-specific gene family and a few members of this family have been characterized for their roles in plant growth, development, and stress tolerance (Fang et al, 2008).

BBC-BPC transcription factors (Kooiker et al, 2005), EIL transcription factors (Yamasaki et al, 2005), HD-ZIP transcription factors (Ariel et al, 2007), and BES1 transcription factors (Yin et al, 2005) only have down-regulated transcripts.

BBC-BPC transcription factors controls ovule identity, and characterized its mechanism of action (Kooiker et al, 2005). The EIL transcription factors regulate various processes related to growth, development, and stress response of higher plants (Yamasaki et al, 2005). The HD-Zip transcription factors are unique to the plant kingdom, which participate in organ and vascular development or meristem maintenance and mediate the action of hormones or are involved in responses to environmental conditions (Ariel et al, 2007). BES1 defines a class of plant-specific transcription factors which regulate plant growth and development (Yin et al, 2005)

WOX transcription factors (Ariel et al, 2007; van der Graaff et al, 2009), TCP transcription factors (Martín-Trillo et al, 2010), Trihelix transcription factors (Fang et al, 2009), CAMTA transcription factors (Finkler et al, 2007), ARF transcription factors (Guilfoyle et al, 2007) only have up-regulated transcripts.

The WOX transcription factors form a plant-specific subclade of the eukaryotic homeobox transcription factor superfamily, which is characterized by the presence of a conserved DNA-binding homeodomain (van et al, 2009). According to the analysis of WOX gene expression and function, WOX family members fulfill specialized functions in key developmental processes in plants, such as embryonic patterning, stem-cell maintenance and organ formation, which can be related to either promotion of cell division activity and/or prevention of premature cell differentiation (van et al, 2009). TCP transcription factors have been found in various plant species, and new roles in plant development have been elucidated, which are the importance of this plant-specific gene family in the evolution and developmental control of plant form (Martín-Trillo et al, 2010). The Trihelix transcription factors are a plant-specific transcription factor family with a DNA-binding domain that binds GT elements, which are highly degenerate cis-elements with A/T-rich core sequences (Fang et al, 2009). The CAMTA transcription factors in various organisms imply a broad range of functions from sensory mechanisms to embryo development and growth control (Finkler et al, 2007). Auxin response factors (ARF) are transcription factors that regulate the expression of auxin response genes (Guilfoyle et al, 2007).

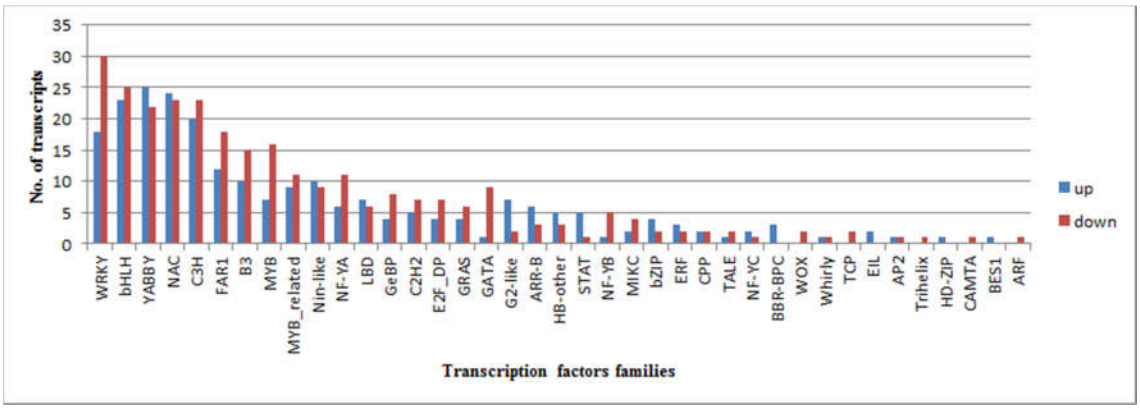


Fig. 3.9. Identification of TF-encoding 518 transcripts of 39 TFs families responsed to anaerobic stress.

Up-regulated genes and down-regulated genes are indicated in blue colour and down colour, respectively.

3.3.3 Transcription factors profiles between wild type and bleached type under aerobic condition

In this study, we identified a total of 906 transcripts of 44 TFs families that were between wild type and bleached type in *E. gracilis* (Shown in Figure 3.10). The bHLH (Toledo-Ortiz et al, 2003) family was the largest group.

The basic/helix-loop-helix (bHLH) proteins are a superfamily of transcription factors that have been well characterized in nonplant eukaryotes, especially in mammalian systems, which are important regulatory components in transcriptional networks in these systems, controlling a diversity of processes from cell proliferation to cell lineage establishment (Toledo-Ortiz et al, 2003).

HD-ZIP transcription factors (Ariel et al, 2007), and S1Fa-like transcription factors (Zhou et al, 1995) only have down-regulated transcripts.

The S1Fa-like transcription factors are highly conserved between dicotyledonous and monocotyledonous plants and may represent a novel class of DNA binding proteins, which are accumulated more in roots and in etiolated seedlings than in green leaves (Zhou et al, 1995). This expression pattern is correlated with the tissue-specific function of the S1F binding site which represses the rps1 promoter preferentially in roots and in etiolated plants (Zhou et al, 1995).

ARF transcription factors (Guilfoyle et al, 2007) only have up-regulated transcripts.

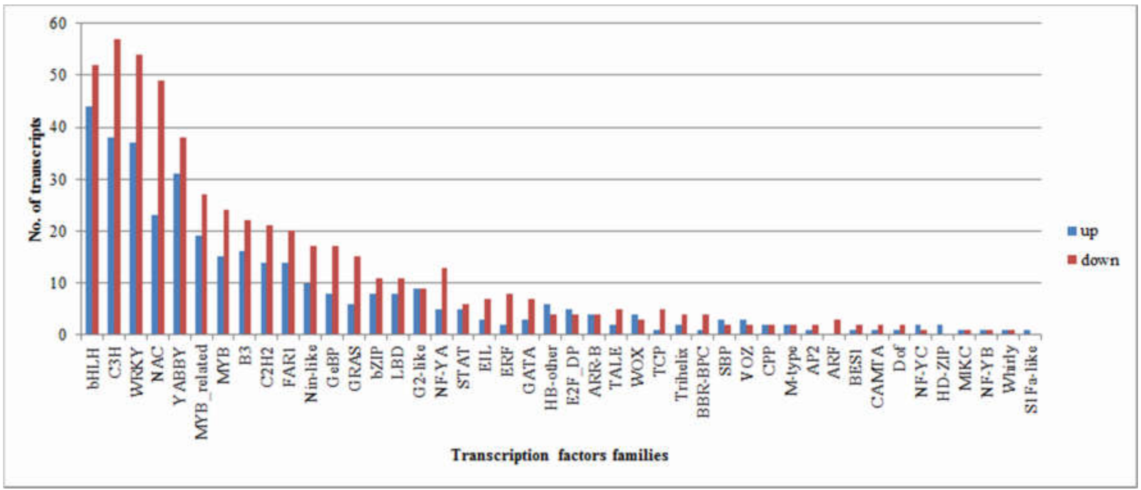


Fig. 3.10. Identification of TF-encoding 906 transcripts of 44 TFs families between wild type and bleached type.

Up-regulated genes and down-regulated genes are indicated in blue colour and down colour, respectively.

3.4.4 Transcription factors profiles between midlogarithmic growth phase and late logarithmic growth phase under aerobic condition

In this study, we identified a total of 792 transcripts of 43 TFs families that were between midlogarithmic growth phase and late logarithmic growth phase in *E. gracilis* (Shown in Figure 3.11). The C3H (Li et al, 1998) family was the largest group.

The C3H transcription factors encode a protein containing a Cys3His zinc finger domain associated with a number of animal and fungal transcription factors, which plays an important role during Arabidopsis embryogenesis, functioning primarily in the apical domain of the embryo.

NF-YC transcription factors (Siefers et al, 2009), RAV transcription factors (Nakano et al, 2006), TALE transcription factors (Hamant et al, 2010) and Whirly transcription factors (Desveaux et al, 2005) only have down-regulated transcripts.

NF-Y transcription factors are likely found in all eukaryotes and have roles in the regulation of diverse genes (Siefers et al, 2009). The members of the RAV family have important functions in the transcriptional regulation of a variety of biological processes related to growth and development, as well as various responses to environmental stimuli, which also are involved in ethylene response and in brassinosteroid response (Nakano et al, 2006). The three-amino-acid-loop-extension (TALE) class of homeoproteins has been shown to control meristem formation and/or maintenance, organ morphogenesis, organ position, and several aspects of the reproductive phase (Hamant et al, 2010).

MIKC transcription factors (Pařenicová et al, 2003), CAMTA transcription factors (Finkler et al, 2007), VOZ transcription factors (Mitsudan et al, 2004), ARF transcription factors (Windhövel et al, 2001), HRT-like transcription factors (Raventós et al, 2008) and SRS transcription factors (Byzova et al, 1999) transcription factors only have up-regulated transcripts.

MIKC transcription factors are key regulators of several plant development processes (Pařenicová et al, 2003). The expression level of VOZ transcription factors is changed at the transcriptional level in response to various environmental conditions and developmental stages for low-temperature, sugar, auxin and pollen-specific (Mitsudan et al, 2004). The HRT-like transcription factors was identified by southwestern screening with baits containing a gibberellin phytohormone response element from an alpha-amylase promoter, which accumulate to low levels in various tissues (Raventós et al, 2008). SRS transcription factors may act as a negative regulator of GA responses through transcriptional control (Byzova et al, 1999).

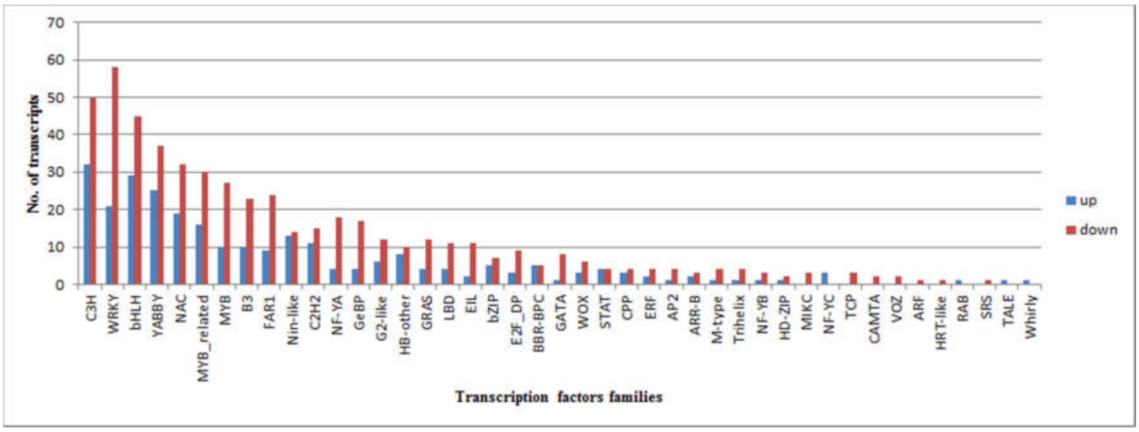


Fig. 3.11. Identification of TF-encoding 792 transcripts of 43 TFs families between midlogarithmic growth phase and late logarithmic growth phase.

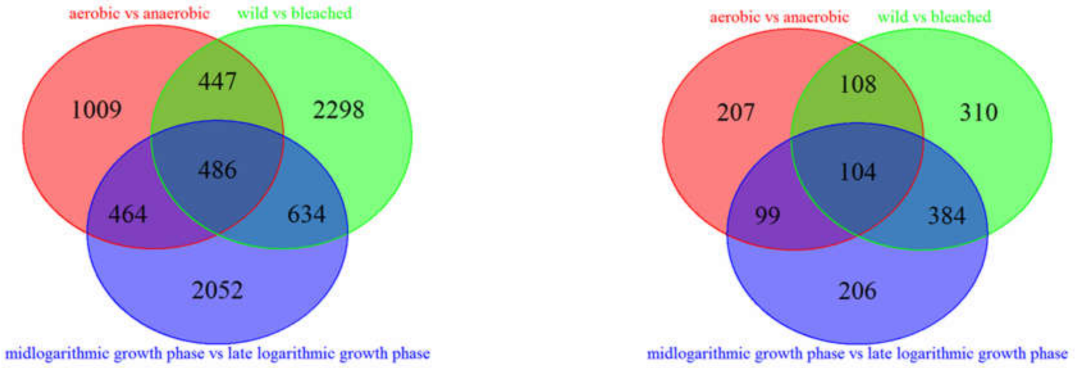
Up-regulated genes and down-regulated genes are indicated in blue colour and down colour, respectively.

3.4.5 TF families and other genes for co-expression by BLSOM

I further had constructed BLSOM with expression change of 7,390 DEGs of the *Euglena gracilis* transcriptome, which were derived from aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase (Figure 3.12, Figure 3.13), include 1,418 transcription factors.

BLSOM map is 350 × 192 lattice map. Lattice points containing transcripts are indicated in black and lattice points without any transcript are indicated in white colour (Figure 3.13).

BLSOM map was analyzed and displayed by R language ggplot2 package (H. Wickham, 2009).



(A) 7,390 DEGs of the *Euglena gracilis* transcriptome from aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase

(B) 1,418 transcription factors of DEGs from aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase

Fig. 3.12. Venn Diagram of 7,390 DEGs and TF-encoding 1,418 transcripts of 47 TFs families between midlogarithmic growth phase and late logarithmic growth phase.

(A) 7,390 DEGs of the *Euglena gracilis* transcriptome from aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase. (B) 1,418 transcription factors of DEGs from aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase

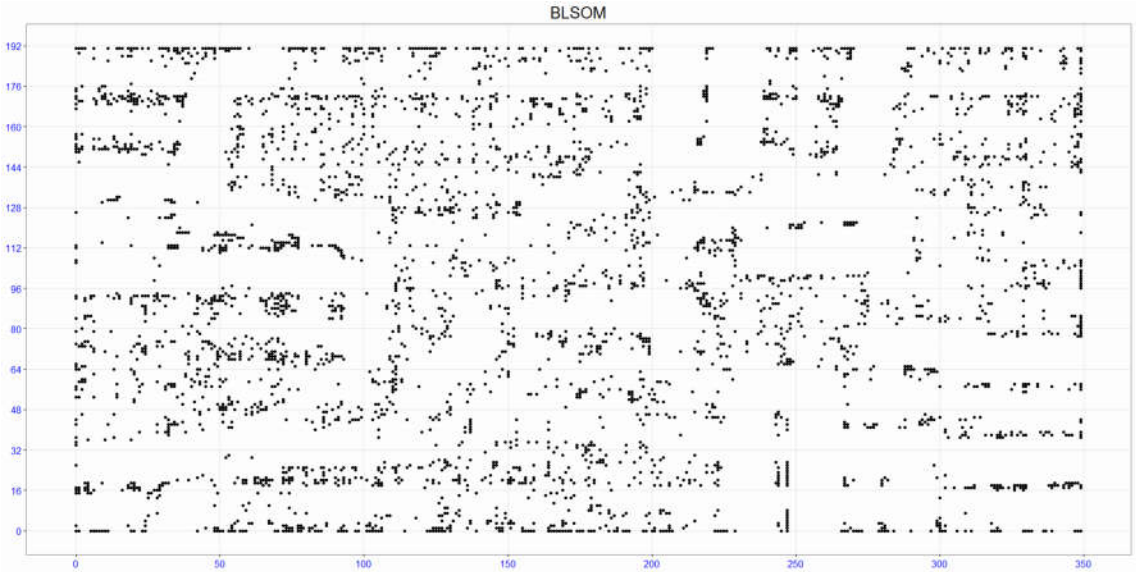
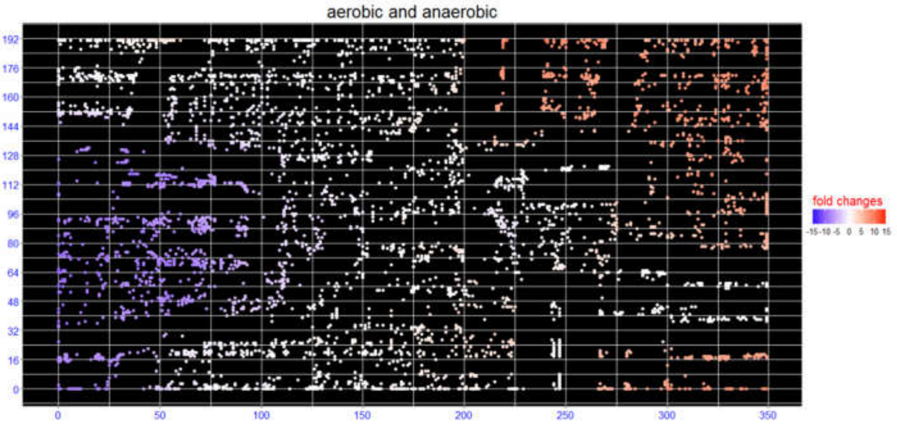


Fig. 3.13. BLSOM for DEGs under different conditions.

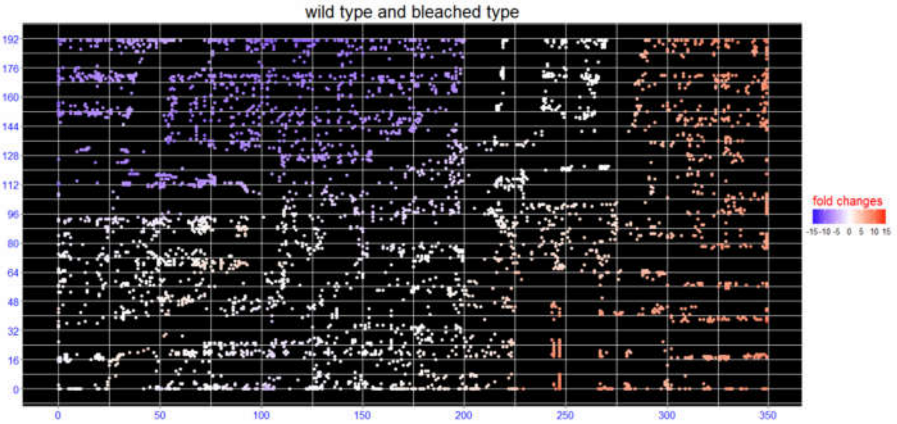
The map is a lattice comprised of 350 × 192. Lattice points containing transcripts are indicated in black and lattice points without any transcript are indicated in white.

In Figure 3.14, the fold changes logarithm value of TF families transcripts and other transcripts were explained in red (up), blue (down), and white (middle).

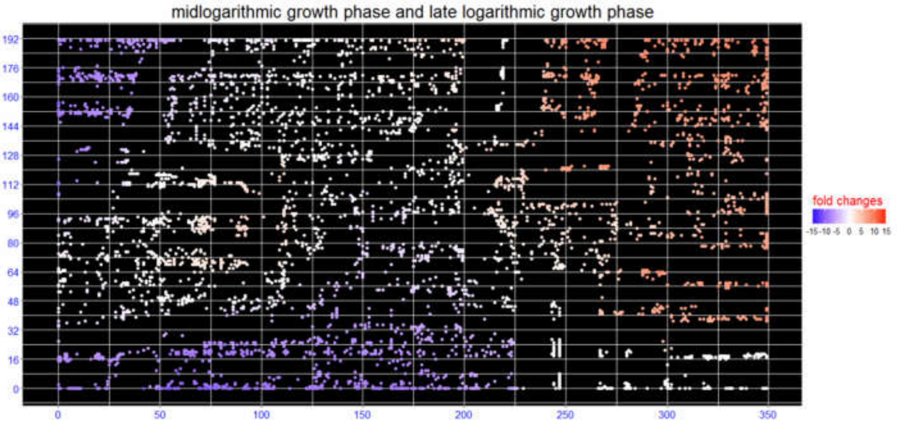
The up-regulated territories are similar for aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase, which belong to KEGG metabolic pathways, include reproductive process, immune effector process, and nitrogen compound metabolic process.



(A) aerobic and anaerobic condition



(B) wild type and bleached type



(C) midlogarithmic growth phase and late logarithmic growth phase

Fig. 3.14. The fold changes logarithm value of TF families transcripts and other transcripts for BLSOM.

(A) aerobic and anaerobic condition, (B) wild type and bleached type, (C) midlogarithmic growth phase and late logarithmic growth phase. The fold changes logarithm value indicated in colors presented in red (up), blue (down), and white (middle); lattice points without any transcript are indicated in black colour.

In Figure 3.15, I further analyzed distribution of transcripts of TF families. Lattice points only with other gene (not TFs) transcripts are indicated in grey and lattice points without any transcript are indicated in white; Lattice points only with different TFs families are indicated in different colors as shown in the bottom of the panel.

In table 3.3, we expressed pattern of DEGs with three types of expression pattern, Up (fold changes > 2), down (fold changes < -2), normal (-2 < fold changes < 2). According to expression patterns of aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase, BLSOM map is divided into 26 areas. Area 1 – area 26 are indicated by different colours presented, respectively. The expression patterns of the corresponding region are shown in Figure 3.17, Figure 3.18 and Figure 3.19. Transcription factors and other genes placed at same area on the map follows the same patterns of regulation changes.

In area 14, expression pattern is down (aerobic and anaerobic conditions, fold changes < -2), down (wild type and bleached type, fold changes < -2), down (midlogarithmic growth phase and late logarithmic growth phase, fold changes < -2). BLSOM cannot classify similar value. So the area 14 and the area 17 (expression pattern is down, normal, down) have a few regions overlap. However, BLSOM can be easy to classify different patterns in the same expression pattern for region C and region D of area 14 (shown in 14 (c) and 14 (d)).

When we consider co-expression between TFs and other genes that can function as important regulatory rules such as the target genes of transcription factors, their the patterns of regulation changes in dynamic transcriptome profile should be biased significantly from different condition. Therefore, the patterns of regulation changes only in a restricted portion of the BLSOM are thought to provide useful information for predicting the biological function significance of TFs, especially when the target genes of the TF of interest are unknown.

Table 3.3. The expression pattern of 26 regions of BLSOM

aerobic vs anaerobic condition regulation	wild type vs bleached type regulation	midlogarithmic growth phase vs late logarithmic growth phase regulation	No. of area
up	up	up	1
up	up	down	2
up	up	normal	3
up	down	up	4
up	down	down	5
up	down	normal	6
up	normal	up	7
up	normal	down	8
up	normal	normal	9
down	up	up	10
down	up	down	11
down	up	normal	12
down	down	up	13
down	down	down	14
down	down	normal	15
down	normal	up	16
down	normal	down	17
down	normal	normal	18
normal	up	up	19
normal	up	down	20
normal	up	normal	21
normal	down	up	22
normal	down	down	23
normal	down	normal	24
normal	normal	up	25
normal	normal	down	26

Up regulation is fold change value >2; down regulaton is fold change value <-2; normal regulation is other fold change value

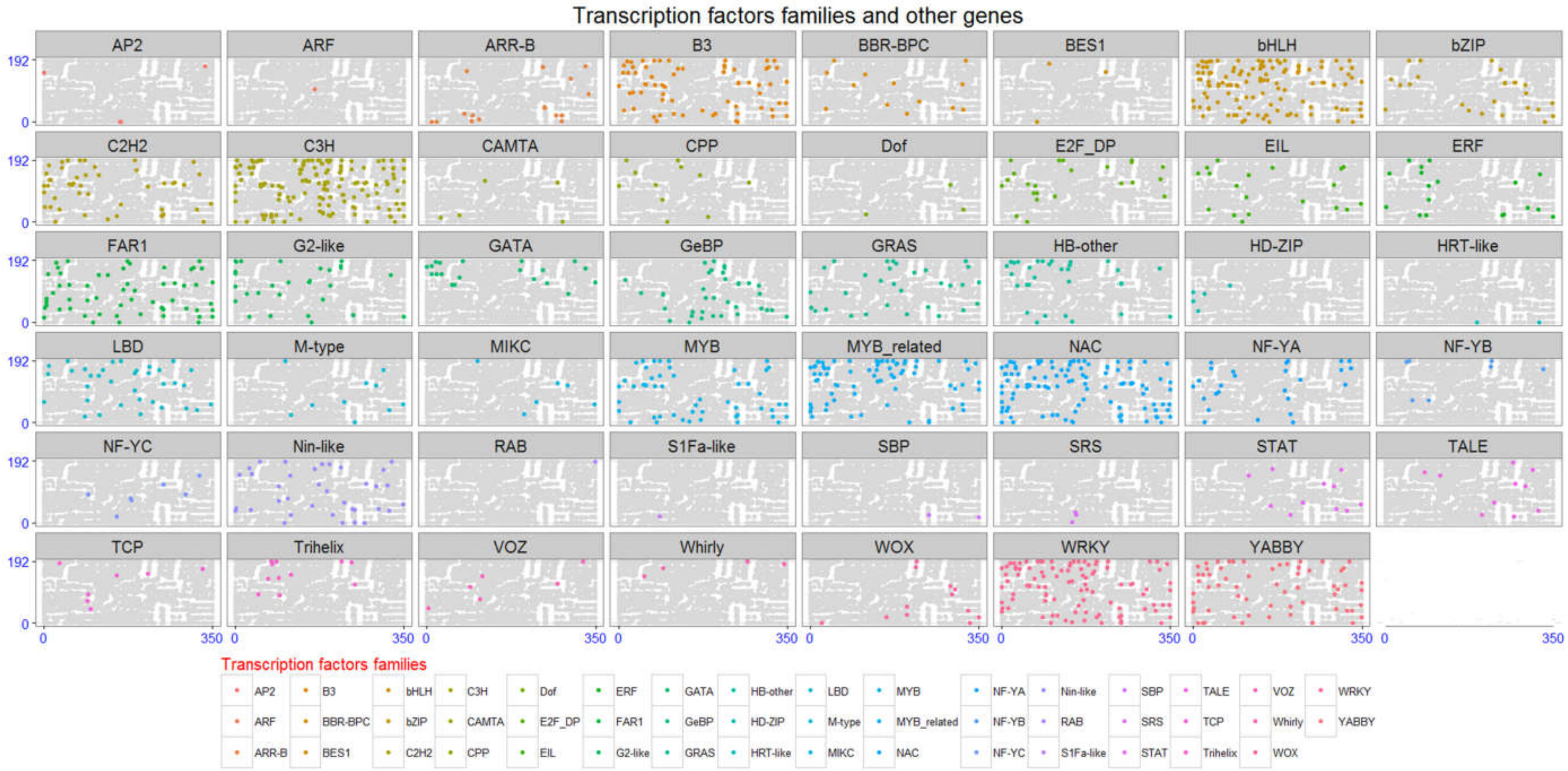


Fig. 3.15. Analysis of TF families transcripts and other gene transcripts for BLSOM.

Lattice points only with other gene (not TFs) transcripts are indicated in grey and lattice points without any transcript are indicated in white; Lattice points only with different TFs families are indicated by different colors.

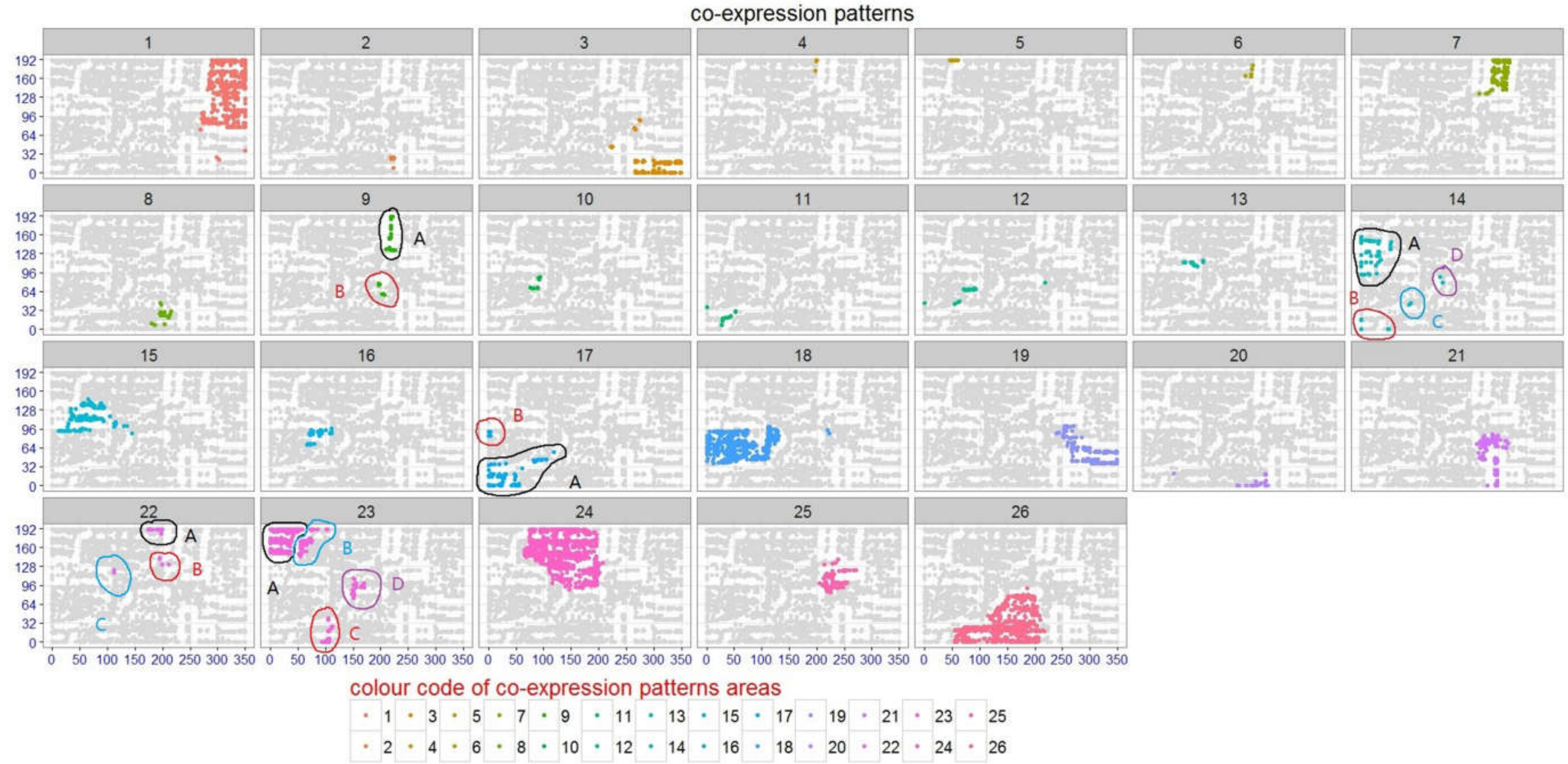
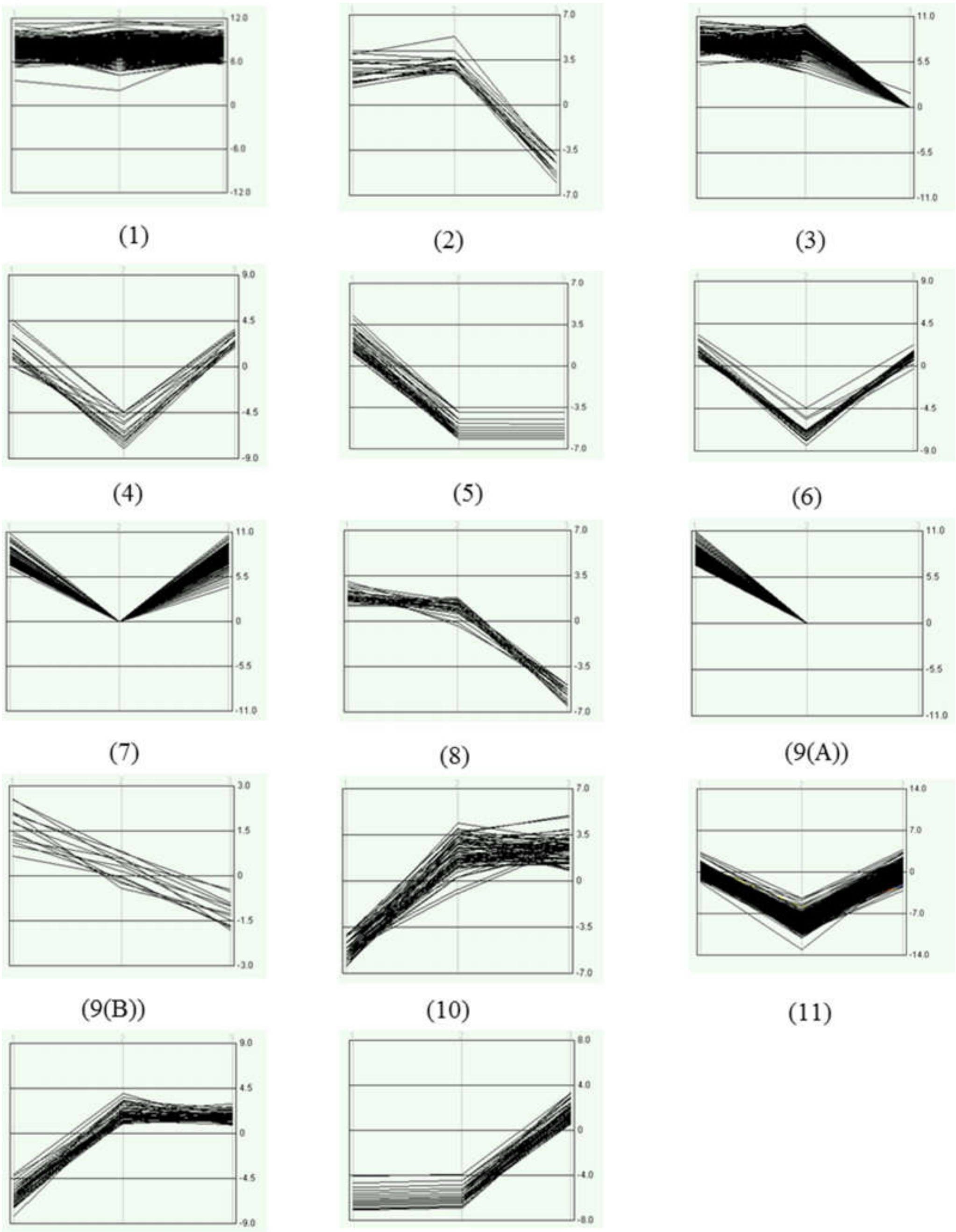


Fig. 3.16. Co-expression patterns analyses of TF families transcripts and other gene transcripts for BLSOM.

According to expression patterns, BLSOM map is divided into several areas. Areas 1-26 are indicated by different colours presented, respectively. Lattice points without any transcript are indicated in white colour.



(1) (2) (3)

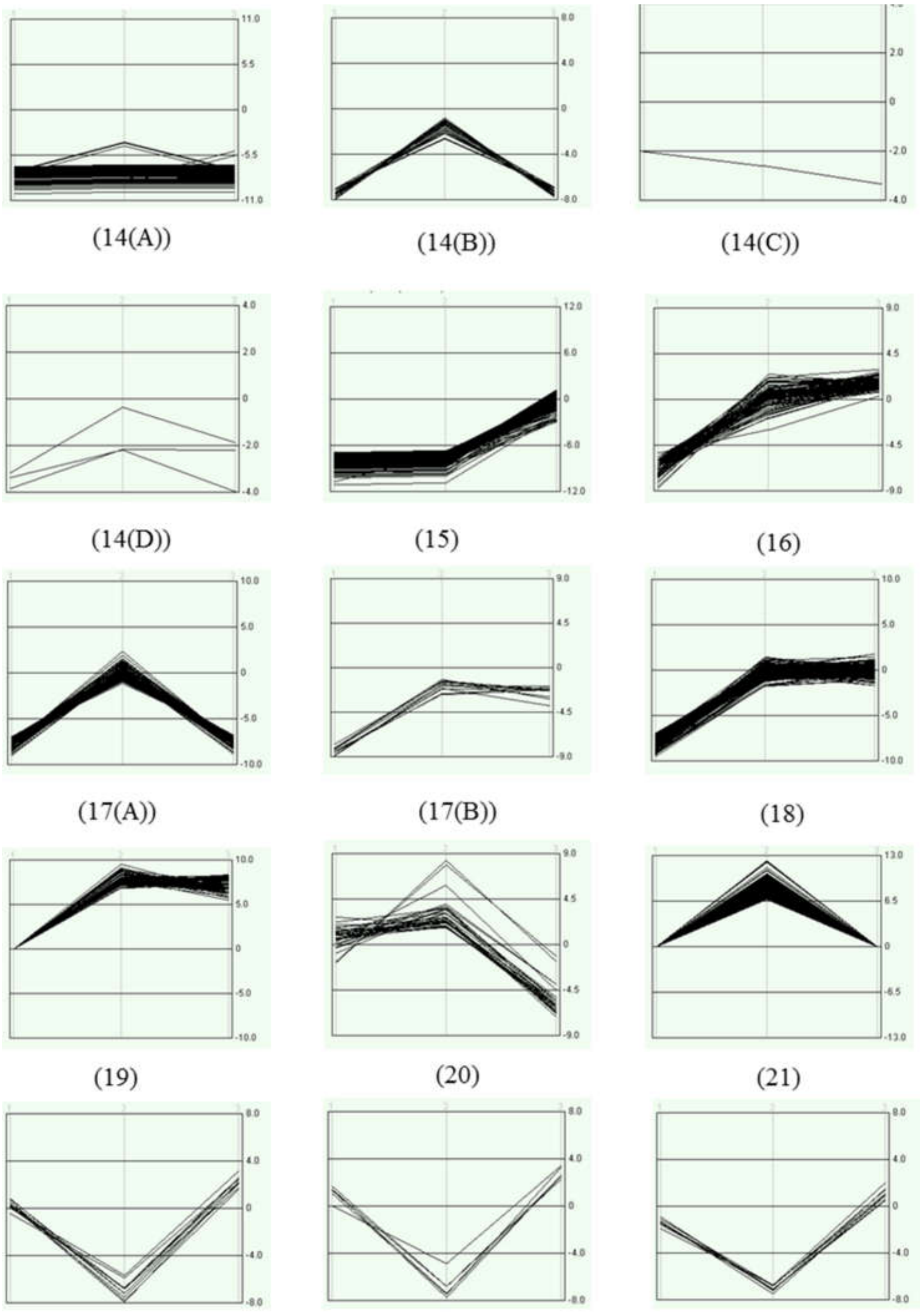
(4) (5) (6)

(7) (8) (9(A))

(9(B)) (10) (11)

(12) (13)

Fig. 3.17. The expression pattern of the corresponding regions from area 1 to area 13.



(14(A)) (14(B)) (14(C))

(14(D)) (15) (16)

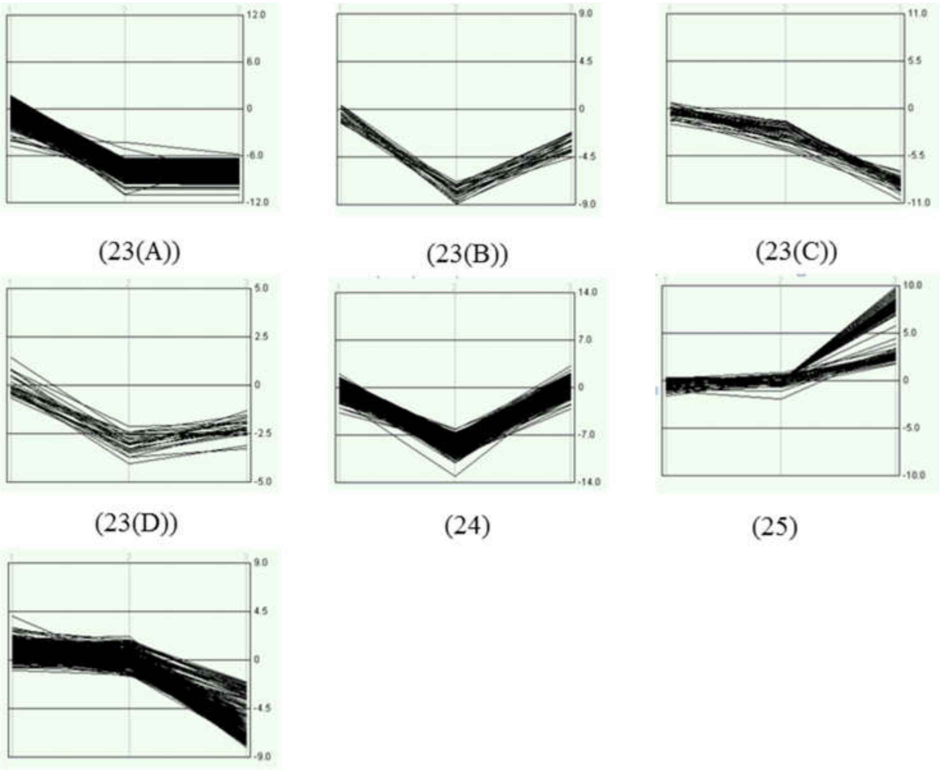
(17(A)) (17(B)) (18)

(19) (20) (21)

(22(A)) (22(B)) (22(C))

Fig. 3.18. The expression pattern of the corresponding regions from area 14 to area 22.

TF are indicated in colourful colour an



(23(A)) (23(B)) (23(C))

(23(D)) (24) (25)

(26)

Fig. 3.19. The expression pattern of the corresponding regions from area 23 to area 26.

3.5 Conclusions

In this study, I performed large-scale transcriptome sequencing of *Euglena gracilis* under anaerobic stress using the Illumina sequencing technology. A total of more than 300 million reads were generated and *de novo* assembled into 358,420 unique transcripts which were further extensively annotated by comparing their sequences to nr databases. We also performed gene expression profiling analysis and identified 518 transcription factors, 906 transcription factors, 792 transcription factors and many well-known stress responsive genes in aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase, respectively. BLSOM analyses showed that regulation changes for transcription factors and other genes. Collectively, our transcriptome sequences can provide a valuable resource for *Euglena gracilis* and research and novel insights into *E. gracilis* responses to anaerobic stress and offer candidate genes or markers that can be used to guide future efforts attempting to anaerobic tolerant *E. gracilis* cultivars.

Chapter 4: Conclusions

We utilized BLSOM to reveal the species specific key combination of oligonucleotide frequencies in each genome, which is called a “genome signature”, and TF families and metabolic pathways enzymes co-expression.

The former part of this research mainly focus on pentanucleotide compositions in 100 kb sequences derived from a wide range of vertebrate genomes and then the compositions in the human and mouse genomes in order to investigate a method for detecting differences between the closely related genomes. BLSOM can recognize the species specific key combination of oligonucleotide frequencies in each genome, which is called a “genome signature,” and the specific regions specifically enriched by transcription-factor binding sequences.

We also get a preliminary research on how the secondary metabolic pathways genes express in the major pathways analysis. Then, we performed gene expression profiling analysis and identified 518 transcription factors, 906 transcription factors, 792 transcription factors and many well-known stress responsive genes in aerobic and anaerobic condition, wild type and bleached type, midlogarithmic growth phase and late logarithmic growth phase, respectively. BLSOM analyses showed that expression regulation changes for transcription factors and other genes and ability to predict the target genes of TFs.

List of publications

Peer review journal paper

1. Yu Bai, Yuki Iwasaki, Shigehiko Kanaya, Yue Zhao, and Toshimichi Ikemura, 2014, A Novel Bioinformatics Method for Efficient Knowledge Discovery by BLSOM from Big Genomic Sequence Data, BioMed Research International, vol. 2014, Article ID 765648, 11 pages, 2014. doi:10.1155/2014/765648

Peer review international conference

1. Yu Bai, Yuki Iwasaki, Shigehiko Kanaya, Yue Zhao, Toshimichi Ikemura, Tadao Sugiura, Md. Altaf-Ui-Amin, TetsuoSato, Naoaki Ono, and Ming Huang; Systematization of genomes based on occurrence of penta-nucleotides sequences; GIW ISCB-ASIA 2014 (The 25th international Conference on Genome Informatic; 2014 December)

Reference

Abbas, O. A. (2008). Comparisons Between Data Clustering Algorithms. Int. Arab J. Inf. Technol., 5(3), 320-325.

Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T. A novel bioinformatic strategy for unveiling hidden genome signatures of eukaryotes: self-organizing map of oligonucleotide frequency. Genome Inform Ser Workshop Genome Inform., 13, 2002, 12-20.

Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T. Informatics for unveiling hidden genome signatures. Genome Res., 13, 2003, 693-702.

Abe T, Kanaya S, Uehara H, Ikemura T. A novel bioinformatics strategy for function prediction of poorly-characterized protein genes obtained from metagenome analyses. DNA Res., 16, 2009, 287-298.

Abe T, Sugawara H, Kanaya S, Ikemura T. Sequences from almost all prokaryotic, eukaryotic, and viral genomes available could be classified according to genomes on a large-scale Self-Organizing Map constructed with the Earth Simulator, Journal of the Earth Simulator 6, 2006, 17-23.

Abe T, Sugawara H, Kinouchi M, Kanaya S, Ikemura T. Novel phylogenetic studies of genomic sequence fragments derived from uncultured microbe mixtures in environmental and clinical samples. DNA Res., 12, 2005, 281-290.

Adl, Sina M., et al. The revised classification of eukaryotes. Journal of Eukaryotic Microbiology 59.5 (2012): 429-514.

Altschul, Stephen F., et al. Basic local alignment search tool. Journal of molecular biology 215.3 (1990): 403-410.

Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev., 59, 1995, 143-169.

Ariel, Federico D., et al. The true story of the HD-Zip family. Trends in plant science 12.9 (2007): 419-426.

Ashburner, Michael, et al. Gene Ontology: tool for the unification of biology. Nature genetics 25.1 (2000): 25-29.

Atteia, Ariane, et al. Anaerobic energy metabolism in unicellular photosynthetic eukaryotes. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1827.2 (2013): 210-223.

Bai, Yu, et al. A novel bioinformatics method for efficient knowledge discovery by BLSOM from big genomic sequence data. BioMed research international 2014 (2014).

Barnes, Barry, and John Dupré. Genomes and what to make of them. University of Chicago Press, 2009.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological), 289-300.

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F. The mosaic genome of warm-blooded vertebrates. Science, 228, 1985, 953-958

Bernardi G. Structural and evolutionary genomics: natural selection in genome evolution. Elsevier, Amsterdam; New York, 2004.

Bikandi, Joseba, et al. In silico analysis of complete bacterial genomes: PCR, AFLP–PCR and endonuclease restriction. Bioinformatics 20.5 (2004): 798-799.

Bolshoy A. DNA sequence analysis linguistic tools: contrast vocabularies, compositional spectra and linguistic complexity. Appl Bioinformatics., 2, 2003, 103-12.

Brunet, J. P., Tamayo, P., Golub, T. R., & Mesirov, J. P. (2004). Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences, 101(12), 4164-4169.

Byzova, Marina V., et al. Arabidopsis STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. Genes & development 13.8 (1999): 1002-1014.

Chattopadhyay, M., Dan, P. K., & MAZUMDAR, S. (2011, June). Principal component analysis and self-organizing map for visual clustering of machine-part cell formation in cellular manufacturing system. In Systems Research Forum (Vol. 5, No. 01, pp. 25-51). World Scientific Publishing Company.

Chevalier, Florian, et al. GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in Arabidopsis. Plant physiology 146.3 (2008): 1142-1154.

Chrispeels, Hanya E., et al. AtZFP1, encoding Arabidopsis thaliana C2H2 zinc-finger protein 1, is expressed downstream of photomorphogenic activation. Plant molecular biology 42.2 (2000): 279-290.

Cock, Peter JA, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25.11 (2009): 1422-1423.

Conesa, Ana, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21.18 (2005): 3674-3676.

DeLong EF. Microbial population genomics and ecology. Curr. Opin. Microbiol., 5, 2002, 520-524.

Desveaux, Darrell, Alexandre Maréchal, and Normand Brisson. Whirly transcription factors: defense gene regulation and beyond. Trends in plant science 10.2 (2005):

95-102.

Desveaux, Darrell, Alexandre Maréchal, and Normand Brisson. Whirly transcription factors: defense gene regulation and beyond. Trends in plant science 10.2 (2005): 95-102.

Devarajan, K. (2008). Nonnegative matrix factorization: an analytical and interpretive tool in computational biology. PLoS Comput Biol, 4(7), e1000029.

Du Y, Topp CN, Dawe RK. DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. PLoS Genet., 6, 2010, e1000835.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proceedings of the National Academy of Sciences, 95(25), 14863-14868.

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57-74.

Fang, Yujie, et al. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. Molecular Genetics and Genomics 280.6 (2008): 547-563.

Finkler, Aliza, Ruth Ashery-Padan, and Hillel Fromm. CAMTAs: calmodulin-binding transcription activators from plants to human. Febs Letters 581.21 (2007): 3893-3898.

Fitter, David W., et al. GLK gene pairs regulate chloroplast development in diverse plant species. The Plant Journal 31.6 (2002): 713-727.

Frantz, Christian, et al. Characterization of trans-splicing in Euglenoids. Current genetics 37.6 (2000): 349-355.

Fridborg, Ingela, et al. The Arabidopsis dwarf mutant shi exhibits reduced gibberellin responses conferred by overexpression of a new putative zinc finger protein. The plant cell 11.6 (1999): 1019-1031.

Fukagawa T, Nogami M, Yoshikawa M et al. Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat Cell Biol 6., 2004, 784 – 791.

Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic acids research 32.suppl 1 (2004): D258-D261.

Gentles AJ, Karlin S. Genome-scale compositional comparisons in eukaryotes. Genome Res., 11, 2001, 540-6.

Goodwin, Sara, John D. McPherson, and W. Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17.6 (2016): 333-351.

Grabherr, Manfred G., et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology 29.7 (2011): 644-652.

Griffiths, Simon, et al. The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant physiology 131.4 (2003): 1855-1867.

Guilfoyle, Tom J., and Gretchen Hagen. Auxin response factors. Current opinion in plant biology 10.5 (2007): 453-460.

Guojian, C., & Yao, A. (2013). Clustering Analysis of Gene Data Based on PCA and SOM Networks. Software Guide, 1, 052.

Haas, Brian J., et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8.8 (2013): 1494-1512.

Hayashi H, Abe T, Sakamoto M et al. Direct cloning of genes encoding novel xylanases from human gut, Can. J. Microbiol., 51, 2005, 251–9.

Hayashi H, Abe T, Sakamoto M, Ohara H, Ikemura T, Sakka K, Benno Y. Direct cloning of genes encoding novel xylanases from human gut. Can. J. Microbiol., 51, 2005, 251-259.

Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., ... & Saito, K. (2004). Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in Arabidopsis thaliana. Proceedings of the National academy of Sciences of the United States of America, 101(27), 10205-10210.

Hirai, M. Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D. B., Yamazaki, Y., ... & Sakurai, N. (2005). Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. Journal of Biological Chemistry, 280(27), 25590-25595.

Hoyer, P. O. (2004). Non-negative matrix factorization with sparseness constraints. Journal of machine learning research, 5(Nov), 1457-1469.

Hugenholtz P, Pace NR. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol., 14, 1996, 190-197.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Ikemura T, Aota S. Global variation in G+C content along vertebrate genome DNA: Possible correlation with chromosome band structures. J. Mol. Biol., 203, 1988, 1-13.

Ikemura T, Wada K. Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. Nucl. Acids Res., 19, 4333-4339, 1991.

Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol., 2, 1985, 13-34.

Inui H, Miyatake K, Nakano Y, Kitaoka S. Wax ester fermentation in *Euglena gracilis*.

FEBS Lett. 1982;150:89–93.

Iwasaki Y, Wada K, Wada Y, Abe T, Ikemura T. 2013. Notable clustering of transcription-factor-binding motifs in human pericentric regions and its biological significance. Chromosome Res., 21, 461-474.

Jiang, An-Long, et al. Genome-wide analysis of the C3H Zinc finger transcription factor family and drought responses of members in Aegilops tauschii. Plant Molecular Biology Reporter 32.6 (2014): 1241-1256.

Kanaya S, Kudo Y, Abe T, Okazaki T, Carlos DC and Ikemura T. Gene classification by self-organization mapping of codon usage in bacteria with completely sequenced genome. Genome Inform Ser Workshop Genome Inform., 13, 1998, 369-371.

Kanehisa, Minoru, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research 42.D1 (2014): D199-D205.

Karlin S, Campbell AM, Mrazek J. Comparative DNA analysis across diverse genomes. Annu Rev Genet., 32, 1998, 185-225.

Karlin S. Global dinucleotide signatures and analysis of genomic heterogeneity. Curr Opin Microbiol., 1, 1998, 598-610.

Kohonen T, Oja E, Simula O, Visa A and Kangas J. Engineering applications of the self-organizing map. Proceedings of the IEEE 84., 1996, 1358 - 1384.

Kohonen T. Self-organized formation of topologically correct feature maps. Biol. Cybern., 43, 1982, 59–69.

Kohonen T. The self-organizing map. Proceedings of the IEEE 78., 1990, 1464-1480.

Kumagai, Takeshi, et al. The common function of a novel subfamily of B-Box zinc finger proteins with reference to circadian-associated events in Arabidopsis thaliana. Bioscience, biotechnology, and biochemistry 72.6 (2008): 1539-1549.

Langmead, Ben, et al. Bowtie: an ultrafast memory-efficient short read aligner. Genome Biol 10.3 (2009): R25.

Läuter, J., Horn, F., Rosołowski, M., & Glimm, E. (2009). High - dimensional data analysis: Selection of variables, data compression and graphics–Application to gene expression. Biometrical journal, 51(2), 235-251.

Lee, Robert Edward. Phycology. Cambridge University Press, 2008.

Li, Zhongsen, and Terry L. Thomas. PEI1, an embryo-specific zinc finger protein gene required for heart-stage embryo formation in Arabidopsis. The Plant Cell 10.3 (1998): 383-398.

Lorenz P, Liebeton K, Niehaus F, Eck J. Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. Curr. Opin. Biotechnol., 13, 2002, 572-577.

Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. Nat Rev Mol Cell Biol., 5, 2004, 296– 304.

Maison C, Bailly D, Peters AH, et al. Higher-orderstructure in pericentromeric heterochromatin involves a distinct pattern of histone modification and an RNA component. Nat Genet., 30, 2002, 329-334.

Maison C, Bailly D, Roche D, et al. SUMOylation promotes de novo targeting of HP1α to pericentromeric heterochromatin. Nat Genet., 43, 2011, 220–227.

Margulis, Lynn. Power to the Protoctists. Slanted Truths. Springer New York, 1997. 75-82.

Martianov I, Brancorsini S, Gansmuller A, Parvinen M, Davidson I, Sassone-Corsi P. Distinct functions of TBP and TLF/TRF2 during spermatogenesis: requirement of TLF for heterochromatic chromocenter formation in haploid round spermatids. Development, 129, 2002, 945-955.

Morozova, Olena, Martin Hirst, and Marco A. Marra. Applications of new sequencing technologies for transcriptome analysis. Annual review of genomics and human genetics 10 (2009): 135-151.

Nakao R, Abe T, Nijhof AM, Yamamoto S, Jongejan F, Ikemura T, Sugimoto C. A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks, ISME J., Vol.7, 2013, 1003 – 1015.

Nielsen, Rasmus. Genomics: In search of rare human variants. Nature 467.7319 (2010): 1050-1051.

Nussinov R. Doublet frequencies in evolutionary distinct groups. Nucleic Acids Res., 12, 1984, 1749-63.

O'Neill, Ellis C., et al. The transcriptome of *Euglena gracilis* reveals unexpected metabolic capabilities for carbohydrate and natural product biochemistry. Molecular BioSystems 11.10 (2015): 2808-2820.

Ortiz, W., & Wilson, C. J. (1988). Induced changes in chloroplast protein accumulation during heat bleaching in Euglena gracilis. *Plant physiology*,*86*(2), 554-561.

Panigrahi, P. P., & Singh, T. R. (2013). Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases. Journal of theoretical biology, 334, 109-121.

Pechenik, J. A. Biology of the invertebrates. (2014).

Peng, Jiajie, et al. Extending gene ontology with gene association networks. Bioinformatics 32.8 (2016): 1185-1194.

Peng, Jiajie, et al. Measuring semantic similarities by combining gene ontology annotations and gene co-function networks. BMC bioinformatics 16.1 (2015): 1.

Phillips GJ, Arnold J, Ivarie R. Mono- through hexanucleotide composition of the Escherichia coli genome: a Markov chain analysis. Nucleic Acids Res., 15, 1987, 2611-2626.

Probst AV, Almouzni G. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. Trends Genet., 27, 2011, 192-206.

Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. Nat Rev Mol Cell Biol., 10, 2009, 192-206.

Probst AV, Okamoto I, Casanova M, Marjou FE, Baccon PL, Almouzni G. A strand-specific burst in transcription of pericentromeric satellites is required for chromocenter formation and early mouse development. Dev Cell., 19, 2010, 625-638.

Raventós, Dora, et al. HRT, a novel zinc finger, transcriptional repressor from barley. Journal of Biological Chemistry 273.36 (1998): 23313-23320.

Reiner, Anat, Daniel Yekutieli, and Yoav Benjamini. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19.3 (2003): 368-375.

Robinson, Mark D., Davis J. McCarthy, and Gordon K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26.1 (2010): 139-140.

Rocha EP, Viari A, Danchin A. Oligonucleotide bias in Bacillus subtilis: general trends and taxonomic comparisons. Nucleic Acids Res., 26, 1998, 2971-80.

Rondon MR, August PR, Bettermann AD et al. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol., 66, 2000, 2541-2547.

Russell, Anthony G., et al. Unusual features of fibrillarin cDNA and gene structure in *Euglena gracilis*: evolutionary conservation of core proteins and structural predictions for methylation-guide box C/D snoRNPs throughout the domain Eucarya. Nucleic acids research 33.9 (2005): 2781-2791.

Schauser, Leif, et al. A plant regulator controlling development of symbiotic root nodules. Nature 402.6758 (1999): 191-195.

Schloss PD, Handelsman J. Biotechnological prospects from metagenomics. Curr. Opin. Biotechnol., 14, 2003, 303-310.

Siefers, Nicholas, et al. Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. Plant physiology 149.2 (2009): 625-641.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., ... & Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proceedings of the National Academy of Sciences, 96(6), 2907-2912.

Teakle, Graham R., et al. Arabidopsis thaliana GATA factors: organisation, expression and DNA-binding characteristics. Plant molecular biology 50.1 (2002): 43-56.

Tessier, L. H., et al. Short leader sequences may be transferred from small RNAs to pre-mature mRNAs by trans-splicing in Euglena. The EMBO journal 10.9 (1991): 2621.

Uehara H, Iwasaki Y, Wada C, Ikemura T., Abe T. A novel bioinformatics strategy for searching industrially useful genome resources from metagenomic sequence libraries. Genes & Genetic Systems, 86, 2011, 53-66.

Ultsch A. Self organized feature maps for monitoring and knowledge acquisition of a chemical process. In Proc. ICANN’93, Int. Conf. on Artificial Neural Networks, edited by S Gielen, B Kappen. London: Springer, 1993, 864-867.

Van Den Heuvel, Sander, and Nicholas J. Dyson. Conserved functions of the pRB and E2F families. Nature reviews Molecular cell biology 9.9 (2008): 713-724.

van der Graaff, Eric, Thomas Laux, and Stefan A. Rensing. The WUS homeobox-containing (WOX) protein family. Genome biology 10.12 (2009): 1.

Vinga S, Almeida J. Alignment-free sequence comparison-a review. Bioinformatics, 19, 2003, 513-23.

Wagner, Günter P., Koryu Kin, and Vincent J. Lynch. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory in Biosciences 131.4 (2012): 281-285.

Wang, J., Delabie, J., Aasheim, H. C., Smeland, E., & Myklebost, O. (2002). Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. BMC bioinformatics, 3(1), 1.

Wilkinson, L., & Friendly, M. (2012). The history of the cluster heat map. The American Statistician.

Windhövel, Andrea, et al. Characterization of a novel class of plant homeodomain proteins that bind to the C4 phosphoenolpyruvate carboxylase gene of Flaveria trinervia. Plant molecular biology 45.2 (2001): 201-214.

Wirth, H., Löffler, M., von Bergen, M., & Binder, H. (2011). Expression cartography of human tissues using self organizing maps. BMC bioinformatics, 12(1), 1.

Wirth H, von Bergen M, Binder H. Mining SOM expression portraits: Feature selection and integrating concepts of molecular function[J]. BioData mining, 2012, 5(1): 1.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. Chemometrics and intelligent laboratory systems, 2(1-3), 37-52.

Wong LH, Brettingham-Moore KH, Chan L, et al. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. Genome Res., 17, 2011, 1146–1160.

Wu, Chin-Lee, et al. In vivo association of E2F and DP family proteins. Molecular and cellular biology 15.5 (1995): 2536-2546.

Xu, Yanjie, et al. Transcriptome sequencing and whole genome expression profiling of chrysanthemum under dehydration stress. Bmc Genomics 14.1 (2013): 662.

Ye, Jia, et al. WEGO: a web tool for plotting GO annotations. Nucleic acids research 34.suppl 2 (2006): W293-W297.

Yin, Yanhai, et al. A new class of transcription factors mediates brassinosteroid-regulated gene expression in Arabidopsis. Cell 120.2 (2005): 249-259.

Yoshida, Yuta, et al. De novo assembly and comparative transcriptome analysis of *Euglena gracilis* in response to anaerobic conditions. BMC genomics 17.1 (2016): 1.

Zamani, N., Russell, P., Lantz, H., Hoeppner, M. P., Meadows, J. R., Vijay, N., ... & Grabherr, M. G. (2013). Unsupervised genome-wide recognition of local relationship patterns. BMC genomics, 14(1), 1.

Zhang, He, et al. PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic acids research 39.suppl 1 (2011): D1114-D1117.

Appendix

Appendix table 1 genes name of area 1

(a) List of TF families (151 transcripts of 37 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp181858_g1_i3	B3	2.60E-10
2	comp188973_g3_i10	C3H	1.90E-07
3	comp188973_g3_i11	C3H	1.80E-07
4	comp188973_g3_i13	C3H	2.00E-07
5	comp189137_g1_i3	C3H	2.20E-51
6	comp189137_g1_i4	C3H	3.00E-51
7	comp189137_g2_i8	C3H	2.30E-51
8	comp190570_g2_i5	C3H	5.20E-21
9	comp189698_g1_i4	FAR1	5.20E-114
10	comp194101_g1_i2	FAR1	1.20E-08
11	comp194101_g1_i3	FAR1	1.10E-08
12	comp170831_g3_i3	FAR1	3.60E-34
13	comp186291_g1_i11	G2-like	6.40E-14
14	comp177576_g1_i1	GATA	4.40E-32
15	comp175324_g1_i5	GATA	6.70E-07
16	comp194735_g1_i2	GRAS	9.50E-39
17	comp194735_g1_i3	GRAS	9.60E-39
18	comp193786_g1_i2	FAR1	1.80E-55
19	comp184547_g1_i2	NF-YA	5.80E-16
20	comp184547_g2_i9	NF-YA	5.60E-16
21	comp178049_g1_i2	HB-other	1.20E-13
22	comp178174_g1_i3	MIKC	4.00E-35
23	comp163436_g1_i1	MYB_related	3.80E-26
24	comp194663_g1_i5	ARR-B	6.10E-09
25	comp179095_g1_i2	NAC	1.40E-14
26	comp180164_g1_i2	NAC	1.10E-32
27	comp190079_g2_i3	NF-YB	5.60E-21
28	comp191474_g1_i6	NF-YC	2.70E-12
29	comp188608_g1_i4	NF-YC	1.20E-12
30	comp189602_g1_i8	Nin-like	1.00E-08

31	comp187312_g1_i6	Nin-like	3.50E-78
32	comp195309_g1_i6	Nin-like	2.20E-33
33	comp162315_g1_i1	AP2	1.50E-52
34	comp162315_g1_i6	AP2	1.50E-52
35	comp175007_g2_i6	C2H2	4.30E-07
36	comp185168_g1_i4	FAR1	4.10E-26
37	comp176412_g1_i2	EIL	1.50E-09
38	comp188870_g1_i1	YABBY	4.30E-64
39	comp180808_g2_i1	YABBY	2.00E-18
40	comp180068_g1_i3	MYB_related	3.40E-37
41	comp182421_g1_i2	STAT	3.70E-07
42	comp178065_g1_i1	B3	5.00E-08
43	comp182244_g1_i2	B3	3.20E-35
44	comp182244_g1_i3	B3	4.00E-35
45	comp182244_g1_i4	B3	3.70E-35
46	comp182244_g1_i5	B3	4.60E-35
47	comp179593_g1_i1	WRKY	4.00E-19
48	comp179593_g1_i3	WRKY	4.30E-19
49	comp179740_g1_i2	YABBY	3.60E-35
50	comp179740_g1_i5	YABBY	3.30E-35
51	comp178912_g1_i1	bHLH	4.80E-15
52	comp162315_g1_i2	AP2	1.90E-52
53	comp162315_g1_i4	AP2	1.80E-52
54	comp194663_g1_i1	ARR-B	4.10E-09
55	comp194663_g1_i10	ARR-B	5.60E-09
56	comp182244_g1_i8	B3	4.20E-35
57	comp182244_g1_i10	B3	3.70E-35
58	comp189855_g1_i1	BBR-BPC	4.90E-16
59	comp187492_g1_i6	BBR-BPC	1.40E-15
60	comp176830_g1_i1	C2H2	1.40E-12
61	comp177546_g1_i3	C2H2	4.20E-09
62	comp188973_g3_i3	C3H	1.90E-07
63	comp188973_g3_i5	C3H	1.90E-07
64	comp188973_g3_i6	C3H	1.90E-07
65	comp188973_g3_i8	C3H	2.00E-07

66	comp188973_g3_i12	C3H	1.90E-07
67	comp190570_g2_i4	C3H	4.20E-21
68	comp169510_g1_i1	E2F_DP	3.10E-08
69	comp169510_g1_i3	E2F_DP	3.70E-08
70	comp174717_g2_i2	E2F_DP	1.70E-08
71	comp184653_g1_i2	EIL	4.10E-21
72	comp175907_g1_i3	ERF	5.40E-09
73	comp185168_g1_i2	FAR1	5.20E-26
74	comp172018_g1_i2	FAR1	1.30E-14
75	comp173266_g1_i2	FAR1	5.40E-41
76	comp175150_g1_i1	FAR1	1.50E-58
77	comp175324_g1_i2	GATA	6.40E-07
78	comp175324_g1_i6	GATA	6.70E-07
79	comp192585_g1_i9	GRAS	2.90E-09
80	comp174585_g2_i3	GRAS	3.80E-18
81	comp174585_g2_i5	GRAS	4.30E-18
82	comp190700_g2_i8	GeBP	2.80E-09
83	comp195585_g1_i7	GeBP	2.10E-07
84	comp186003_g1_i2	LBD	7.90E-24
85	comp187693_g1_i3	LBD	8.30E-146
86	comp187750_g1_i5	M-type	4.50E-23
87	comp181145_g1_i3	MYB	1.30E-132
88	comp182710_g1_i3	MYB	1.20E-09
89	comp166665_g1_i3	MYB_related	4.60E-12
90	comp163565_g1_i1	MYB_related	4.80E-07
91	comp168201_g1_i2	MYB_related	1.70E-40
92	comp180068_g1_i5	MYB_related	4.00E-37
93	comp181387_g1_i1	NAC	1.10E-15
94	comp178688_g1_i2	NAC	1.90E-22
95	comp180908_g1_i3	NAC	2.90E-28
96	comp182293_g3_i9	NF-YA	1.80E-13
97	comp184908_g1_i1	NF-YA	1.10E-36
98	comp186468_g1_i3	NF-YA	8.60E-37
99	comp192871_g1_i2	TCP	1.30E-08
100	comp178141_g1_i1	WOX	6.90E-33

101	comp181755_g1_i1	WOX	9.60E-63
102	comp179009_g2_i4	WRKY	9.50E-84
103	comp179009_g2_i8	WRKY	7.90E-84
104	comp179186_g1_i1	WRKY	1.10E-18
105	comp179637_g1_i1	YABBY	3.10E-20
106	comp179431_g1_i1	YABBY	5.20E-19
107	comp190197_g3_i7	YABBY	7.80E-07
108	comp179631_g3_i1	bHLH	1.70E-13
109	comp180387_g1_i1	bHLH	3.40E-23
110	comp180387_g1_i2	bHLH	4.70E-23
111	comp185754_g1_i1	bZIP	4.90E-09
112	comp185754_g1_i12	bZIP	5.00E-09
113	comp180548_g1_i2	B3	1.90E-10
114	comp188973_g3_i9	C3H	2.00E-07
115	comp189137_g1_i10	C3H	2.40E-51
116	comp189137_g2_i9	C3H	2.10E-51
117	comp180042_g1_i1	Dof	1.30E-08
118	comp189698_g1_i6	FAR1	2.00E-114
119	comp186009_g4_i2	FAR1	1.10E-08
120	comp174843_g1_i4	FAR1	3.30E-34
121	comp194799_g1_i7	GRAS	8.80E-39
122	comp180515_g3_i1	NAC	1.80E-41
123	comp190835_g1_i1	M-type	6.00E-11
124	comp182710_g1_i4	MYB	1.10E-09
125	comp180388_g1_i1	MYB_related	3.90E-13
126	comp178688_g1_i4	NAC	2.20E-22
127	comp180481_g1_i3	NAC	1.30E-33
128	comp187817_g1_i2	Nin-like	4.10E-07
129	comp187817_g1_i3	Nin-like	5.10E-07
130	comp195309_g1_i2	Nin-like	1.70E-33
131	comp176412_g1_i3	EIL	1.50E-09
132	comp189100_g1_i2	YABBY	1.20E-20
133	comp180626_g1_i7	MYB	4.50E-136
134	comp180143_g1_i4	MYB_related	4.20E-37
135	comp182842_g1_i9	TALE	1.20E-10

136	comp182842_g1_i11	TALE	9.60E-11
137	comp180981_g1_i4	STAT	4.80E-15
138	comp183284_g1_i3	VOZ	6.40E-13
139	comp181755_g1_i2	WOX	1.40E-64
140	comp179009_g2_i7	WRKY	2.10E-93
141	comp179593_g1_i2	WRKY	4.20E-19
142	comp179593_g1_i4	WRKY	4.30E-19
143	comp175892_g3_i1	Whirly	7.30E-08
144	comp189823_g1_i3	YABBY	1.40E-22
145	comp189762_g1_i1	YABBY	5.80E-28
146	comp179529_g1_i1	bHLH	2.40E-31
147	comp179443_g1_i1	bHLH	7.80E-25
148	comp178347_g1_i2	bHLH	2.10E-45
149	comp179184_g2_i5	bHLH	5.80E-16
150	comp180387_g1_i3	bHLH	3.20E-23
151	comp180387_g2_i4	bHLH	3.10E-23

(b) List of other genes names (579 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp101701_g1_i1	RS24_DICDI	3.50E-25
2	comp102129_g1_i2	ADP-ribosyltransferase 1 precursor	8.90E-29
3	comp102495_g1_i1	MIFH_TRITR	2.30E-11
4	comp105272_g1_i2	EFTU_EUGGR	0.00E+00
5	comp108746_g1_i2	Golgi SNAP receptor complex member 2	2.40E-09
6	comp109717_g1_i1	hypothetical protein DQ04_09981010	3.60E-13
7	comp109800_g1_i1	RS24_DICDI	6.30E-25
8	comp11357_g1_i1	AAEL017413- partial	6.40E-08
9	comp115089_g1_i1	KAT_DICDI	1.60E-81
10	comp115089_g1_i2	kynurenine--oxoglutarate transaminase 1 isoform X1	3.30E-123
11	comp116378_g1_i2	EFTU_EUGGR	0.00E+00
12	comp116569_g1_i1	ZRAB2_CHICK	1.70E-09
13	comp116744_g1_i1	hypothetical protein Ctob_007964	9.30E-17
14	comp117008_g1_i1	TRAF6_HUMAN	2.60E-15
15	comp122456_g2_i2	PREDICTED: uncharacterized protein LOC18444466	1.20E-07

16	comp122705_g1_i1	YL477_MIMIV	1.10E-31
17	comp126696_g1_i2	hypothetical protein T9A_01030	3.70E-47
18	comp127343_g1_i2	CTR1_SOLIN	3.80E-51
19	comp127814_g1_i1	COX2_LASSP	1.50E-77
20	comp127851_g1_i3	VIS_VIBS1	1.70E-07
21	comp127935_g1_i2	PSBA_EUGGR	0.00E+00
22	comp127935_g1_i3	PSBA_EUGGR	1.60E-31
23	comp128383_g1_i1	NLRC3_HUMAN	7.90E-40
24	comp129541_g1_i1	HOS4_YEAST	1.90E-07
25	comp130210_g1_i1	CHLI_EUGGR	0.00E+00
26	comp130265_g1_i2	DNA-directed RNA polymerase II subunit RPB7	3.40E-48
27	comp133101_g1_i1	ATPA_EUGGR	0.00E+00
28	comp133109_g1_i2	YIW2_SCHPO	3.30E-42
29	comp134569_g1_i1	Y1181_ARATH	1.10E-31
30	comp134569_g1_i2	Y1181_ARATH	1.50E-31
31	comp134876_g1_i4	hypothetical protein TRSC58_03537	7.40E-07
32	comp135084_g1_i1	CHLI_EUGGR	0.00E+00
33	comp135330_g1_i1	hypothetical protein GUITHDRAFT_161344, partial	8.60E-13
34	comp138222_g1_i1	CTR1_SOLIN	1.40E-61
35	comp138431_g1_i1	PSBB_EUGGR	0.00E+00
36	comp138431_g1_i2	PSBB_EUGGR	0.00E+00
37	comp138431_g1_i3	PSBB_EUGGR	9.30E-135
38	comp138470_g1_i2	WDR92_MOUSE	2.50E-135
39	comp140221_g1_i2	RR4_EUGGR	2.00E-132
40	comp140221_g2_i1	RR4_EUGGR	2.30E-132
41	comp140373_g1_i1	PSBK_EUGGR	2.50E-15
42	comp140387_g1_i2	WD40 repeat-containing	4.90E-08
43	comp140574_g1_i2	bifunctional -like	1.70E-35
44	comp140734_g1_i1	RENT1_CAEEL	3.40E-25
45	comp140961_g1_i1	ATP synthase F0 subunit 6 (mitochondrion)	1.30E-18
46	comp140961_g2_i1	ATP synthase F0 subunit 6 (mitochondrion)	1.80E-12
47	comp141180_g1_i3	CSPA_STIAD	1.00E-18
48	comp141217_g1_i1	uv radiation resistance-associated	6.00E-16
49	comp141780_g1_i2	PX24D_DICDI	1.50E-09
50	comp141931_g1_i1	PSAC_EUGGR	1.10E-33

51	comp142258_g1_i1	DLRB2_MOUSE	5.70E-39
52	comp142531_g1_i1	4-isomerase	3.80E-45
53	comp142531_g2_i1	4-isomerase	3.90E-45
54	comp142981_g1_i1	mitochondrial carrier family	8.50E-37
55	comp142981_g1_i3	PNC1_ARATH	6.90E-25
56	comp143008_g1_i1	BnaC09g29270D	7.20E-42
57	comp143128_g1_i2	hypothetical protein TRSC58_03537	7.40E-07
58	comp143128_g1_i3	hypothetical protein TRSC58_03537	7.10E-07
59	comp143128_g1_i4	unnamed protein product	2.30E-08
60	comp143128_g1_i5	hypothetical protein TRSC58_03537	7.00E-07
61	comp143235_g1_i1	TRAF6_HUMAN	2.60E-15
62	comp143255_g2_i1	RL34A_YEAST	1.20E-19
63	comp143474_g1_i1	RL34A_YEAST	1.90E-20
64	comp143492_g1_i1	hypothetical protein	4.10E-16
65	comp143711_g1_i2	ANT1_YEAST	1.30E-21
66	comp143978_g1_i1	OTU domain-containing 7B	3.50E-30
67	comp144079_g1_i1	rubrerythrin	4.40E-27
68	comp144297_g1_i1	ILV5_SCHPO	4.60E-99
69	comp144388_g1_i2	ANK2_HUMAN	2.70E-17
70	comp144388_g1_i4	ANK2_HUMAN	4.60E-18
71	comp144519_g1_i1	YCX1_EUGGR	1.50E-169
72	comp144924_g1_i1	hypothetical protein NAEGRDRAFT_58984	3.00E-09
73	comp145585_g1_i2	NLRC3_HUMAN	4.10E-34
74	comp145907_g1_i1	ILV5_SCHPO	7.90E-100
75	comp146107_g1_i3	dihydrolipoamide dehydrogenase	6.10E-52
76	comp146407_g1_i1	fungal specific transcription	5.30E-13
77	comp146534_g3_i2	ATPI_EUGGR	2.30E-149
78	comp146646_g1_i3	Uncharacterized conserved	5.10E-19
79	comp146646_g1_i4	Uncharacterized conserved	5.40E-19
80	comp146874_g1_i1	CCL11_ORYSJ	1.00E-31
81	comp146874_g1_i2	CCL11_ORYSJ	8.00E-32
82	comp147074_g1_i1	hypothetical protein VOLCADRAFT_100878	7.40E-60
83	comp147074_g1_i2	hypothetical protein VOLCADRAFT_100878	6.40E-60
84	comp147265_g1_i1	beta-type IP39	7.90E-20
85	comp147297_g1_i1	fungal specific transcription	5.30E-13

86	comp147834_g1_i1	ZNHI1_HUMAN	1.00E-16
87	comp148135_g1_i1	RENT1_CAEEL	3.60E-25
88	comp148201_g1_i1	mitochondrial tryptophan rich sensory 1	7.90E-09
89	comp148285_g1_i1	RNZ1_HUMAN	1.30E-56
90	comp148315_g1_i1	NLRC3_HUMAN	2.50E-27
91	comp148589_g1_i1	DENR_MAGO7	4.50E-25
92	comp148809_g1_i2	DNA endonuclease RBBP8	7.40E-13
93	comp148809_g1_i4	DNA endonuclease RBBP8	1.40E-12
94	comp149134_g1_i1	adenylate cyclase	4.00E-15
95	comp149282_g1_i1	Y108_SYNY3	4.00E-10
96	comp149348_g1_i1	hypothetical protein GUITHDRAFT_166518	4.60E-24
97	comp149582_g1_i2	RIBX_MIMIV	1.60E-47
98	comp149586_g1_i1	AB1F_ARATH	0.00E+00
99	comp149586_g1_i4	AB1F_ARATH	0.00E+00
100	comp149586_g1_i5	AB1F_ARATH	0.00E+00
101	comp149586_g1_i6	ABC transporter	0.00E+00
102	comp149588_g1_i1	TXN4A_HUMAN	7.30E-80
103	comp149630_g1_i2	YHCV_BACSU	3.60E-15
104	comp149630_g2_i2	YHCV_BACSU	4.70E-15
105	comp149648_g1_i2	phosphatase 2A regulatory subunit PR55	1.90E-143
106	comp149858_g1_i1	AZOR_DECAR	1.20E-23
107	comp149921_g1_i2	LYRM4_TAEGU	7.50E-18
108	comp149989_g1_i1	PLDA1_TOBAC	4.70E-168
109	comp150070_g1_i2	KAP0_HUMAN	2.10E-09
110	comp150091_g1_i1	CYSG_NITEC	8.10E-94
111	comp150234_g1_i3	COM1_MOUSE	1.50E-11
112	comp150399_g1_i3	YCX9_EUGGR	0.00E+00
113	comp150399_g1_i4	YCX9_EUGGR	0.00E+00
114	comp150399_g1_i5	YCX9_EUGGR	0.00E+00
115	comp150423_g1_i1	RBL_EUGGR	0.00E+00
116	comp150527_g1_i1	PYRB_THEMA	7.40E-150
117	comp150532_g1_i1	S35B1_DICDI	5.20E-27
118	comp150581_g1_i1	beta-type IP39	9.20E-20
119	comp150619_g1_i1	REEP6_DANRE	4.90E-11
120	comp150691_g1_i2	riken cdna 2700097o09	5.00E-55

121	comp151007_g1_i3	TRX1_YEAST	3.60E-27
122	comp151206_g1_i1	PSBB_EUGGR	0.00E+00
123	comp151206_g1_i2	PSBB_EUGGR	0.00E+00
124	comp151206_g1_i3	PSBB_EUGGR	1.40E-135
125	comp151533_g1_i1	YRHG_BACSU	1.10E-19
126	comp151721_g1_i1	WD40 repeat-containing ,	7.00E-08
127	comp151802_g1_i2	transmembrane 56 isoform X1	6.80E-21
128	comp151977_g1_i1	chromatin assembly factor 1 subunit FAS2	2.00E-68
129	comp152060_g1_i3	NLRC3_HUMAN	3.60E-34
130	comp152092_g1_i1	Y108_SYNY3	4.00E-10
131	comp152118_g1_i3	YAS2_SCHPO	7.10E-15
132	comp152293_g1_i2	MMP10_RAT	1.10E-16
133	comp152332_g1_i1	AZOR_AROAE	1.20E-23
134	comp152332_g1_i2	AZOR_DECAR	2.10E-67
135	comp152441_g2_i1	NAD-dependent alcohol dehydrogenase	1.70E-09
136	comp152452_g1_i1	RH5_ORYSJ	4.40E-136
137	comp152590_g1_i3	ASAH1_MACFA	1.40E-30
138	comp152869_g1_i4	TRX1_YEAST	5.70E-26
139	comp152886_g1_i1	mitochondrial tryptophan rich sensory 1	8.30E-09
140	comp152886_g1_i2	mitochondrial tryptophan rich sensory 1	1.90E-08
141	comp153060_g1_i2	#NAME?	2.00E-125
142	comp153060_g1_i5	#NAME?	3.10E-125
143	comp153204_g1_i2	hypothetical protein Naga_100030g15	2.00E-09
144	comp153236_g1_i1	hypothetical protein PBRA_008002	7.60E-58
145	comp153239_g1_i2	UFAA1_MYCTU	2.60E-50
146	comp153254_g1_i2	DDX20_HUMAN	4.00E-92
147	comp153362_g1_i3	VIS_VIBS1	2.00E-07
148	comp153383_g1_i1	rubrerythrin	5.20E-27
149	comp153481_g1_i2	RIBX_SHIFL	4.10E-19
150	comp153483_g1_i3	hypothetical protein EMIHUDRAFT_247554	2.60E-36
151	comp153567_g1_i3	hypothetical protein GUITHDRAFT_166518	1.10E-23
152	comp153597_g1_i1	phospholipase	7.00E-33
153	comp153597_g1_i2	phospholipase	4.40E-33
154	comp153609_g1_i1	palmitoyltransferase ZDHHC15 isoform X2	1.80E-28
155	comp153609_g1_i2	palmitoyltransferase ZDHHC15 isoform X2	3.90E-28

156	comp153856_g1_i3	hypothetical protein CHLNCDRAFT_145050	5.30E-58
157	comp153984_g1_i1	zinc finger with UFM1-specific peptidase domain	3.60E-53
158	comp153984_g1_i2	ZUFSP_BOVIN	6.00E-35
159	comp154015_g1_i1	kinase domain	6.50E-25
160	comp154037_g1_i4	OGFR_MOUSE	3.90E-36
161	comp154198_g1_i1	Hypothetical protein, putative	1.40E-09
162	comp154198_g1_i3	Hypothetical protein, putative	2.50E-09
163	comp154359_g1_i1	NLRC3_HUMAN	2.50E-45
164	comp154563_g1_i2	hypothetical protein GUITHDRAFT_109974	6.30E-46
165	comp154614_g1_i1	RPGR_CANFA	9.00E-19
166	comp154777_g1_i1	TGS1_MYCTO	7.80E-12
167	comp154783_g1_i1	OGFR_MOUSE	8.80E-36
168	comp154783_g1_i3	OGFR_MOUSE	5.10E-36
169	comp154819_g1_i1	Zinc finger Ran-binding domain-containing 2	1.70E-09
170	comp154886_g1_i3	#NAME?	1.40E-125
171	comp154886_g1_i4	#NAME?	1.20E-125
172	comp154886_g1_i5	#NAME?	5.00E-125
173	comp154889_g1_i1	YRHG_BACSU	1.00E-19
174	comp155077_g1_i4	hypothetical protein CHLNCDRAFT_145050	5.90E-59
175	comp155077_g1_i5	hypothetical protein CHLNCDRAFT_145050	8.30E-59
176	comp155100_g1_i1	TL19_ARATH	9.30E-21
177	comp155211_g1_i2	KLHL4_HUMAN	1.30E-27
178	comp155230_g1_i1	7-cyano-7-deazaguanine synthase	2.20E-29
179	comp155291_g1_i2	tetratricopeptide repeat 1	1.20E-23
180	comp155310_g1_i1	CBS domain containing	1.30E-18
181	comp155310_g1_i2	CBS domain containing	1.10E-18
182	comp155365_g1_i4	rho GDP-dissociation inhibitor 1-like	4.10E-09
183	comp155365_g1_i5	rho GDP-dissociation inhibitor 1-like	3.60E-09
184	comp155380_g1_i2	ADT3_BOVIN	8.20E-89
185	comp155420_g1_i1	NLRC3_HUMAN	3.20E-40
186	comp155515_g1_i1	glycerol-3-phosphate dehydrogenase 1	4.60E-123
187	comp155556_g1_i4	NLRC3_HUMAN	1.30E-49
188	comp155556_g1_i5	NLRC3_HUMAN	2.60E-45
189	comp155632_g1_i2	NEK1_HUMAN	4.30E-50
190	comp155692_g1_i1	amino acid transporter,	3.20E-51

191	comp155697_g1_i1	YCX91_PHAAORecName: Full=Uncharacterized protein ORF91 (chloroplast)	1.40E-18
192	comp155736_g1_i2	TPC6B_HUMAN	5.90E-22
193	comp155782_g1_i3	FIS1_DEBHA	2.20E-08
194	comp155796_g1_i4	Mitochondrial transcription termination factor family	6.30E-09
195	comp155993_g1_i5	predicted protein	1.40E-76
196	comp156303_g1_i2	malate synthase-isocitrate lyase	1.60E-45
197	comp156342_g1_i1	STX6_CHICK	8.50E-07
198	comp156345_g1_i2	NLRC3_HUMAN	7.50E-25
199	comp156363_g1_i2	hypothetical protein GUITHDRAFT_166346	5.80E-73
200	comp156380_g1_i2	MFS transporter, FLVCR family, disrupted in renal carcinoma 2	1.20E-58
201	comp156404_g1_i2	YKT61_ARATH	1.10E-44
202	comp156422_g1_i2	RBL_EUGGR	0.00E+00
203	comp156488_g1_i3	YAS2_SCHPO	7.30E-15
204	comp156508_g1_i1	YCX8_EUGGR	3.70E-54
205	comp156508_g1_i3	YCX8_EUGGR	1.10E-53
206	comp156508_g2_i1	YCXA_EUGGR	0.00E+00
207	comp156508_g4_i1	YCXA_EUGGR	0.00E+00
208	comp156508_g6_i1	YCX7_EUGGR	1.30E-20
209	comp156557_g1_i1	F151B_HUMAN	1.30E-21
210	comp156649_g1_i4	ATE1_ARATH	4.00E-43
211	comp156684_g1_i3	TSSC1_DANRE	4.80E-08
212	comp156743_g1_i2	dynein light chain	1.70E-80
213	comp156890_g1_i1	LUL3_ARATH	1.60E-22
214	comp156890_g1_i2	LUL3_ARATH	1.70E-22
215	comp156941_g1_i8	DEAD DEAH box helicase	2.90E-30
216	comp157038_g1_i1	tartrate-resistant acid phosphatase type 5	1.20E-53
217	comp157038_g1_i2	PPA5_RABIT	8.10E-08
218	comp157059_g1_i2	PUR7_BACFN	1.20E-62
219	comp157088_g1_i2	Y182_METEA	1.50E-17
220	comp157089_g1_i2	ANR54_RAT	1.90E-10
221	comp157136_g1_i1	transmembrane 128-like	3.00E-08
222	comp157182_g1_i1	hypothetical protein PBRA_008009	6.40E-18
223	comp157236_g1_i2	ITPA_SORBI	1.80E-86
224	comp157268_g1_i2	RMGL_PICST	7.40E-22
225	comp157268_g1_i3	RMGL_PICST	8.20E-22

226	comp157300_g1_i1	CYC6_EUGVI	3.00E-32
227	comp157300_g1_i2	CYC6_EUGVI	3.90E-32
228	comp157403_g1_i5	CYA1_RHIME	1.40E-21
229	comp157432_g1_i2	OTUD3_MOUSE	1.10E-31
230	comp157487_g1_i2	cyclin K	6.40E-12
231	comp157565_g1_i1	GST1_SCHPO	2.10E-08
232	comp157569_g1_i1	kelch repeat ,	2.30E-25
233	comp157624_g1_i2	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	5.90E-105
234	comp157799_g1_i1	ATG26_YEAS7	2.00E-13
235	comp157913_g1_i3	SSM4 like ring finger	2.30E-11
236	comp157913_g1_i4	SSM4 like ring finger	1.80E-11
237	comp158051_g1_i1	CYTB_OPLFA	7.70E-18
238	comp158064_g1_i2	inner membrane PPF-1, chloroplastic	6.10E-42
239	comp158144_g1_i2	MCES1_ORYSJ	1.50E-39
240	comp158183_g1_i1	ODC_DICDI	4.00E-65
241	comp158183_g1_i3	ODC_DICDI	3.00E-65
242	comp158289_g1_i1	pterin-4-alpha-carbinolamine dehydratase	3.00E-17
243	comp158289_g1_i3	pterin-4-alpha-carbinolamine dehydratase	3.10E-17
244	comp158481_g1_i1	METL4_MOUSE	3.80E-21
245	comp158481_g1_i3	METL4_MOUSE	7.30E-21
246	comp158577_g1_i5	zinc finger with UFM1-specific peptidase domain	7.40E-53
247	comp158577_g1_i6	zinc finger with UFM1-specific peptidase domain	9.60E-53
248	comp158590_g1_i2	LENG8_XENLA	3.40E-43
249	comp158793_g2_i2	hypothetical protein CAOG_03935	1.40E-15
250	comp158810_g1_i2	hydroxyacylglutathione hydrolase	2.70E-52
251	comp158867_g2_i1	F26_ARATH	3.60E-66
252	comp158867_g2_i2	F26_ARATH	3.50E-66
253	comp158867_g2_i3	F26_ARATH	3.60E-66
254	comp158867_g3_i3	F26_ARATH	1.30E-66
255	comp158868_g1_i3	MBF1A_ARATH	3.00E-39
256	comp158874_g1_i2	bacterial leucyl aminopeptidase	1.20E-07
257	comp158911_g1_i2	KSHB_MYCS2	3.90E-19
258	comp158945_g1_i3	mitochondrial transcription termination factor family	7.40E-10
259	comp159029_g1_i2	constitutive coactivator of peroxisome proliferator-activated receptor gamma	3.00E-09
260	comp159075_g1_i2	s-adenosylmethionine mitochondrial carrier	1.90E-15

261	comp159124_g1_i2	AOX4_ARATH	1.30E-64
262	comp159124_g1_i3	plastid terminal oxidase	1.60E-92
263	comp159148_g1_i2	pterin-4-alpha-carbinolamine dehydratase	3.70E-17
264	comp159163_g1_i1	AIR9_ARATH	3.80E-14
265	comp159201_g1_i4	KAP0_HUMAN	3.60E-09
266	comp159204_g1_i2	AB17A_RAT	1.40E-18
267	comp159221_g1_i2	6-pyruvoyl tetrahydrobiopterin synthase	7.00E-32
268	comp159242_g1_i2	major facilitator superfamily	4.20E-90
269	comp159286_g1_i1	predicted protein	1.10E-08
270	comp159304_g1_i2	Adenylosuccinate lyase	3.00E-10
271	comp159340_g1_i2	hydroxyacylglutathione hydrolase	1.80E-52
272	comp159355_g1_i2	RHBGB_XENLA	6.80E-79
273	comp159355_g1_i6	ammonium transporter Rh type A	2.10E-92
274	comp159360_g1_i3	AB1F_ARATH	0.00E+00
275	comp159485_g1_i2	DB10_NICSY	6.60E-53
276	comp159493_g1_i1	COX1_CHOPI	4.80E-148
277	comp159493_g1_i4	COX1_CHOFU	1.30E-137
278	comp159493_g1_i5	COX1_CHOFU	6.90E-126
279	comp159559_g1_i2	vacuolar sorting-associated 22 homolog 1	2.60E-52
280	comp159621_g1_i1	HTSF1_HUMAN	3.00E-44
281	comp159676_g1_i1	ubiquitin ligase	4.90E-16
282	comp159676_g1_i4	ubiquitin ligase	4.90E-16
283	comp159680_g1_i2	containing DUF1963	9.80E-28
284	comp159709_g1_i1	exostosin	1.30E-23
285	comp159812_g1_i2	MCES1_ORYSJ	2.20E-28
286	comp159846_g1_i2	ATE1_ARATH	6.50E-45
287	comp159854_g1_i1	hypothetical protein GLOINDRAFT_312092	2.30E-19
288	comp159854_g1_i2	hypothetical protein GLOINDRAFT_312092	5.00E-19
289	comp159950_g1_i2	AK1_DICDI	1.30E-44
290	comp160021_g1_i2	plastid terminal oxidase	1.50E-92
291	comp160024_g1_i1	ferredoxin-NADP+ reductase	2.10E-94
292	comp160043_g1_i4	hypothetical protein F443_00246	9.90E-09
293	comp160061_g1_i1	polyketide cyclase	4.90E-26
294	comp160156_g1_i1	membrane-bound adenylyl cyclase	4.70E-76
295	comp160179_g1_i2	procollagen galactosyltransferase 1	4.30E-15

296	comp160211_g1_i4	pseudouridine synthase, family	1.00E-47
297	comp160278_g1_i1	SIN3A_MOUSE	1.10E-35
298	comp160278_g1_i3	paired amphipathic helix Sin3a isoform X1	2.50E-48
299	comp160291_g1_i2	SGT1A_ARATH	2.90E-40
300	comp160355_g1_i1	endo-1,3-beta-glucanase	6.20E-43
301	comp160438_g1_i2	AZOR_AROAE	2.40E-23
302	comp160438_g1_i3	AZOR_DECAR	2.20E-67
303	comp160506_g1_i1	Ribosomal S21	1.20E-15
304	comp160609_g1_i3	NSMA2_MOUSE	4.60E-08
305	comp160629_g1_i1	ubiquitin family	2.40E-07
306	comp160629_g1_i2	ubiquitin family	2.20E-07
307	comp160660_g1_i1	DB10_NICSY	6.60E-53
308	comp160801_g1_i1	RTPR_EUGGR	1.20E-18
309	comp160863_g1_i1	Y1017_SYNY3	3.00E-12
310	comp160864_g1_i4	SAMC1_ARATH	1.20E-10
311	comp161092_g1_i1	SETD3_MOUSE	2.10E-07
312	comp161095_g1_i2	PHO88_YEAST	8.00E-13
313	comp161120_g1_i1	e3 ubiquitin- ligase	4.80E-16
314	comp161181_g1_i2	CYSKM_ARATH	1.40E-78
315	comp161194_g1_i1	FACT complex subunit SSRP1-like	1.20E-44
316	comp161194_g1_i2	SSRP1_XENLA	8.10E-45
317	comp161194_g1_i3	SSRP1_XENLA	1.10E-44
318	comp161194_g1_i4	FACT complex subunit SSRP1-like	1.10E-44
319	comp161194_g1_i6	SSRP1_XENLA	9.40E-45
320	comp161199_g1_i2	WSDU1_DANRE	3.10E-09
321	comp161211_g1_i2	MCE_FOWPN	9.30E-13
322	comp161280_g1_i1	hypothetical protein F443_00243	3.00E-11
323	comp161295_g2_i2	BACD3_HUMAN	5.20E-15
324	comp161345_g1_i1	PCS3_LOTJA	1.50E-67
325	comp161345_g1_i3	PCS3_LOTJA	7.60E-68
326	comp161352_g1_i3	predicted protein	3.00E-08
327	comp161419_g1_i1	ferredoxin-NADP+ reductase	2.20E-94
328	comp161429_g1_i1	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	2.60E-53
329	comp161429_g1_i3	LPAT4_ARATH	9.60E-40
330	comp161429_g1_i5	LPAT4_ARATH	1.50E-39

331	comp161429_g1_i6	LPAT4_ARATH	1.20E-39
332	comp161443_g1_i1	dihydropyrimidinase	0.00E+00
333	comp161524_g3_i1	ATPB_EUGGR	0.00E+00
334	comp161573_g1_i1	HIR3_ARATH	5.80E-76
335	comp161583_g1_i1	ZRAB2_PONAB	4.80E-10
336	comp161585_g1_i3	TBCA_ARATH	7.00E-08
337	comp161603_g1_i1	RHBGB_XENLA	8.80E-79
338	comp161632_g1_i1	PH domain-containing	1.40E-19
339	comp161697_g1_i1	APM1_ARATH	1.40E-120
340	comp161699_g1_i1	CFA58_CHLRE	1.80E-46
341	comp161708_g1_i2	branch point binding	5.20E-30
342	comp161719_g1_i2	CYC6_EUGVI	3.40E-32
343	comp161732_g1_i1	GORS1_RAT	7.60E-31
344	comp161776_g1_i1	N2B_HAEIR	4.30E-75
345	comp161863_g1_i3	hypothetical protein PRUPE_ppa021045mg	6.20E-08
346	comp161874_g1_i2	GST1_SCHPO	2.10E-08
347	comp161874_g1_i3	GST1_SCHPO	2.10E-08
348	comp161903_g1_i1	calcium homeostasis endoplasmic reticulum	2.50E-11
349	comp161920_g1_i1	glycosyl family 43	3.00E-50
350	comp161920_g1_i2	glycosyl family 43	2.40E-50
351	comp161932_g2_i1	hypothetical protein SELMODRAFT_54741, partial	1.50E-20
352	comp161932_g2_i2	hypothetical protein SELMODRAFT_102324	1.70E-15
353	comp161981_g1_i1	endo-1,3-beta-glucanase	1.20E-07
354	comp162052_g1_i2	PIN1_ARATH	1.40E-37
355	comp162090_g1_i1	PIN1_ARATH	2.80E-37
356	comp162097_g1_i1	LPSA_DICNO	1.20E-27
357	comp162101_g1_i1	hypothetical protein	2.00E-12
358	comp162110_g2_i3	alpha-type IP39	1.00E-131
359	comp162141_g2_i2	AGAL1_PEDPE	4.30E-111
360	comp162141_g2_i3	AGALG_ASPTN	8.80E-93
361	comp162171_g1_i1	SPEF1_XENLA	1.20E-34
362	comp162177_g1_i1	LPAT4_ARATH	2.70E-39
363	comp162177_g1_i3	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	9.60E-54
364	comp162177_g1_i4	LPAT4_ARATH	1.90E-39
365	comp162177_g1_i5	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	1.40E-53

366	comp162236_g1_i2	hypothetical protein GUITHDRAFT_149743	3.50E-20
367	comp162292_g1_i4	domain containing	2.70E-107
368	comp162363_g3_i2	VPS39_HUMAN	2.30E-57
369	comp162368_g1_i1	glycosyl transferase	9.70E-93
370	comp162368_g1_i2	glycosyl transferase	5.70E-111
371	comp162419_g1_i3	cyclin-K	2.80E-16
372	comp162429_g1_i2	NIMA-related kinase	6.60E-10
373	comp162429_g1_i3	NIMA-related kinase	6.60E-10
374	comp162429_g1_i4	NIMA-related kinase	6.80E-10
375	comp162434_g1_i2	HECD2_PONAB	9.50E-113
376	comp162489_g1_i1	conserved unknown protein	3.50E-14
377	comp162489_g1_i4	conserved unknown protein	3.50E-14
378	comp162519_g1_i2	BET1_MOUSE	6.70E-09
379	comp162540_g1_i4	pap fibrillin family	8.80E-16
380	comp162540_g1_i5	pap fibrillin family	9.00E-16
381	comp162559_g1_i3	PR2E2_ORYSJ	9.90E-33
382	comp162631_g1_i1	alpha-(1,6)-fucosyltransferase	2.20E-18
383	comp162631_g1_i3	FUT8_MOUSE	2.20E-22
384	comp162660_g1_i2	DBP5_LODEL	4.80E-48
385	comp162660_g1_i3	DBP5_LODEL	2.20E-48
386	comp162679_g1_i2	testis-expressed sequence 11 -like	5.50E-10
387	comp162703_g2_i1	AP4S_ARATH	2.00E-59
388	comp162703_g3_i1	AP4S_ARATH	1.20E-59
389	comp162756_g1_i2	Trigger factor	2.40E-23
390	comp162865_g1_i3	2ABD_CHICK	3.80E-118
391	comp162973_g1_i1	NLRC3_HUMAN	1.10E-14
392	comp163001_g1_i2	Tellurite resistance methyltransferase, , core	1.70E-18
393	comp163029_g1_i10	citramalate synthase	3.90E-175
394	comp163029_g1_i11	CIMA_GEOSL	3.30E-149
395	comp163029_g1_i2	CIMA_GEOSL	4.20E-149
396	comp163029_g1_i3	citramalate synthase	1.40E-174
397	comp163029_g1_i4	CIMA_GEOSL	9.90E-149
398	comp163029_g1_i5	CIMA_GEOSL	3.70E-149
399	comp163029_g1_i8	citramalate synthase	2.80E-175
400	comp163041_g1_i1	YCX91_PHAAORecName: Full=Uncharacterized protein ORF91 (chloroplast)	2.10E-18

401	comp163062_g1_i1	2ABD_CHICK	8.60E-116
402	comp163062_g1_i4	phosphatase PP2A regulatory subunit B	2.00E-126
403	comp163062_g1_i5	2ABD_CHICK	1.00E-118
404	comp163063_g1_i1	N-formylglutamate amidohydrolase	6.10E-18
405	comp163063_g1_i2	N-formylglutamate amidohydrolase	5.70E-18
406	comp163086_g1_i1	predicted protein	2.30E-08
407	comp163104_g1_i1	PCS3_LOTJA	1.60E-67
408	comp163128_g1_i1	SCRN3_BOVIN	1.90E-87
409	comp163128_g1_i2	secernin-3	2.90E-91
410	comp163175_g1_i2	TAR3_ARATH	1.60E-70
411	comp163175_g1_i5	TAR3_ARATH	2.50E-70
412	comp163197_g1_i2	RBMX2_RAT	3.00E-47
413	comp163223_g1_i1	zinc-binding ,	1.20E-68
414	comp163237_g1_i1	conserved unknown protein	3.00E-14
415	comp163237_g1_i2	conserved unknown protein	2.80E-14
416	comp163237_g1_i5	conserved unknown protein	2.80E-14
417	comp163237_g1_i7	conserved unknown protein	2.70E-14
418	comp163260_g1_i1	ATG12_ASPCL	3.50E-19
419	comp163268_g1_i3	CYSKP_SOLTU	2.20E-86
420	comp163276_g1_i2	GLTP1_ARATH	2.60E-32
421	comp163278_g1_i2	PLSP1_ARATH	1.50E-64
422	comp163285_g1_i1	MTCA2_MYCTO	3.60E-37
423	comp163315_g1_i1	S39AC_MACFA	7.20E-09
424	comp163319_g1_i2	psbP domain-containing 6, chloroplastic	2.70E-61
425	comp163442_g1_i1	U483_DROME	2.90E-21
426	comp163442_g1_i3	UPF0483 GA18864	2.20E-40
427	comp163584_g1_i2	zinc finger , partial	6.40E-36
428	comp163584_g1_i3	zinc finger , partial	6.50E-36
429	comp163597_g1_i2	Y1017_SYNY3	3.50E-12
430	comp163635_g1_i1	amino acid transporter	1.00E-13
431	comp163743_g1_i2	AIDA_DANRE	1.30E-31
432	comp163780_g1_i2	hypothetical protein GUITHDRAFT_158222	2.30E-14
433	comp163784_g1_i3	protein deltex-2	7.90E-94
434	comp163805_g1_i1	ENPP1_MOUSE	3.20E-09
435	comp163805_g2_i1	Type I phosphodiesterase nucleotide pyrophosphatase	1.60E-36

436	comp163805_g2_i4	ENPP1_MOUSE	3.00E-09
437	comp163866_g1_i1	NUDC_CHICK	4.60E-60
438	comp163867_g1_i1	SPXS1_DICDI	2.40E-28
439	comp163883_g1_i1	thioredoxin domain-containing 9-like	2.10E-45
440	comp163883_g1_i4	thioredoxin domain-containing 9-like	1.90E-45
441	comp163945_g1_i3	PSB5_DICDI	7.30E-94
442	comp163952_g1_i4	DOHH_DROME	1.00E-71
443	comp163952_g1_i5	DOHH_DROME	1.40E-71
444	comp163964_g1_i2	Hypothetical protein, putative	3.10E-09
445	comp164077_g1_i1	chloroplast envelope membrane	5.20E-42
446	comp164077_g1_i2	CEMA_NEPOL	2.50E-24
447	comp164091_g1_i1	lateral signaling target 2 homolog	5.40E-07
448	comp164091_g1_i2	lateral signaling target 2 homolog	5.60E-07
449	comp164099_g1_i1	ATPF1_XENTR	4.40E-08
450	comp164102_g1_i2	transmembrane 42	4.00E-09
451	comp164108_g1_i3	spermatogenesis-associated 17	8.30E-20
452	comp164129_g1_i3	ZRAB2_CHICK	2.10E-11
453	comp164157_g1_i6	SGT1 homolog A	4.20E-42
454	comp164172_g1_i1	PHOT2_ARATH	3.30E-23
455	comp164194_g1_i2	PDLI5_RAT	4.90E-15
456	comp164216_g1_i1	CAH1_CHIHA	1.20E-29
457	comp164242_g1_i3	GLO2_MOUSE	4.60E-69
458	comp164252_g1_i4	TRAM LAG1 CLN8 homology domain	2.10E-17
459	comp164252_g1_i5	TRAM LAG1 CLN8 homology domain	2.00E-17
460	comp164252_g1_i6	TRAM LAG1 CLN8 homology domain	2.10E-17
461	comp164262_g1_i1	TLC1_CHLTR	6.80E-11
462	comp164267_g1_i1	AVP_HORVU	4.90E-41
463	comp164300_g1_i2	CYAB_STIAU	8.20E-08
464	comp164336_g1_i2	RING finger 10	3.40E-29
465	comp164344_g1_i4	formamidopyrimidine-DNA glycosylase	1.30E-92
466	comp164390_g1_i2	S-adenosyl-L-methionine-dependent methyltransferases superfamily isoform 1	2.50E-34
467	comp164409_g1_i1	LPAT1_ARATH	1.10E-57
468	comp164413_g1_i2	FLU_ARATH	1.20E-12
469	comp164457_g1_i2	DOHH_DROME	1.60E-71
470	comp164471_g1_i1	predicted protein	5.50E-10

471	comp164481_g2_i2	Appr-1-p processing	8.30E-21
472	comp164527_g1_i2	S14L1_HUMAN	4.10E-07
473	comp164597_g1_i2	AROG_ECO57	4.40E-147
474	comp164626_g1_i2	GDIR_ARATH	2.10E-08
475	comp164653_g1_i3	Hypothetical protein, putative	7.50E-12
476	comp164653_g1_i4	Hypothetical protein, putative	7.70E-12
477	comp164653_g1_i5	Hypothetical protein, putative	7.50E-12
478	comp164725_g1_i3	PREDICTED: uncharacterized protein LOC105646130	3.40E-18
479	comp164733_g1_i1	PTPS_DROME	3.80E-29
480	comp164733_g1_i3	PTPS_DROME	3.10E-29
481	comp164733_g1_i6	PTPS_DROME	7.00E-31
482	comp164734_g1_i1	hypothetical protein TRIADDRAFT_57460	9.40E-11
483	comp164747_g1_i2	hypothetical protein COCSUDRAFT_66025	1.20E-09
484	comp164747_g1_i3	Dpy-30 motif	6.10E-39
485	comp164768_g1_i1	NUCL_XENLA	2.30E-11
486	comp164778_g1_i2	chaperone dnaJ 10-like isoform X1	6.70E-56
487	comp164778_g1_i4	DNJ10_ARATH	1.10E-34
488	comp164778_g1_i5	DNJ10_ARATH	1.90E-34
489	comp164821_g2_i1	beta-type IP39	8.30E-09
490	comp164821_g2_i2	beta-type IP39	7.40E-09
491	comp164821_g2_i3	beta-type IP39	7.60E-09
492	comp164858_g1_i2	ARIA_ARATH	3.10E-09
493	comp164861_g1_i3	translation initiation factor eIF2 alpha subunit	5.10E-91
494	comp164942_g3_i2	5 -AMP-activated kinase subunit gamma-1	8.00E-44
495	comp164955_g1_i3	serine threonine kinase	1.20E-21
496	comp164989_g1_i2	MESA_EMENI	1.30E-13
497	comp165007_g1_i5	NFU1_DROSI	1.40E-57
498	comp165022_g1_i1	PREDICTED: uncharacterized protein LOC105646130	5.80E-19
499	comp165023_g1_i3	FDHC_METTF	1.30E-19
500	comp165023_g2_i1	FDHC_METTF	1.40E-19
501	comp165026_g1_i3	FNTA_SOLLC	1.80E-57
502	comp165040_g1_i3	cytochrome b5	1.90E-17
503	comp165052_g1_i2	WAXS5_ARATH	3.60E-22
504	comp165063_g1_i4	TEA1_SCHPO	1.80E-20
505	comp165149_g1_i4	TM222_HUMAN	6.10E-30

506	comp165267_g1_i1	cyclopropane-fatty-acyl-phospholipid synthase	2.20E-80
507	comp165267_g1_i2	cyclopropane-fatty-acyl-phospholipid synthase	9.70E-81
508	comp165286_g2_i4	DODA_MIRJA	3.70E-12
509	comp165294_g1_i2	EKI_ARATH	5.40E-50
510	comp165297_g2_i1	DDX4_BOVIN	1.50E-69
511	comp165297_g2_i2	DDX4_BOVIN	1.30E-69
512	comp165323_g2_i1	PSBA_EUGGR	1.20E-31
513	comp165323_g2_i2	PSBA_EUGGR	3.50E-53
514	comp165323_g2_i3	PSBA_EUGGR	0.00E+00
515	comp165324_g1_i2	TIC21_ARATH	6.80E-07
516	comp165338_g1_i4	hypothetical protein	1.50E-30
517	comp165338_g1_i5	hypothetical protein	3.80E-30
518	comp165500_g1_i2	-tyrosine phosphatase 3	2.00E-15
519	comp165514_g1_i3	RING finger 10	2.90E-29
520	comp165515_g1_i1	MPK3_ARATH	7.20E-135
521	comp165565_g1_i1	TTC8_MOUSE	0.00E+00
522	comp165608_g1_i1	CAH1_CHIHA	1.00E-29
523	comp165642_g1_i2	hypothetical protein GUITHDRAFT_149743	3.70E-20
524	comp165651_g1_i1	TOC64_PEA	3.20E-13
525	comp165657_g1_i3	cysteine synthase	5.70E-66
526	comp165657_g1_i4	cysteine synthase	4.80E-66
527	comp165672_g1_i3	AT18A_ARATH	4.10E-89
528	comp165739_g1_i1	PSMD4_ARATH	4.70E-68
529	comp165740_g1_i2	homoisocitrate dehydrogenase	3.60E-81
530	comp165740_g2_i1	homoisocitrate dehydrogenase	2.90E-81
531	comp165748_g1_i1	Endonuclease exonuclease phosphatase	9.60E-30
532	comp165748_g1_i3	Endonuclease exonuclease phosphatase	8.30E-30
533	comp165751_g1_i3	TM165_HUMAN	4.20E-48
534	comp165760_g1_i1	Ctr2 family transporter: copper ion CTR-type copper transporter	1.70E-26
535	comp165760_g1_i3	Ctr2 family transporter: copper ion CTR-type copper transporter	1.90E-26
536	comp165761_g1_i1	mitochondrial paraflagellar rod (PFC16)	2.10E-35
537	comp165761_g1_i2	mitochondrial paraflagellar rod (PFC16)	2.30E-35
538	comp165770_g1_i3	Ctr2 family transporter: copper ion CTR-type copper transporter	1.80E-26
539	comp165770_g1_i4	Ctr2 family transporter: copper ion CTR-type copper transporter	2.10E-26
540	comp165914_g1_i2	SYFM_HUMAN	4.60E-56

541	comp165914_g1_i3	SYFM_HUMAN	5.00E-56
542	comp165936_g1_i2	SMYD3_HUMAN	2.70E-10
543	comp165938_g1_i1	Hermansky-Pudlak syndrome 1 isoform X1	7.10E-17
544	comp166014_g2_i2	coiled-coil domain-containing 148-like	3.90E-12
545	comp166048_g1_i1	ERV1_YEAST	3.50E-08
546	comp166049_g1_i2	50S ribosomal L21	1.20E-75
547	comp166049_g1_i4	50S ribosomal L21	9.10E-76
548	comp166051_g1_i1	DLDH_SCHPO	2.50E-39
549	comp166051_g1_i3	DLDH_SCHPO	2.00E-39
550	comp166051_g2_i2	AVT5_SCHPO	2.50E-55
551	comp166051_g2_i4	AVT5_SCHPO	5.40E-55
552	comp166062_g1_i2	TM222_HUMAN	3.40E-30
553	comp166062_g1_i5	TM222_HUMAN	5.90E-30
554	comp166102_g1_i2	hypothetical protein	1.10E-27
555	comp166198_g1_i2	sporulation RMD1	1.60E-47
556	comp166235_g1_i2	RTPR_EUGGR	1.10E-18
557	comp166254_g1_i2	lipopolysaccharide biosynthesis	1.90E-28
558	comp166286_g1_i2	CC189_MOUSE	1.70E-17
559	comp166348_g1_i2	hypothetical protein DICPUDRAFT_99500	1.40E-08
560	comp166376_g1_i2	DESI2_DANRE	3.10E-25
561	comp166376_g1_i3	DESI2_DANRE	3.30E-25
562	comp166391_g1_i2	NLRC3_HUMAN	8.60E-44
563	comp166442_g1_i2	LHTL6_ARATH	1.70E-11
564	comp166447_g1_i1	GUP1_YEAST	7.30E-63
565	comp166448_g1_i2	60S ribosomal L26	1.50E-20
566	comp166556_g1_i12	DYRK2_DICDI	4.40E-63
567	comp166556_g1_i2	Os03g0719500, partial	2.20E-158
568	comp166556_g1_i4	Os03g0719500, partial	4.70E-159
569	comp166556_g1_i5	Os03g0719500, partial	3.30E-158
570	comp166556_g1_i8	Os03g0719500, partial	5.40E-158
571	comp166556_g1_i9	Os03g0719500, partial	1.60E-157
572	comp166572_g1_i1	glycine cleavage system R	1.60E-15
573	comp166574_g1_i2	MGDP1_HUMAN	4.40E-40
574	comp166574_g1_i3	MGDP1_HUMAN	4.10E-40
575	comp166606_g1_i1	leucine rich repeat	3.50E-15

576	comp166606_g1_i2	PIRL5_ORYSJ	1.10E-12
577	comp166620_g1_i1	predicted protein	4.70E-19
578	comp166620_g1_i2	predicted protein	4.20E-19
579	comp166630_g1_i2	CRLA_DICDI	1.00E-08

Appendix table 2 genes name of area 2

(a) List of TF family (1 transcripts of 1 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp196057_g1_i4	C3H	9.80E-15

(b) List of other genes name (13 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp176341_g3_i1	hypothetical protein GUITHDRAFT_163095	2.80E-09
2	comp177968_g1_i3	hypothetical protein H310_02371	5.50E-40
3	comp168069_g1_i2	two-pore potassium channel 3-like	4.60E-27
4	comp188052_g2_i6	predicted protein, partial	1.30E-23
5	comp179399_g1_i4	thiol protease aleurain-like	1.80E-104
6	comp193142_g1_i2	PLCG1_MOUSE	1.10E-09
7	comp192108_g2_i3	UGGG_DROME	8.10E-173
8	comp193925_g1_i3	tRNA pseudouridine synthase A	5.70E-38
9	comp156342_g1_i1	STX6_CHICK	8.50E-07
10	comp176098_g1_i2	SLBP2_XENLA	2.70E-15
11	comp180673_g1_i9	YDK2_SCHPO	6.20E-18
12	comp192097_g2_i4	RIBBA_UNCTG	1.00E-47
13	comp168258_g1_i4	isoform A	1.70E-10

Appendix table 3 genes name of area 3

(a) List of TF family (73 transcripts of 31 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193951_g1_i6	ARR-B	1.20E-07
2	comp193951_g1_i7	ARR-B	1.20E-07
3	comp189382_g1_i3	ARR-B	1.20E-08
4	comp172316_g1_i2	C2H2	3.50E-09
5	comp188239_g1_i5	C3H	3.40E-08
6	comp188239_g1_i6	C3H	3.40E-08
7	comp182505_g1_i1	E2F_DP	4.40E-11
8	comp194799_g1_i4	GRAS	1.00E-38
9	comp193105_g2_i5	FAR1	4.90E-74
10	comp172891_g1_i3	NAC	1.80E-46
11	comp190140_g2_i11	LBD	3.00E-44
12	comp171421_g1_i1	NAC	3.30E-20
13	comp172402_g1_i2	NAC	4.90E-09
14	comp172567_g1_i3	NAC	4.80E-09
15	comp187312_g1_i3	Nin-like	2.70E-79
16	comp185491_g1_i3	NAC	9.50E-23
17	comp174049_g1_i1	G2-like	4.30E-09
18	comp172727_g1_i2	MYB	1.50E-138
19	comp192254_g1_i5	FAR1	2.90E-45
20	comp180619_g1_i2	TALE	1.50E-13
21	comp174861_g1_i5	B3	4.10E-52
22	comp174861_g1_i7	B3	4.10E-52
23	comp190700_g2_i6	GeBP	2.60E-09
24	comp175353_g1_i7	WOX	8.90E-33
25	comp189630_g1_i1	YABBY	1.40E-16
26	comp191991_g1_i1	YABBY	2.70E-36
27	comp174861_g1_i4	B3	3.50E-52
28	comp187492_g1_i2	BBR-BPC	1.30E-15
29	comp171906_g1_i1	C2H2	1.00E-07
30	comp188354_g1_i2	C3H	3.60E-08
31	comp188239_g1_i4	C3H	3.60E-08
32	comp169524_g1_i3	EIL	4.20E-56

33	comp170244_g1_i4	FAR1	4.80E-19
34	comp165906_g1_i2	GRAS	3.10E-08
35	comp190700_g2_i2	GeBP	2.60E-09
36	comp178049_g1_i1	HB-other	5.20E-11
37	comp181802_g1_i8	LBD	1.40E-44
38	comp187413_g1_i4	M-type	2.70E-23
39	comp174072_g1_i7	MYB	5.40E-56
40	comp179774_g3_i2	MYB	9.10E-16
41	comp193440_g1_i1	MYB_related	1.60E-30
42	comp162796_g1_i1	MYB_related	3.20E-20
43	comp177508_g1_i4	NAC	1.70E-33
44	comp177746_g1_i1	WRKY	1.30E-114
45	comp177746_g1_i5	WRKY	1.70E-113
46	comp178399_g2_i2	WRKY	2.10E-28
47	comp177523_g1_i2	YABBY	1.40E-10
48	comp174261_g1_i2	bHLH	5.60E-11
49	comp182329_g2_i2	bZIP	2.20E-08
50	comp185161_g1_i1	bZIP	3.90E-11
51	comp188354_g1_i3	C3H	3.50E-08
52	comp188354_g1_i5	C3H	3.30E-08
53	comp179240_g1_i1	CAMTA	1.30E-11
54	comp179347_g4_i3	Whirly	4.70E-08
55	comp168481_g2_i1	ERF	9.60E-15
56	comp179143_g1_i1	Dof	1.70E-29
57	comp179058_g1_i4	E2F_DP	1.40E-25
58	comp169411_g1_i2	ERF	9.40E-08
59	comp175288_g1_i1	MYB	3.60E-34
60	comp173869_g1_i2	NAC	1.20E-12
61	comp178433_g1_i4	Nin-like	2.00E-11
62	comp192254_g1_i9	FAR1	3.40E-45
63	comp168201_g1_i4	MYB_related	2.20E-40
64	comp190745_g1_i3	SBP	5.90E-18
65	comp180891_g1_i1	STAT	5.00E-15
66	comp180619_g1_i1	TALE	1.50E-13
67	comp175353_g1_i8	WOX	7.30E-33

68	comp193382_g1_i3	WRKY	2.00E-31
69	comp189385_g2_i1	YABBY	3.10E-26
70	comp176883_g1_i1	bHLH	2.10E-50
71	comp176730_g1_i5	bHLH	4.30E-15
72	comp175060_g2_i1	bHLH	1.30E-25
73	comp180899_g1_i2	bZIP	9.00E-20

(b) List of other genes name (344 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp179810_g1_i2	UPF0418 C6orf94	3.40E-12
2	comp184908_g1_i5	phospho phosphatase	7.60E-79
3	comp189402_g2_i1	K0556_XENLARecName: Full=Uncharacterized protein KIAA0556 homolog	4.90E-77
4	comp178379_g2_i2	ANK3_HUMAN	3.30E-40
5	comp180208_g1_i1	SIPL3_ARATH	6.50E-27
6	comp174735_g1_i3	Hypothetical protein, putative	1.80E-12
7	comp187609_g1_i9	1-acyl-sn-glycerol-3-phosphate acyltransferase delta isoform X1	1.20E-67
8	comp171613_g2_i2	CCB3_ARATH	3.00E-11
9	comp184602_g1_i6	CRTI_NARPS	0.00E+00
10	comp166630_g1_i2	CRLA_DICDI	1.00E-08
11	comp165007_g1_i5	NFU1_DROSI	1.40E-57
12	comp170558_g2_i2	conserved hypothetical protein, fragment	4.80E-15
13	comp189181_g1_i1	predicted protein, partial	2.90E-59
14	comp181987_g1_i2	Hypothetical protein, putative	3.60E-12
15	comp170699_g1_i1	PAPD5_MOUSE	3.60E-32
16	comp170275_g1_i5	Poly (A) RNA polymerase cid14	4.20E-51
17	comp174505_g1_i2	Poly(ADP-ribose) polymerase catalytic domain containing	2.80E-89
18	comp149586_g1_i6	ABC transporter	0.00E+00
19	comp185769_g1_i6	SLBP1_XENLA	2.50E-13
20	comp176652_g2_i1	hypothetical protein GUITHDRAFT_80007	3.40E-11
21	comp194882_g1_i2	ELAV 1 isoform X2	3.20E-12
22	comp194071_g4_i3	family methyltransferase	2.10E-08
23	comp183323_g1_i3	Coatomer subunit beta -3	1.20E-24
24	comp187790_g1_i14	carbamoyl phosphate synthase large subunit	3.00E-75

25	comp190494_g1_i1	S38A5_MOUSE	1.20E-17
26	comp189904_g1_i3	MSRB5_ARATH	2.10E-40
27	comp171014_g1_i2	FLU_ARATH	1.80E-13
28	comp195122_g1_i2	transmembrane protein, putative	1.70E-36
29	comp185393_g1_i1	hypothetical protein RMCBS344292_08877	8.50E-08
30	comp170207_g1_i1	CALM_EUGGR	1.40E-09
31	comp196175_g1_i5	Multi-sensor Hybrid Histidine Kinase	8.00E-13
32	comp155230_g1_i1	7-cyano-7-deazaguanine synthase	2.20E-29
33	comp164471_g1_i1	predicted protein	5.50E-10
34	comp193688_g2_i3	MKT1_YEAST	2.90E-11
35	comp176024_g1_i4	calcium-dependent cytoplasmic cysteine ase, papain	1.00E-98
36	comp195290_g2_i2	photosystem I assembly Ycf3	6.40E-10
37	comp155736_g1_i2	TPC6B_HUMAN	5.90E-22
38	comp194004_g1_i2	ADCYA_HUMAN	3.90E-19
39	comp182367_g1_i2	membrane-associated ,	3.70E-34
40	comp192177_g1_i5	KAPR_ASPFU	7.30E-08
41	comp182788_g1_i2	MYND finger family partial	4.10E-09
42	comp193409_g2_i1	THOP1_MOUSE	5.00E-27
43	comp190165_g1_i4	Plastid lipid-associated fibrillin conserved domain	5.80E-118
44	comp189216_g2_i5	alpha beta hydrolase	1.40E-10
45	comp180076_g1_i1	RPGR_CANFA	2.20E-17
46	comp188193_g2_i3	hypothetical protein THAOC_02569	1.50E-09
47	comp173942_g1_i1	RAD51_HUMAN	0.00E+00
48	comp182569_g2_i3	NUCL_HUMAN	3.00E-12
49	comp190241_g1_i1	LRC49_MOUSE	1.60E-48
50	comp179066_g1_i2	thiol protease aleurain-like	5.00E-105
51	comp153481_g1_i2	RIBX_SHIFL	4.10E-19
52	comp192253_g1_i9	leucine-rich repeat-containing C10orf11 homolog	4.40E-19
53	comp178657_g1_i3	NLRC3_MOUSE	3.30E-41
54	comp160278_g1_i3	paired amphipathic helix Sin3a isoform X1	2.50E-48
55	comp101701_g1_i1	RS24_DICDI	3.50E-25
56	comp178720_g2_i6	THIO_EMENI	2.10E-12
57	comp178213_g1_i1	GALM_ACICA	1.00E-101
58	comp156649_g1_i4	ATE1_ARATH	4.00E-43
59	comp192932_g2_i2	HECD1_HUMAN	2.00E-09

60	comp161345_g1_i3	PCS3_LOTJA	7.60E-68
61	comp183289_g1_i4	uncharacterized aarF domain-containing kinase At4g31390, chloroplastic	0.00E+00
62	comp189758_g2_i10	AARA_DICDI	2.10E-11
63	comp167503_g1_i2	INT11_DICDI	0.00E+00
64	comp182012_g1_i1	Fibronectin, type III domain and Immunoglobulin-like fold domain-containing	5.70E-08
65	comp185960_g1_i6	AT3g61870 F21F14_40	3.40E-55
66	comp196182_g1_i7	CYAA_TRYEQ	6.20E-24
67	comp166572_g1_i1	glycine cleavage system R	1.60E-15
68	comp183610_g1_i3	AK1_DICDI	4.50E-40
69	comp180209_g2_i3	Cytidine deaminase	2.60E-22
70	comp186830_g1_i1	DIG1_CAEEL	2.40E-07
71	comp187022_g1_i4	ABCA1_MOUSE	2.70E-10
72	comp196030_g1_i4	EXPR_XANCP	3.20E-65
73	comp189636_g2_i2	IP3KA_HUMAN	6.30E-37
74	comp180260_g1_i5	50S ribosomal L34	1.40E-09
75	comp162110_g2_i3	alpha-type IP39	1.00E-131
76	comp166632_g1_i1	hypothetical protein THAOC_10539	5.30E-11
77	comp189887_g3_i2	ABC transporter substrate-binding	1.30E-22
78	comp141931_g1_i1	PSAC_EUGGR	1.10E-33
79	comp192378_g1_i2	CYAB_STIAU	3.30E-12
80	comp175372_g1_i5	RRC1L_ARATH	4.90E-48
81	comp181202_g1_i1	FKBP3_MOUSE	4.10E-28
82	comp182507_g1_i4	D-alanine--D-alanine ligase	1.40E-141
83	comp195154_g2_i5	endonuclease V isoform X1	2.60E-25
84	comp196132_g1_i5	ef hand family	3.00E-36
85	comp184416_g3_i5	EXGA_NEOFI	3.70E-27
86	comp181140_g1_i2	ACD11_CHICK	2.20E-131
87	comp189834_g1_i1	MTAP2_RAT	2.70E-12
88	comp189755_g1_i4	AMP deaminase	0.00E+00
89	comp195858_g1_i3	proteasome activator complex subunit 4	1.40E-42
90	comp190765_g1_i4	Tbc2 translation chloroplastic	1.40E-43
91	comp159493_g1_i1	COX1_CHOPI	4.80E-148
92	comp193627_g2_i7	EF-hand calcium-binding domain-containing 6	4.20E-12
93	comp179373_g1_i2	UPTG_PEA	1.60E-150
94	comp174805_g1_i2	hypothetical protein	7.10E-07

95	comp187464_g1_i1	ATP-binding cassette transporter	4.60E-21
96	comp158481_g1_i3	METL4_MOUSE	7.30E-21
97	comp189528_g1_i4	MPCP_YEAST	2.40E-22
98	comp165748_g1_i1	Endonuclease exonuclease phosphatase	9.60E-30
99	comp175697_g1_i2	MTFP1_HUMAN	1.90E-09
100	comp185089_g1_i3	membrane-associated partial	8.00E-20
101	comp191720_g3_i1	NLRC3_HUMAN	1.40E-34
102	comp185389_g4_i5	CY43_TRYBB	2.10E-14
103	comp190973_g2_i13	histone-lysine N-methyltransferase SETMAR-like	2.50E-30
104	comp185316_g2_i3	transmembrane ,	9.40E-55
105	comp168433_g1_i1	salt tolerance 5	3.40E-16
106	comp182081_g1_i4	type II inositol 1,4,5-trisphosphate 5-phosphatase	3.10E-53
107	comp187997_g1_i4	WDR35_HUMAN	3.30E-22
108	comp187197_g1_i3	hypothetical protein UT61_C0050G0009	1.50E-12
109	comp165761_g1_i1	mitochondrial paraflagellar rod (PFC16)	2.10E-35
110	comp188259_g3_i1	Mitochondrial transcription termination factor family	6.10E-09
111	comp195560_g1_i4	voltage-gated ion channel superfamily	3.60E-14
112	comp191199_g1_i7	DTW domain-containing	1.30E-16
113	comp143474_g1_i1	RL34A_YEAST	1.90E-20
114	comp170275_g1_i10	Poly (A) RNA polymerase cid14	8.30E-51
115	comp165026_g1_i3	FNTA_SOLLC	1.80E-57
116	comp175952_g1_i1	PROD_HUMAN	3.30E-102
117	comp192000_g1_i2	ADT3_BOVIN	2.50E-20
118	comp187755_g2_i3	hypothetical protein CHLREDRAFT_206193	5.50E-07
119	comp171637_g1_i1	GLRX1_RHIID	1.50E-11
120	comp188699_g1_i3	calcium sensing receptor, chloroplastic	7.00E-21
121	comp186460_g1_i3	hnRNPL PTB hephaestus splicing factor subfamily	8.40E-75
122	comp186529_g1_i4	CHO2_ASPFU	3.20E-101
123	comp195456_g1_i9	kinase domain	5.90E-28
124	comp190508_g1_i2	PBP_BRUA2	9.70E-09
125	comp193191_g1_i2	Presequence protease 1, chloroplastic mitochondrial	0.00E+00
126	comp192602_g1_i2	CY42_TRYBB	1.40E-47
127	comp179493_g3_i3	KAPR_YARLI	5.70E-14
128	comp195499_g1_i1	SC4AA_TETNG	1.40E-08
129	comp191224_g1_i2	glutathione S-transferase C-terminal domain-containing	1.10E-22

130	comp191048_g1_i2	CHI11_METAN	9.70E-09
131	comp164242_g1_i3	GLO2_MOUSE	4.60E-69
132	comp187555_g3_i9	AK1_DICDI	2.00E-34
133	comp183983_g1_i5	hypothetical protein EMCG_09124	5.90E-07
134	comp163128_g1_i2	secernin-3	2.90E-91
135	comp191463_g1_i7	VMAT1_MOUSE	1.40E-17
136	comp195432_g2_i1	PK2L1_HUMAN	7.40E-14
137	comp182118_g1_i3	AARA_DICDI	2.10E-17
138	comp190428_g4_i4	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1 isoform X2	6.80E-85
139	comp166447_g1_i1	GUP1_YEAST	7.30E-63
140	comp184137_g1_i1	GGPPS_MUCCL	2.70E-35
141	comp187400_g2_i9	PDE9A_HUMAN	4.30E-24
142	comp188998_g1_i3	violaxanthin de- chloroplastic	1.70E-89
143	comp195767_g2_i6	OTU7B_HUMAN	1.30E-22
144	comp186080_g5_i1	RBBP9_HUMAN	8.80E-39
145	comp183741_g1_i1	calmodulin mutant syncam9	8.60E-09
146	comp195112_g3_i1	Multi-sensor hybrid histidine kinase	4.20E-73
147	comp176000_g1_i6	ZDH13_ARATH	3.80E-43
148	comp172210_g3_i1	family methyltransferase	2.10E-23
149	comp184457_g1_i4	E3 ubiquitin- ligase rififylin-like	9.70E-09
150	comp193226_g2_i1	CY43_TRYBB	1.70E-11
151	comp182609_g2_i1	Y0701_DICDI	1.80E-24
152	comp186814_g1_i3	ABCD4_MOUSE	1.90E-66
153	comp173314_g1_i2	SETD4_MOUSE	4.60E-13
154	comp166967_g1_i5	probable G- coupled receptor 157	2.80E-16
155	comp187210_g1_i9	NEK1_MOUSE	2.70E-45
156	comp186830_g1_i6	DIG1_CAEEL	2.50E-07
157	comp189611_g5_i3	hypothetical protein ACA1_090140	1.30E-17
158	comp183983_g1_i7	TPA: vesicle-mediated transport (Imh1), (AFU_orthologue AFUA_4G02990)	3.50E-07
159	comp186161_g2_i1	PLCD4_HUMAN	6.10E-63
160	comp190696_g2_i1	flagellar associated	1.10E-67
161	comp192391_g2_i7	hypothetical protein AURANDRAFT_67549	3.60E-13
162	comp166774_g1_i11	PTBP3_ARATH	6.90E-08
163	comp190304_g1_i7	hypothetical protein LOTGIDRAFT_190059	5.40E-07
164	comp181515_g1_i2	hypothetical protein	1.00E-07

165	comp187483_g1_i8	transcriptional regulator	1.10E-07
166	comp192399_g1_i19	Hypothetical protein, putative	3.80E-16
167	comp195922_g1_i11	CYA1_RHIME	4.20E-24
168	comp194234_g1_i2	vacuolar sorting-associated vps13	1.30E-43
169	comp116744_g1_i1	hypothetical protein Ctob_007964	9.30E-17
170	comp179243_g1_i5	ECE1_MOUSE	2.60E-103
171	comp190016_g1_i4	FBX11_RAT	5.60E-46
172	comp183216_g1_i1	CRNS1_HUMAN	2.40E-101
173	comp189611_g2_i3	udp-n-acetylglucosamine--peptide n-acetylglucosaminyltransferase	2.30E-17
174	comp181283_g1_i2	ENG2_SCHPO	3.30E-42
175	comp192503_g1_i1	cation channel family	1.80E-27
176	comp189876_g4_i6	predicted protein	1.10E-17
177	comp182433_g2_i3	hypothetical protein GUITHDRAFT_131487	6.10E-22
178	comp176141_g1_i3	Mitochondrial transcription termination factor family	5.10E-10
179	comp185293_g1_i12	MPPD1_HUMAN	1.10E-55
180	comp196048_g1_i1	NLRC3_HUMAN	1.10E-21
181	comp186243_g2_i5	glutathione S-transferase	1.70E-30
182	comp188859_g1_i12	hypothetical protein H310_00261	2.10E-10
183	comp195714_g2_i5	uracil-DNA glycosylase	3.30E-16
184	comp192179_g1_i8	membrane-bound adenylyl cyclase	1.00E-40
185	comp183648_g1_i3	NLRC3_HUMAN	2.40E-36
186	comp188065_g1_i2	SLBP2_XENLA	9.90E-09
187	comp157432_g1_i2	OTUD3_MOUSE	1.10E-31
188	comp175173_g1_i3	cell adhesion domain-containing	2.70E-37
189	comp168047_g1_i2	PAS domain-containing sensor histidine kinase	4.00E-16
190	comp177734_g1_i2	hypothetical protein AMSG_09933	2.30E-51
191	comp166198_g1_i2	sporulation RMD1	1.60E-47
192	comp192692_g3_i1	DNA glycosylase	3.70E-46
193	comp167300_g1_i3	KAPR_DICDI	8.30E-75
194	comp190372_g1_i4	SET and MYND domain-containing 4	3.00E-09
195	comp180968_g1_i2	CD027_MOUSE	2.70E-50
196	comp185960_g1_i1	AT3g61870 F21F14_40	9.20E-55
197	comp176998_g1_i3	phosphatidate cytidylyltransferase	7.70E-20
198	comp175843_g1_i3	MKAR_COPC7	3.30E-73
199	comp191177_g2_i3	PGML1_ARATH	4.60E-21

200	comp175448_g1_i10	hypothetical protein CHLNCDRAFT_138448	2.60E-10
201	comp196159_g1_i4	PBP_BRUA2	9.90E-09
202	comp185248_g1_i3	TM53B_XENLA	8.10E-09
203	comp182816_g1_i4	OAC1_YEAST	2.90E-47
204	comp170477_g1_i9	citramalate synthase	8.50E-175
205	comp185995_g1_i2	transporter, major facilitator superfamily superfamily	4.20E-51
206	comp187484_g3_i8	luc7 3	2.40E-125
207	comp182454_g1_i5	DPOL_DHBV1	2.20E-11
208	comp194165_g1_i7	Adenylate cyclase	3.60E-37
209	comp184814_g1_i8	ATPase AAA	1.40E-46
210	comp182701_g1_i3	T9SS C-terminal target domain-containing	2.20E-16
211	comp194810_g1_i1	leucine-rich repeat	2.90E-11
212	comp161863_g1_i3	hypothetical protein PRUPE_ppa021045mg	6.20E-08
213	comp163883_g1_i1	thioredoxin domain-containing 9-like	2.10E-45
214	comp168114_g1_i2	RH46_ARATH	1.10E-73
215	comp172243_g1_i4	VP33_APLCA	2.90E-08
216	comp189717_g1_i2	CY43_TRYBB	1.10E-10
217	comp188911_g1_i1	probable alpha-1,6-mannosyltransferase MNN10	3.20E-23
218	comp167679_g1_i6	folate pteridine transporter	2.90E-16
219	comp189991_g2_i1	single-stranded DNA-binding	3.10E-17
220	comp182366_g1_i2	bile Acid:Na+ symporter family	6.20E-39
221	comp180579_g1_i5	DDL_RALME	3.10E-09
222	comp186398_g1_i5	tyrosine phosphatase	1.90E-66
223	comp162177_g1_i5	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	1.40E-53
224	comp193426_g1_i2	glycerol-3-phosphate dehydrogenase, mitochondrial	4.40E-145
225	comp196317_g1_i1	inositol 1,4,5-trisphosphate receptor isoform X1	1.00E-112
226	comp174989_g1_i5	probable S-acyltransferase 4	8.50E-44
227	comp176899_g1_i5	CSE_ARATH	4.60E-43
228	comp189162_g2_i2	phytoene desaturase	0.00E+00
229	comp182446_g1_i2	metallophosphoesterase domain-containing 1	7.00E-62
230	comp182868_g2_i3	Amine oxidase	4.40E-77
231	comp184253_g2_i2	hypothetical protein COCSUDRAFT_56625	2.10E-24
232	comp186752_g1_i2	nicotinate-nucleotide adenylyltransferase	9.20E-53
233	comp174836_g1_i6	Hypothetical protein, putative	1.10E-11
234	comp149630_g1_i2	YHCV_BACSU	3.60E-15

235	comp169595_g1_i1	calmodulin-binding ,	1.70E-14
236	comp193679_g1_i3	KAPR_APLCA	2.20E-47
237	comp187682_g1_i2	UDP-galactopyranose mutase	2.90E-135
238	comp187483_g1_i7	transcriptional regulator	9.80E-08
239	comp191652_g3_i3	3 , 5 -cyclic nucleotide phosphodiesterase	1.10E-109
240	comp194738_g2_i7	TYW1_ORYSJ	1.50E-105
241	comp184016_g1_i1	hypothetical protein CHLNCDRAFT_53531	8.40E-13
242	comp187090_g1_i2	TRPV1_CANFA	9.50E-14
243	comp192267_g3_i1	exported lipo	2.30E-30
244	comp178824_g1_i5	YPL1_DROME	1.00E-09
245	comp183521_g1_i4	manganese transporter	3.20E-146
246	comp182421_g1_i5	GGPPS_MUCCL	3.10E-35
247	comp186139_g1_i5	methyltransferase family	8.90E-34
248	comp188701_g2_i1	hypothetical protein GUITHDRAFT_163095	3.90E-10
249	comp184396_g1_i2	tRNAHis guanylyltransferase	2.50E-46
250	comp188998_g1_i5	violaxanthin de- chloroplastic	1.10E-89
251	comp194103_g2_i6	CYNT_MYCS2	6.70E-26
252	comp189464_g2_i4	FPG_ARATH	8.80E-52
253	comp188537_g1_i2	presequence protease 1, chloroplastic mitochondrial-like	0.00E+00
254	comp189142_g4_i3	transmembrane ,	9.30E-30
255	comp191827_g2_i2	6-4 photolyase	3.50E-159
256	comp163237_g1_i7	conserved unknown protein	2.70E-14
257	comp188055_g2_i1	ANR50_HUMAN	9.70E-37
258	comp192221_g1_i4	AB31G_ARATH	1.50E-140
259	comp177994_g1_i1	hypothetical protein SPPG_02418	1.50E-12
260	comp164778_g1_i5	DNJ10_ARATH	1.90E-34
261	comp187747_g1_i2	MCFE_DICDI	6.80E-39
262	comp193157_g1_i2	peptidyl-prolyl cis-trans isomerase FKBP62-like isoform X2	2.40E-23
263	comp165608_g1_i1	CAH1_CHIHA	1.00E-29
264	comp173296_g1_i3	metallophosphoesterase domain-containing	4.20E-23
265	comp165740_g1_i2	homoisocitrate dehydrogenase	3.60E-81
266	comp185155_g2_i2	MYCB2_MOUSE	1.90E-33
267	comp187283_g6_i6	predicted protein, partial	2.70E-26
268	comp175817_g1_i2	MUTYH_MOUSE	8.50E-77
269	comp190372_g1_i2	SET and MYND domain-containing 4	3.00E-09

270	comp172240_g2_i7	DNA polymerase epsilon subunit 2	1.80E-87
271	comp179716_g1_i4	coiled-coil domain containing 39	1.80E-162
272	comp193681_g1_i2	GEFF_DICDI	6.50E-15
273	comp190949_g1_i5	AARA_DICDI	4.80E-07
274	comp177520_g1_i1	flagellar associated	3.60E-30
275	comp193627_g1_i9	EF-hand calcium-binding domain-containing 6	3.30E-12
276	comp190028_g2_i2	nonspecific lipid-transfer	1.90E-114
277	comp175191_g2_i2	hypothetical protein EMIHUDRAFT_453971	3.00E-14
278	comp185073_g1_i7	hypothetical protein CY34DRAFT_805584	7.40E-11
279	comp191929_g2_i1	ESYT3_XENTR	2.00E-43
280	comp192633_g1_i6	WDFY3_HUMAN	6.90E-105
281	comp189398_g1_i3	CNOT7_MOUSE	8.70E-84
282	comp186653_g3_i4	serine threonine- kinase 2 19-like	4.30E-36
283	comp188759_g1_i2	IF4E_DICDI	1.30E-14
284	comp195141_g1_i5	SLN1_YEAST	9.00E-14
285	comp194992_g1_i5	ANR50_HUMAN	1.50E-16
286	comp194950_g1_i2	ANM7_XENLA	2.20E-18
287	comp191635_g1_i4	cyclic nucleotide-binding domain	1.40E-08
288	comp179471_g1_i1	metallophosphoesterase domain-containing	9.10E-23
289	comp186130_g1_i2	BAT1_ORYSJ	4.20E-86
290	comp188870_g1_i3	snf1-related kinase catalytic subunit alpha kin10	1.40E-136
291	comp190681_g1_i1	PREDICTED: nucleolin	3.70E-07
292	comp148201_g1_i1	mitochondrial tryptophan rich sensory 1	7.90E-09
293	comp187283_g6_i4	YDJX_ECOLI	4.70E-07
294	comp195734_g5_i2	mediator of RNA polymerase II transcription subunit 18	2.10E-08
295	comp194189_g1_i8	5NTD_LUTLO	3.00E-32
296	comp188754_g3_i1	DPOLL_MACFA	2.40E-60
297	comp182251_g1_i4	BACH_MOUSE	1.40E-21
298	comp178727_g1_i3	CASTO_ORYSJ	2.40E-85
299	comp167787_g1_i2	hypothetical protein MNEG_13922	8.30E-16
300	comp190086_g2_i2	1A12_CUCMA	7.80E-71
301	comp159304_g1_i2	Adenylosuccinate lyase	3.00E-10
302	comp258859_g1_i1	YCX7_EUGGR	3.00E-171
303	comp177968_g1_i1	hypothetical protein H310_02371	5.70E-40
304	comp173352_g1_i2	AB5F_ARATH	2.10E-147

305	comp172725_g1_i1	probable G- coupled receptor 157	3.10E-16
306	comp175317_g1_i3	LAMA5_MOUSE	6.70E-08
307	comp187124_g1_i5	brucei -specific	2.00E-12
308	comp182250_g2_i1	UPF0586 C9orf41 homolog isoform X2	4.50E-60
309	comp185938_g1_i3	WDR90_HUMAN	5.70E-40
310	comp183773_g1_i5	UVH3_ARATH	6.30E-38
311	comp193325_g1_i4	DEAD DEAH box helicase	9.70E-14
312	comp195823_g1_i2	membrane-bound adenylyl	1.70E-08
313	comp195254_g3_i6	OTU family cysteine protease	2.20E-24
314	comp134569_g1_i2	Y1181_ARATH	1.50E-31
315	comp184319_g1_i5	HNRPR_HUMAN	1.80E-08
316	comp173437_g1_i2	ABCC8_HUMAN	4.60E-15
317	comp176383_g1_i2	hypothetical protein GUITHDRAFT_162439	9.10E-08
318	comp194030_g3_i1	kinase domain	1.70E-12
319	comp194001_g1_i5	hypothetical protein	4.60E-20
320	comp192781_g2_i2	CPT2_DANRE	6.80E-78
321	comp191241_g1_i1	GT15_ORYSJ	2.00E-14
322	comp196159_g1_i2	ABC transporter substrate-binding	9.80E-21
323	comp176517_g2_i2	FLNA_DROME	1.00E-12
324	comp151206_g1_i1	PSBB_EUGGR	0.00E+00
325	comp193117_g1_i1	tRNA pseudouridine synthase A	3.50E-17
326	comp174011_g2_i1	hydroxyacylglutathione hydrolase	4.60E-112
327	comp193547_g1_i2	CID4_ARATH	3.10E-08
328	comp190883_g1_i3	molecular chaperone	3.80E-64
329	comp171450_g1_i3	PSBD_EUGGR	1.70E-86
330	comp185453_g1_i6	GTP pyrophosphokinase	5.20E-113
331	comp183097_g1_i4	IDH_AZOVI	0.00E+00
332	comp192100_g1_i1	Mannosyltransferase OCH1 and related enzymes	2.70E-13
333	comp182313_g1_i3	glycosyltransferase aer61	9.20E-25
334	comp194775_g2_i3	serine threonine- kinase STN7, chloroplastic	6.90E-103
335	comp226995_g1_i1	RS29_WHEAT	1.70E-15
336	comp173575_g1_i4	TTD14_DROME	2.90E-70
337	comp184482_g1_i2	STT3A_ORYSJ	1.40E-21
338	comp192399_g1_i3	Hypothetical protein, putative	2.80E-16
339	comp188689_g3_i3	YIID_ECO57RecName: Full=Uncharacterized protein YiiD	1.50E-08

340	comp175632_g1_i5	B3GTF_ARATH	2.50E-10
341	comp185198_g1_i3	YIDE_SCHPO	1.30E-22
342	comp186126_g1_i4	NLRC3_MOUSE	5.50E-13
343	comp191360_g2_i1	NADB_PSEAE	1.80E-125
344	comp167724_g1_i4	Hypothetical protein, putative	7.20E-12

Appendix table 4 genes name of area 4

List of other genes name (22 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp161199_g1_i2	WSDU1_DANRE	3.10E-09
2	comp185483_g1_i9	CRNS1_CHICK	3.10E-18
3	comp194344_g1_i13	hybrid sensor histidine kinase response regulator	7.10E-10
4	comp175256_g1_i3	aspartic protease	1.80E-13
5	comp193223_g1_i2	FIXL_BRADU	7.60E-09
6	comp158481_g1_i1	METL4_MOUSE	3.80E-21
7	comp180718_g3_i2	Methyltransferase like 11A, related	1.50E-35
8	comp184316_g1_i1	COPDA_ARATH	0.00E+00
9	comp172118_g1_i1	AOX_CRYNH	5.30E-53
10	comp193590_g1_i3	glutathione S-transferase	9.40E-10
11	comp187322_g1_i8	TBL16_ARATH	6.60E-11
12	comp163964_g1_i2	Hypothetical protein, putative	3.10E-09
13	comp172946_g1_i2	FLU_ARATH	7.20E-13
14	comp180128_g1_i5	TBC1 domain family member 31	4.20E-19
15	comp109717_g1_i1	hypothetical protein DQ04_09981010	3.60E-13
16	comp194293_g1_i10	inositol hexakisphosphate kinase 3	1.30E-30
17	comp195344_g1_i9	e3 ubiquitin- ligase	9.30E-24
18	comp175564_g1_i3	PFR1_TRYBB	6.70E-14
19	comp180260_g1_i2	50S ribosomal L34	1.80E-09
20	comp186806_g2_i7	nitrate reductase	4.70E-51
21	comp195984_g1_i6	nxn protein	0.00E+00
22	comp178461_g1_i2	hypothetical protein COCSUDRAFT_63432	4.90E-14

Appendix table 5 genes name of area 5

(a) List of TF family (1 transcripts of 1 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp174158_g1_i4	C3H	1.40E-29

(b) List of other genes name (32 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193275_g2_i5	transcriptional regulator	9.90E-13
2	comp181051_g1_i2	RL2_ANAPZ	6.80E-44
3	comp190182_g3_i1	AP3B1_BOVIN	2.20E-138
4	comp186810_g1_i4	phosphatidylinositol class b	1.00E-52
5	comp155692_g1_i1	amino acid transporter,	3.20E-51
6	comp183162_g1_i3	CB2_DUNSA	3.60E-17
7	comp177081_g1_i2	YGB0_YEAST	1.90E-14
8	comp177997_g2_i3	SDP1L_ARATH	8.60E-63
9	comp190606_g2_i12	poly(ADP-ribose) glycohydrolase-like	2.00E-87
10	comp193816_g1_i5	NOXA_DICDI	3.40E-26
11	comp192920_g2_i5	hypothetical protein Ctob_010150	4.70E-09
12	comp166606_g1_i1	leucine rich repeat	3.50E-15
13	comp176024_g1_i2	calcium-dependent cytoplasmic cysteine ase, papain	1.20E-98
14	comp190336_g2_i2	ALG12_DROME	3.00E-63
15	comp172871_g1_i5	hypothetical protein H257_10789	1.20E-18
16	comp187283_g6_i11	YDJX_ECOLI	4.90E-07
17	comp173909_g2_i4	UN104_CAEEL	1.40E-10
18	comp165760_g1_i3	Ctr2 family transporter: copper ion CTR-type copper transporter	1.90E-26
19	comp190038_g1_i3	RTCA_MOUSE	1.90E-74
20	comp190696_g2_i3	CE104_HUMAN	7.00E-35
21	comp193633_g1_i2	hypothetical protein, variant	6.60E-45
22	comp187198_g3_i1	CLPB_PARUW	8.80E-26
23	comp176156_g1_i2	S6A12_HUMAN	3.80E-11
24	comp182206_g1_i2	ALR_HUMAN	9.10E-19
25	comp191941_g1_i2	von Willebrand factor A domain-containing 8	0.00E+00
26	comp194883_g2_i1	hypothetical protein Naga_100042g35	1.70E-08

27	comp177405_g1_i3	D4FAD_CHLRE	6.00E-130
28	comp191912_g1_i13	AARA_DICDI	1.10E-10
29	comp186677_g1_i5	PUM5_ARATH	4.70E-22
30	comp177589_g1_i1	alpha-type kinase	1.10E-46
31	comp194165_g1_i4	CYAA_ANACY	1.20E-20
32	comp185268_g3_i3	gonidia-specific KA_k47	3.20E-23

Appendix table 6 genes name of area 6

List of other genes name (31 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp178248_g1_i2	hypersensitive-induced response 3-like	3.40E-12
2	comp190889_g1_i4	histidine kinase	1.70E-57
3	comp194154_g2_i4	3 ,5 -cyclic-nucleotide phosphodiesterase	3.10E-123
4	comp187208_g2_i2	DUS1_ARATH	2.20E-27
5	comp172725_g1_i2	GP157_MOUSE	1.90E-08
6	comp185163_g1_i3	mitochondrial transcription termination factor family	2.40E-13
7	comp188922_g1_i3	KIF28_RAT	1.80E-19
8	comp181873_g4_i1	membrane-bound adenylyl cyclase	2.10E-104
9	comp161352_g1_i3	predicted protein	3.00E-08
10	comp186370_g1_i2	tetratricopeptide repeat domain	8.10E-36
11	comp189863_g1_i5	COBS_LEPCP	4.90E-21
12	comp167862_g2_i3	family methyltransferase	6.20E-38
13	comp181515_g1_i10	hypothetical protein	7.30E-08
14	comp191840_g2_i3	ITS3_SCHPO	5.00E-54
15	comp194827_g2_i6	FDHC_METTF	2.20E-17
16	comp156557_g1_i1	F151B_HUMAN	1.30E-21
17	comp191158_g1_i2	SODF_PLAFX	3.00E-58
18	comp188827_g1_i3	Tumor susceptibility gene 101	1.50E-28
19	comp178824_g1_i8	YPL1_DROME	1.00E-09
20	comp173909_g2_i2	UN104_CAEEL	1.40E-10
21	comp180132_g1_i6	glycosyltransferase family 15	4.60E-33
22	comp194761_g1_i2	AARA_DICDI	1.20E-47
23	comp180429_g2_i2	CYAA_TRYEQ	1.10E-33
24	comp195338_g1_i5	DRE1_CAEEL	1.70E-42
25	comp193861_g1_i5	hypothetical protein AMSG_03072	9.80E-10
26	comp194289_g3_i5	DNA2_MOUSE	3.40E-112
27	comp186197_g1_i2	alpha beta-hydrolase	3.90E-31
28	comp195414_g1_i2	PABP6_ARATH	1.50E-13
29	comp187387_g1_i3	multiple C2 and transmembrane domain-containing 1	2.60E-09
30	comp195165_g1_i4	NHL repeat containing	2.30E-30
31	comp150234_g1_i3	COM1_MOUSE	1.50E-11

Appendix table 7 genes name of area 7

(a) List of TF family (23 transcripts of 15 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp192285_g1_i2	C3H	5.90E-15
2	comp192285_g1_i6	C3H	7.10E-15
3	comp190065_g1_i6	LBD	3.00E-20
4	comp184981_g1_i3	MYB	2.40E-21
5	comp184885_g1_i3	NAC	5.90E-33
6	comp189365_g1_i2	YABBY	1.10E-31
7	comp190938_g1_i4	YABBY	7.80E-31
8	comp190925_g1_i13	bZIP	3.50E-13
9	comp191409_g1_i3	GRAS	3.30E-08
10	comp162219_g1_i1	HB-other	1.90E-22
11	comp184733_g2_i4	NAC	4.10E-30
12	comp184945_g1_i3	NAC	1.20E-10
13	comp163941_g1_i2	Trihelix	7.90E-10
14	comp181811_g1_i3	bHLH	8.40E-23
15	comp180941_g1_i2	bHLH	3.90E-21
16	comp183118_g2_i2	B3	4.50E-21
17	comp176881_g1_i5	FAR1	1.50E-08
18	comp186096_g2_i5	GATA	5.00E-20
19	comp194663_g1_i9	ARR-B	5.20E-09
20	comp192782_g1_i1	C3H	1.20E-37
21	comp184885_g1_i6	NAC	3.70E-36
22	comp183768_g1_i3	TALE	2.00E-10
23	comp181755_g1_i4	bHLH	1.90E-28

(b) List of other genes name (206 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp172725_g1_i4	GP157_MOUSE	1.70E-08
2	comp177058_g1_i2	Y233_STRP6	1.40E-08
3	comp188879_g1_i12	cAMP specific phosphodiesterase	3.70E-172
4	comp187304_g1_i2	Kynurenine 3-monooxygenase and related flavo monooxygenases (ISS)	1.70E-28

5	comp185910_g2_i2	core-2 i-branching beta--n-acetylglucosaminyltransferase family	2.50E-14
6	comp93419_g1_i2	PREDICTED: uncharacterized protein LOC105619251	3.10E-39
7	comp166556_g1_i5	Os03g0719500predicted protein, partial	3.30E-158
8	comp193858_g1_i3	Y1181_ARATH	5.10E-121
9	comp196310_g3_i6	CYAA_TRYEQ	9.20E-47
10	comp183592_g1_i2	LON peptidase N-terminal domain and RING finger 2predicted protein, partial	4.70E-46
11	comp192851_g1_i3	PDE9A_RAT	1.10E-16
12	comp194745_g2_i3	METH_BOVIN	0.00E+00
13	comp165651_g1_i1	TOC64_PEA	3.20E-13
14	comp157038_g1_i1	tartrate-resistant acid phosphatase type 5	1.20E-53
15	comp146407_g1_i1	fungal specific transcription	5.30E-13
16	comp194859_g2_i2	CYAA_LEIDO	9.90E-31
17	comp192736_g2_i1	BARA_ECO57	3.30E-37
18	comp192399_g1_i2	Hypothetical protein, putative	4.70E-16
19	comp179921_g1_i3	LAAT1_CAEEL	5.50E-22
20	comp180715_g1_i2	Y544_RICPR	1.10E-69
21	comp194343_g1_i2	CYA1_RHIME	3.00E-18
22	comp195409_g1_i3	DNA J-binding	1.30E-113
23	comp166442_g1_i2	LHTL6_ARATH	1.70E-11
24	comp181650_g1_i3	NLRC3_HUMAN	1.30E-32
25	comp195141_g1_i9	SLN1_YEAST	8.80E-14
26	comp194451_g7_i1	CYA1_RHIME	1.10E-42
27	comp191055_g2_i3	hypothetical protein, variant 1	3.00E-78
28	comp179723_g1_i2	Y777_CHLL2	2.80E-11
29	comp196105_g2_i2	beta subunit of photoactivated adenylyl cyclase	2.10E-16
30	comp189761_g1_i1	hypothetical protein SDRG_06217	6.00E-40
31	comp187819_g1_i2	CBSX2_ARATH	4.50E-22
32	comp162973_g1_i1	NLRC3_HUMAN	1.10E-14
33	comp34761_g1_i1	HDDC2_XENLA	1.50E-12
34	comp175933_g1_i1	serine threonine kinase A	1.10E-08
35	comp172200_g1_i3	hypothetical protein UT61_C0050G0009	2.50E-08
36	comp174181_g1_i4	YF2E_SCHPO	1.80E-31
37	comp195323_g1_i1	ANM7_XENLA	2.10E-18
38	comp195653_g1_i5	GFA1_CANAL	9.40E-155
39	comp187579_g1_i2	sterol 3-beta-glucosyltransferase,	9.50E-148

40	comp195596_g1_i2	ADP-ribosyltransferase 1 precursor	9.00E-11
41	comp193058_g3_i4	CYAA_LEIDO	1.70E-43
42	comp168509_g1_i1	N-acetyl-D-glucosaminylphosphatidylinositol de-N-acetylase	2.10E-52
43	comp194688_g1_i8	AB31G_ARATH	5.40E-141
44	comp186560_g2_i3	XYLB_MOUSE	7.30E-119
45	comp193064_g1_i7	hypothetical protein	2.60E-58
46	comp177716_g1_i3	formate transporter	1.00E-113
47	comp187129_g2_i5	ATPG_ODOSI	8.70E-77
48	comp170477_g1_i5	citramalate synthase	2.90E-175
49	comp194489_g2_i2	Y663_METTH	6.10E-76
50	comp181532_g1_i3	LANC1_DANRE	2.30E-11
51	comp183962_g1_i2	RAY1_ARATH	1.50E-09
52	comp167764_g1_i3	VF307_IIV3RecName: Full=Uncharacterized protein 033L	4.50E-24
53	comp189697_g1_i3	hypothetical protein THAOC_09497	4.50E-21
54	comp188085_g1_i6	kinase,	3.30E-49
55	comp193185_g1_i2	ef hand family	4.20E-32
56	comp195531_g1_i12	FER3_CYACA	2.00E-10
57	comp190712_g1_i3	hypothetical protein COCSUDRAFT_31865	1.50E-19
58	comp196104_g4_i4	membrane-bound adenylyl cyclase	1.30E-99
59	comp174391_g1_i3	Subunit of Golgi mannosyltransferase complex (ISS)	2.90E-15
60	comp187777_g1_i4	SSB_CAUCR	3.30E-07
61	comp159485_g1_i2	DB10_NICSY	6.60E-53
62	comp153239_g1_i2	UFAA1_MYCTU	2.60E-50
63	comp172645_g1_i3	Hypothetical protein, putative	8.70E-19
64	comp171933_g1_i3	Y471_RICPRRecName: Full=Uncharacterized protein RP471	1.00E-13
65	comp188023_g4_i1	EKI_ARATH	9.10E-49
66	comp184636_g1_i6	solute carrier family 25 (mitochondrial folate transporter), member 32	1.30E-15
67	comp194149_g2_i3	TEMPT_APLCA	5.40E-25
68	comp174674_g1_i1	HMCN1_HUMAN	3.90E-08
69	comp195850_g1_i2	Proteophosphoglycan ppg4	2.80E-36
70	comp184555_g1_i5	LONF1_HUMAN	1.10E-43
71	comp186094_g4_i2	PCYAA_EUGLO	0.00E+00
72	comp182382_g1_i2	E3 ubiquitin- ligase rififylin-like	1.10E-08
73	comp178007_g1_i5	DYH1_HUMAN	0.00E+00
74	comp166919_g1_i3	SRPR_DROME	1.00E-118

75	comp161429_g1_i3	LPAT4_ARATH	9.60E-40
76	comp164252_g1_i4	TRAM LAG1 CLN8 homology domain	2.10E-17
77	comp188303_g1_i7	hypothetical protein Ctob_016390	1.50E-08
78	comp168870_g1_i5	CY42_TRYBB	4.30E-13
79	comp184908_g1_i6	phospho phosphatase	1.40E-78
80	comp191293_g1_i2	adenylyl cyclase	2.70E-44
81	comp181083_g2_i3	EXD1_MOUSE	3.00E-09
82	comp187129_g2_i1	ATPG_ODOSI	1.00E-77
83	comp192683_g1_i5	DNJA6_ARATH	9.00E-102
84	comp188423_g1_i6	-glutamine gamma-glutamyltransferase 4	8.00E-73
85	comp192124_g2_i2	PLDA2_ARATH	2.50E-163
86	comp185440_g2_i2	VTI11_ARATH	2.50E-17
87	comp184397_g6_i4	membrane-associated ,	4.50E-147
88	comp189471_g1_i4	kelch diablo	1.30E-16
89	comp174929_g1_i6	RFA1_HUMAN	3.40E-91
90	comp186389_g2_i2	Peptidase lon N-terminal	4.70E-43
91	comp189603_g1_i2	50S ribosomal L31, chloroplastic	1.70E-22
92	comp185647_g1_i6	PREDICTED: uncharacterized protein LOC100828535	8.30E-20
93	comp193119_g1_i5	IF4E3_DANRE	6.90E-13
94	comp174569_g1_i2	AIG2 family	1.10E-10
95	comp194768_g1_i1	HEMH2_ORYSJ	5.40E-126
96	comp190203_g1_i5	MTP9_ARATH	2.10E-37
97	comp192517_g1_i2	FIXL_RHIME	1.20E-13
98	comp192730_g2_i2	Y683_METJARecName: Full=Uncharacterized protein MJ0683	3.60E-09
99	comp163029_g1_i10	citramalate synthase	3.90E-175
100	comp185403_g2_i4	UPF0554 C2orf43 homolog isoform X1	9.00E-33
101	comp189630_g1_i6	P2C10_ORYSJ	1.20E-40
102	comp192179_g1_i18	membrane-bound adenylyl cyclase	6.10E-41
103	comp190368_g1_i6	KAPR_YARLI	1.30E-12
104	comp191327_g2_i2	NUF2_DANRE	3.60E-07
105	comp169913_g1_i2	PTBP3_ARATH	7.00E-08
106	comp189924_g2_i1	X-ray repair cross-complementing 5-like	4.40E-60
107	comp194859_g2_i1	CYAA_LEIDO	1.50E-30
108	comp116378_g1_i2	EFTU_EUGGR	0.00E+00
109	comp194343_g1_i1	CYA1_RHIME	2.70E-18

110	comp190179_g3_i1	cyclic nucleotide-binding	2.00E-22
111	comp181039_g1_i1	RPO2B_TOBAC	3.40E-137
112	comp176030_g1_i2	BACH_HUMAN	6.90E-66
113	comp192446_g1_i7	alpha-L-rhamnosidase	2.20E-63
114	comp182889_g1_i1	hypothetical cytosolic	4.10E-29
115	comp194856_g3_i2	KPRS_HELPY	2.10E-37
116	comp171920_g1_i3	TMEM5_HUMAN	2.90E-12
117	comp182572_g1_i4	PROP_STAAC	6.50E-13
118	comp188575_g1_i3	exostosin family	1.10E-27
119	comp188315_g2_i2	regulator of telomere elongation helicase 1 rtel1,	0.00E+00
120	comp183397_g1_i2	ef hand family	9.50E-138
121	comp194361_g1_i3	hypothetical protein EMIHUDRAFT_237418	1.70E-09
122	comp177739_g1_i2	adenylyl cyclase	1.70E-08
123	comp194898_g1_i8	hypothetical protein ACA1_171730	5.30E-11
124	comp172899_g2_i6	folate pteridine transporter	9.10E-16
125	comp174652_g1_i2	NLRC3_HUMAN	6.40E-35
126	comp179132_g2_i1	VATL_HELVI	6.10E-17
127	comp183962_g1_i10	UDP-galactose:fucoside alpha-3-galactosyltransferase	1.60E-10
128	comp191840_g2_i2	phosphatidylinositol-4-phosphate-5-kinase-likep ro tein	5.80E-60
129	comp171335_g1_i2	flagellar basal body	4.30E-20
130	comp187964_g3_i4	predicted protein	6.30E-14
131	comp175274_g1_i2	hypothetical protein	3.60E-13
132	comp194269_g1_i7	MSH1_ARATH	5.90E-85
133	comp190288_g2_i6	QKY_ARATH	3.90E-23
134	comp179109_g1_i9	Y575_SYNY3	8.00E-23
135	comp187411_g2_i6	LOW PSII ACCUMULATION 3, chloroplastic	3.80E-40
136	comp166668_g2_i5	YG75_SCHPO	1.00E-37
137	comp180058_g1_i1	peptidoglycan-binding domain-containing	4.50E-16
138	comp186014_g2_i2	6GAL_HYPRU	3.50E-27
139	comp185191_g1_i4	ZC21C_MOUSE	2.60E-11
140	comp195863_g1_i11	adenylate cyclase	5.60E-13
141	comp192939_g1_i8	alpha subunit of photoactivated adenylyl cyclase	3.80E-43
142	comp157038_g1_i2	PPA5_RABIT	8.10E-08
143	comp171390_g1_i3	hypothetical protein AMJ87_05245	2.70E-08
144	comp182315_g1_i1	polyadenylate-binding -interacting 11-like	1.90E-36

145	comp192979_g1_i8	Heat shock	0.00E+00
146	comp190899_g1_i5	Y663_CHLTE	1.20E-11
147	comp179643_g1_i1	CYA1_RHIME	2.20E-23
148	comp192233_g1_i11	YAFJ_ECOLI	2.00E-76
149	comp157913_g1_i4	SSM4 like ring finger	1.80E-11
150	comp146874_g1_i1	CCL11_ORYSJ	1.00E-31
151	comp190751_g1_i3	NID1_MOUSE	2.30E-07
152	comp177600_g2_i3	hypothetical protein AOL_s00054g109	3.10E-09
153	comp187185_g1_i3	UFSP2_MOUSE	1.40E-55
154	comp194964_g1_i5	adenylate guanylate cyclase	2.10E-08
155	comp190752_g2_i4	metallophosphoesterase	7.90E-156
156	comp184056_g1_i2	amino acid	8.50E-29
157	comp187148_g3_i3	ATP-binding cassette transporter,	7.40E-57
158	comp190925_g1_i4	UVB31_ARATH	1.20E-07
159	comp185845_g2_i9	VATL2_CANGA	5.70E-26
160	comp172444_g1_i4	tRNA rRNA methyltransferase	1.20E-63
161	comp180616_g3_i7	THI4_SCHPO	8.70E-74
162	comp181202_g1_i6	FKBP3_MOUSE	5.90E-29
163	comp188599_g3_i1	DNA topoisomerase 3-beta-1	2.10E-47
164	comp180616_g1_i2	thiazole biosynthetic enzyme	2.20E-101
165	comp186814_g1_i2	ABC transporter ATP-binding	8.40E-71
166	comp176506_g1_i1	uracil-DNA glycosylase	4.30E-32
167	comp194695_g1_i6	TRI37_MOUSE	4.60E-89
168	comp177425_g1_i3	NLRC3 isoform X1	9.00E-15
169	comp166556_g1_i12	DYRK2_DICDI	4.40E-63
170	comp188606_g1_i4	F264_RAT	8.50E-57
171	comp190631_g2_i1	MTMR2_HUMAN	2.00E-91
172	comp189557_g1_i6	CYAA_PODAS	1.40E-25
173	comp176864_g1_i3	Y233_STRP6RecName: Full=Uncharacterized protein M6_Spy0233	1.10E-08
174	comp184170_g1_i5	mannosyl-oligosaccharide 1,2-alpha-mannosidase	1.40E-128
175	comp176825_g1_i2	predicted protein	8.30E-18
176	comp190506_g1_i1	lipolytic G-D-S-L family	1.10E-09
177	comp161095_g1_i2	PHO88_YEAST	8.00E-13
178	comp179938_g1_i1	hypothetical protein SDRG_13758	1.90E-09
179	comp189916_g2_i2	F16P1_BRANA	1.50E-79

180	comp194100_g1_i5	HPT1_ARATH	5.70E-38
181	comp181083_g3_i2	EXD1_MOUSE	4.60E-09
182	comp188052_g2_i3	predicted protein, partial	9.50E-24
183	comp190889_g1_i9	DHKJ_DICDI	8.60E-25
184	comp181527_g1_i3	ASPG1_ECO57	5.80E-66
185	comp182105_g1_i1	Y4757_DICDI	1.30E-08
186	comp182044_g1_i1	NEK3_HUMAN	2.60E-19
187	comp194458_g2_i3	hypothetical protein DQ04_01931010	6.00E-15
188	comp183559_g1_i3	TUB_HUMAN	8.70E-72
189	comp175064_g1_i5	FCA1_TRYRA	1.00E-20
190	comp187103_g2_i2	predicted protein	3.70E-20
191	comp170829_g3_i1	permease	1.30E-10
192	comp166696_g1_i2	triose phosphate phosphate translocator	2.20E-23
193	comp193858_g3_i3	Y1181_ARATH	7.30E-61
194	comp190240_g1_i5	DDX3_DROME	1.20E-15
195	comp168870_g1_i2	CY42_TRYBB	5.00E-13
196	comp187411_g2_i1	LOW PSII ACCUMULATION chloroplastic	1.00E-40
197	comp183844_g2_i3	AMFR_MOUSE	1.40E-19
198	comp187424_g2_i3	glutathione S-transferase	6.20E-10
199	comp184179_g2_i3	RFWD2_MOUSE	2.30E-07
200	comp184547_g2_i6	F-box-like WD repeat-containing TBL1X	2.00E-152
201	comp189904_g1_i2	MSRB5_ARATH	2.00E-40
202	comp160609_g1_i3	NSMA2_MOUSE	4.60E-08
203	comp184855_g1_i2	embryogenesis-associated EMB8	2.10E-71
204	comp189117_g1_i2	CBPC1_CAEEL	1.60E-98
205	comp192976_g1_i6	ASHR1_ARATH	2.30E-08
206	comp169519_g1_i3	alpha-soluble nsf attachment	3.30E-69

Appendix table 8 genes name of area 8

(a) List of TF family (19 transcripts of 16 TF families)

No.	Seqname of Euglena	Annotation	Evalue
1	comp173735_g2_i1	MIKC	5.00E-51
2	comp188263_g1_i3	FAR1	1.00E-40
3	comp161284_g2_i5	GRAS	4.80E-18
4	comp174907_g1_i2	LBD	7.80E-15
5	comp169716_g2_i2	YABBY	1.70E-30
6	comp173208_g1_i1	TALE	2.00E-37
7	comp167367_g1_i3	B3	2.50E-51
8	comp195198_g1_i2	C3H	9.20E-92
9	comp189100_g1_i6	FAR1	3.90E-26
10	comp166380_g1_i2	STAT	2.10E-16
11	comp189385_g1_i2	YABBY	2.60E-26
12	comp169611_g1_i1	bHLH	4.60E-12
13	comp181747_g1_i2	BBR-BPC	9.50E-22
14	comp183358_g2_i4	Nin-like	1.50E-09
15	comp189839_g1_i2	NF-YA	1.30E-08
16	comp165526_g1_i2	FAR1	3.70E-32
17	comp175353_g1_i3	WOX	4.50E-33
18	comp170518_g1_i1	WRKY	1.80E-105
19	comp163298_g2_i1	bZIP	1.20E-164

(b) List of other genes name (64 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp184665_g3_i3	phosphoglycerate mutase	7.60E-27
2	comp140221_g1_i2	RR4_EUGGR	2.00E-132
3	comp196119_g2_i5	NOTC1_HUMAN	8.50E-50
4	comp195512_g1_i13	predicted protein	2.50E-09
5	comp181483_g3_i1	TPA: RING zinc finger domain superfamily	1.10E-12
6	comp174505_g3_i2	HOS4_YEAST	1.50E-07
7	comp194535_g1_i2	chromosome partitioning	2.00E-09
8	comp171042_g1_i5	chloride channel 2	3.60E-63

9	comp195127_g1_i3	probable beta-1,3-galactosyltransferase 19	2.80E-16
10	comp194293_g1_i5	IP6K3_MOUSE	2.70E-30
11	comp170544_g1_i2	SGNH hydrolase	1.10E-29
12	comp185637_g1_i1	SAM-dependent methyltransferase	3.00E-09
13	comp174958_g1_i11	cystatin-A	1.00E-19
14	comp190380_g2_i1	NAD-dependent deacetylase sirtuin-6	1.10E-25
15	comp192036_g1_i3	peptidase	4.60E-42
16	comp138222_g1_i1	CTR1_SOLIN	1.40E-61
17	comp185260_g1_i7	KAP0_BOVIN	1.10E-32
18	comp180639_g2_i4	Serine threonine kinase	2.20E-34
19	comp189696_g3_i3	hypothetical protein EMIHUDRAFT_240965	4.50E-08
20	comp172016_g4_i6	DPOE2_CHICK	2.70E-80
21	comp185603_g2_i1	homeodomain transcription factor 2 isoform X1	2.60E-43
22	comp185654_g4_i3	ubiquitin carboxyl-terminal hydrolase 49	1.60E-43
23	comp187053_g1_i3	F264_RAT	6.00E-57
24	comp195732_g1_i4	UPL3_ARATH	1.10E-15
25	comp193720_g2_i2	DUF1683 family	2.90E-33
26	comp191912_g1_i7	AARA_DICDI	4.50E-11
27	comp187553_g3_i2	STR14_ARATH	5.50E-25
28	comp190368_g2_i1	cyclic nucleotide-binding	1.80E-19
29	comp180162_g1_i1	expressed unknown protein	4.80E-55
30	comp181772_g2_i3	SAMP_MESAU	7.80E-08
31	comp180182_g1_i2	hypothetical protein PPTG_04275	1.70E-49
32	comp183789_g1_i3	YY46_NOSS1	1.80E-19
33	comp195833_g1_i1	ALA3_ARATH	1.30E-47
34	comp174776_g1_i3	replication A 32 kDa subunit	2.30E-30
35	comp187065_g1_i4	AK1_DICDI	2.50E-20
36	comp161295_g2_i2	BACD3_HUMAN	5.20E-15
37	comp192670_g1_i3	Kinesin KIF13A	5.70E-29
38	comp144388_g1_i4	ANK2_HUMAN	4.60E-18
39	comp192253_g2_i2	CJ011_MOUSE	1.50E-14
40	comp187615_g1_i1	PFLA_CLOPA	8.30E-31
41	comp193640_g1_i4	FIXL_RHIME	5.90E-08
42	comp162177_g1_i1	LPAT4_ARATH	2.70E-39
43	comp191570_g1_i3	hypothetical protein PBRA_008009	5.70E-11

44	comp180927_g1_i3	ESIB_ECOL6	2.10E-18
45	comp178617_g1_i3	glycosyl transferase	3.10E-21
46	comp179825_g1_i3	MUTYH_MOUSE	3.30E-78
47	comp168326_g1_i2	PP1R8_BOVIN	2.10E-17
48	comp150399_g1_i3	YCX9_EUGGR	0.00E+00
49	comp171450_g1_i21	MAT1_EUGGR	0.00E+00
50	comp172015_g1_i6	WD repeat-containing 43-like	1.40E-08
51	comp161632_g1_i1	PH domain-containing	1.40E-19
52	comp182658_g1_i3	TKL kinase	3.00E-22
53	comp166724_g1_i1	VKOR_SYNJB	9.10E-09
54	comp190070_g2_i2	SPNS1_ARATH	3.20E-37
55	comp191751_g1_i4	OTU7A_MOUSE	3.00E-25
56	comp180865_g1_i3	phospholipid:diacylglycerol acyltransferase	6.20E-136
57	comp168376_g1_i1	glucosephosphate isomerase	0.00E+00
58	comp175834_g1_i2	glycosyl transferase family GT77 cell wall polysaccharide biosynthesis	7.20E-27
59	comp149586_g1_i5	AB1F_ARATH	0.00E+00
60	comp172657_g1_i3	NAD-dependent deacetylase sirtuin-2	2.00E-90
61	comp166102_g1_i2	hypothetical protein	1.10E-27
62	comp176088_g1_i2	ACY1_DICDI	6.90E-132
63	comp186281_g1_i2	STT7_CHLRE	4.10E-90
64	comp183613_g3_i12	IF4E_RAT	3.00E-45

Appendix table 9 genes name of area 9

(a) List of TF family of part A (60 transcripts of 22 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193265_g1_i4	YABBY	2.90E-30
2	comp194235_g1_i4	NAC	4.10E-11
3	comp179973_g1_i5	MYB_related	8.50E-13
4	comp194693_g1_i3	bHLH	3.40E-09
5	comp181262_g1_i1	BES1	3.00E-21
6	comp189590_g1_i2	GRAS	7.20E-08
7	comp190449_g2_i1	GeBP	8.00E-10
8	comp192755_g1_i4	TCP	4.10E-71
9	comp190881_g1_i2	bHLH	8.70E-80
10	comp187693_g1_i5	LBD	6.70E-120
11	comp194461_g3_i2	B3	8.50E-10
12	comp171050_g1_i2	C3H	1.10E-18
13	comp194772_g1_i2	bHLH	1.20E-10
14	comp187070_g2_i2	YABBY	6.70E-24
15	comp194235_g1_i5	NAC	3.70E-11
16	comp191565_g2_i3	bHLH	1.00E-18
17	comp171050_g1_i3	C3H	1.20E-18
18	comp186915_g1_i1	E2F_DP	8.20E-17
19	comp191923_g1_i1	WRKY	1.40E-18
20	comp191904_g1_i1	bHLH	1.10E-25
21	comp194899_g3_i7	B3	8.60E-08
22	comp171050_g1_i4	C3H	9.80E-19
23	comp192966_g2_i3	G2-like	1.30E-40
24	comp190700_g1_i1	GeBP	3.40E-09
25	comp195262_g1_i2	NAC	1.80E-12
26	comp193265_g1_i6	YABBY	6.40E-30
27	comp195081_g1_i10	bHLH	2.30E-14
28	comp193263_g1_i10	MYB_related	4.70E-58
29	comp195082_g1_i3	Nin-like	7.80E-14
30	comp195091_g2_i2	B3	1.30E-56
31	comp193452_g1_i9	MYB_related	1.00E-57
32	comp140357_g1_i1	WOX	1.10E-43

33	comp184108_g1_i5	YABBY	8.60E-27
34	comp193370_g2_i1	YABBY	2.00E-23
35	comp192128_g4_i13	WRKY	1.20E-41
36	comp178096_g1_i1	NF-YB	3.30E-63
37	comp193852_g1_i1	YABBY	1.10E-27
38	comp192208_g1_i5	WRKY	2.30E-20
39	comp190700_g1_i10	GeBP	3.50E-09
40	comp195502_g1_i3	B3	3.00E-39
41	comp172005_g1_i2	C3H	5.10E-28
42	comp193281_g1_i1	G2-like	1.00E-20
43	comp187949_g1_i2	LBD	3.80E-18
44	comp162401_g1_i1	WOX	3.20E-78
45	comp192111_g1_i2	bHLH	8.30E-21
46	comp180068_g1_i2	MYB_related	3.20E-37
47	comp170769_g1_i12	GATA	3.10E-07
48	comp189417_g1_i1	NF-YA	9.60E-09
49	comp187479_g1_i1	C3H	3.00E-83
50	comp188880_g1_i1	GeBP	2.40E-08
51	comp183797_g1_i2	NF-YB	9.20E-105
52	comp195803_g2_i2	B3	5.30E-12
53	comp172414_g2_i1	C3H	3.10E-10
54	comp193281_g1_i6	G2-like	1.00E-20
55	comp180231_g1_i2	HB-other	1.20E-08
56	comp162401_g1_i2	WOX	2.40E-78
57	comp192497_g1_i1	WRKY	1.10E-09
58	comp193852_g1_i11	YABBY	1.20E-27
59	comp187643_g1_i2	Trihelix	6.60E-16
60	comp191709_g3_i2	bZIP	5.40E-09

(b) List of other genes name of part A (188 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp194695_g1_i1	E3 ubiquitin- ligase TRIM37predicted protein, partial	3.60E-90
2	comp179403_g1_i7	MYBPP_HUMAN	1.00E-11
3	comp180616_g3_i8	THI4_SCHPO	6.50E-74

4	comp191510_g1_i3	predicted protein, partial	7.10E-43
5	comp179393_g1_i2	appr-1-p processing domain	7.90E-40
6	comp175871_g1_i2	predicted protein, partial	3.80E-64
7	comp189522_g2_i5	MFS transporter	3.50E-09
8	comp102495_g1_i1	MIFH_TRITR	2.30E-11
9	comp269188_g1_i1	RS29_WHEAT	1.70E-15
10	comp185179_g2_i1	SPNA_DICDI	7.80E-31
11	comp180056_g2_i2	RWA1_ARATH	9.70E-112
12	comp189163_g1_i3	PIGL_DICDI	3.80E-40
13	comp175251_g1_i2	tyrosine phosphatase	9.50E-105
14	comp99114_g1_i2	hypothetical protein Ctob_007964	8.80E-17
15	comp195855_g1_i2	SMC6_DICDI	1.00E-44
16	comp180501_g1_i2	hypothetical protein PBRA_000501	6.10E-19
17	comp192793_g3_i3	erythroid differentiation-related factor 1	4.60E-41
18	comp194361_g2_i2	hypothetical protein EMIHUDRAFT_222043	2.20E-09
19	comp192486_g1_i12	Proteasome, alpha subunit	1.60E-19
20	comp168939_g3_i1	hypothetical protein Ctob_006924	8.10E-24
21	comp155420_g1_i1	NLRC3_HUMAN	3.20E-40
22	comp186240_g1_i4	tubulin binding	4.30E-33
23	comp189074_g4_i4	tubulin polyglutamylase TTLL4-like	8.10E-48
24	comp188575_g1_i5	exostosin family	1.20E-27
25	comp191570_g2_i1	hypothetical protein PBRA_008009	4.20E-11
26	comp168109_g1_i1	hypothetical protein M569_17729	7.30E-42
27	comp193187_g1_i1	HMA2_ARATH	7.00E-44
28	comp191887_g2_i6	FUT8_HUMAN	6.00E-13
29	comp183962_g1_i4	RAY1_ARATH	1.40E-09
30	comp166774_g1_i4	polypyrimidine tract-binding homolog 2-like	1.10E-08
31	comp194698_g1_i1	PRMS_MAIZE	2.10E-11
32	comp165914_g1_i2	SYFM_HUMAN	4.60E-56
33	comp191293_g2_i3	CYAG_DICDI	1.80E-38
34	comp191847_g2_i1	TSGAD_HALVD	4.50E-40
35	comp162660_g1_i2	DBP5_LODEL	4.80E-48
36	comp189197_g3_i3	kinase domain	1.40E-20
37	comp191827_g1_i1	UVR3_ORYSJ	5.20E-128
38	comp186398_g1_i3	Y328_SYNY3	4.00E-26

39	comp169706_g1_i2	TAR3_ARATH	6.10E-71
40	comp194215_g1_i7	hypothetical protein AURANDRAFT_67549	3.20E-13
41	comp178883_g1_i7	phosphoglycerate mutase family domain containing	8.00E-16
42	comp191790_g1_i4	NEDD4-like E3 ubiquitin- ligase WWP2 isoform X3	3.50E-39
43	comp189471_g3_i6	KLHDB_DROSI	2.00E-18
44	comp187208_g1_i1	DUS1_ARATH	2.60E-27
45	comp185409_g1_i5	ATPbinding Cassette (ABC) superfamily	1.60E-08
46	comp195334_g2_i1	glutathione-s-transferase (ISS)	2.80E-31
47	comp189181_g2_i1	NLRC3_MOUSE	4.90E-40
48	comp169152_g1_i2	ADPRM_ORYSJ	1.80E-66
49	comp175280_g1_i2	Cytosolic carboxypeptidase 1predicted protein, partial	1.80E-89
50	comp194947_g1_i4	magnesium chelatase	0.00E+00
51	comp189059_g1_i11	STK11_CHICK	3.60E-53
52	comp186389_g1_i5	Peptidase S16, lon N-terminal	7.40E-41
53	comp182698_g1_i3	isocitrate dehydrogenase	0.00E+00
54	comp178883_g1_i4	phosphoglycerate mutase family domain containing	8.00E-16
55	comp185144_g1_i2	WD-40 repeat	2.40E-31
56	comp191912_g1_i8	AARA_DICDI	1.20E-10
57	comp181635_g1_i2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase eta-2	1.10E-80
58	comp192215_g1_i5	NUCL_XENLA	3.90E-14
59	comp164858_g1_i2	ARIA_ARATH	3.10E-09
60	comp183647_g3_i3	membrane-associated ,	4.50E-08
61	comp191936_g1_i2	UDP-galactose:fucoside alpha-3-galactosyltransferase	4.20E-23
62	comp188375_g1_i2	phosphatidylinositol 3-kinase, root isoform isoform X1	2.50E-167
63	comp190696_g1_i1	Centrosomal of 104 kDapredicted protein, partial	6.20E-14
64	comp188567_g1_i1	conserved hypothetical protein	1.20E-39
65	comp188369_g1_i2	TDA3_YEAST	2.80E-16
66	comp193767_g1_i5	receptor-type adenylate cyclase a	6.20E-31
67	comp190973_g2_i11	TPR domain-containing	1.40E-11
68	comp178385_g1_i5	ubiquinone biosynthesis	2.00E-81
69	comp179701_g1_i8	duf676 domain containing hydrolase	1.50E-21
70	comp186796_g1_i2	TMEM5_HUMAN	4.60E-14
71	comp161194_g1_i1	FACT complex subunit SSRP1-like	1.20E-44
72	comp182118_g1_i10	Armadillo betacatenin-like repeat domain containing	1.40E-24
73	comp195163_g1_i6	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07

74	comp211196_g1_i1	predicted protein	9.50E-41
75	comp177733_g1_i2	adenylate cyclase	5.90E-09
76	comp195975_g2_i1	HERC4_MOUSE	8.20E-14
77	comp175260_g1_i2	membrane-associated protein, putative	1.30E-86
78	comp180535_g1_i3	PR40A_ARATH	2.20E-49
79	comp171929_g1_i1	neuralized 4	2.40E-30
80	comp193653_g1_i7	adenylate cyclase	3.00E-12
81	comp194969_g1_i1	AT133_HUMAN	4.80E-32
82	comp186389_g1_i4	Peptidase lon N-terminal	4.30E-29
83	comp184790_g1_i2	HEAT repeat containing 2	3.90E-150
84	comp167503_g1_i1	INT11_DICDI	7.60E-174
85	comp168035_g1_i4	receptor-type adenylate cyclase a	1.60E-14
86	comp189274_g1_i4	NUCL_HUMAN	1.50E-14
87	comp172640_g1_i3	threonine dehydratase	3.70E-158
88	comp191026_g1_i2	MGDG_TOBAC	1.90E-60
89	comp178380_g1_i4	membrane-associated protein, putative	2.30E-10
90	comp180477_g1_i2	YU88_ARATH	2.20E-29
91	comp175716_g1_i7	Tetratricopeptide-like helical	3.00E-46
92	comp188956_g1_i6	vacuolar sorting vps16	1.90E-171
93	comp158051_g1_i1	CYTB_OPLFA	7.70E-18
94	comp159854_g1_i1	hypothetical protein GLOINDRAFT_312092	2.30E-19
95	comp171929_g1_i5	NEUL4_HUMAN	2.20E-11
96	comp182446_g1_i10	metallophosphoesterase domain-containing 1	2.00E-62
97	comp186187_g1_i3	CHO2_ASPFU	4.80E-101
98	comp193119_g1_i2	eukaryotic translation initiation factor 4E type 3	3.80E-16
99	comp189431_g1_i2	calcium-binding mitochondrial carrier S -3	6.70E-59
100	comp187136_g2_i7	glycine betaine ABC transporter	2.10E-27
101	comp159148_g1_i2	pterin-4-alpha-carbinolamine dehydratase	3.70E-17
102	comp133101_g1_i1	ATPA_EUGGR	0.00E+00
103	comp194330_g1_i4	Ankyrin repeat	5.30E-89
104	comp170974_g1_i1	Y1758_AQUAE	1.40E-22
105	comp180626_g1_i21	DYRK2_DICDI	1.50E-63
106	comp186398_g1_i9	Y328_SYNY3	2.70E-26
107	comp183363_g1_i2	hypothetical protein PBRA_000501	2.00E-24
108	comp195163_g1_i8	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.70E-07

109	comp181360_g1_i9	KIN10_ARATH	1.20E-125
110	comp186524_g4_i3	VKOR_SYNJB	3.50E-07
111	comp193325_g1_i12	DEAD DEAH box helicase	9.90E-14
112	comp191641_g1_i8	hypothetical protein COCSUDRAFT_45614	1.10E-14
113	comp187660_g1_i5	C2 domain-containing	5.40E-14
114	comp190688_g1_i3	Aspartic peptidase	4.20E-21
115	comp192874_g2_i4	orofacial cleft 1 candidate gene 1	2.40E-28
116	comp193962_g1_i2	CY42_TRYBB	1.90E-14
117	comp183767_g1_i3	efflux pump,	9.70E-70
118	comp192932_g1_i8	E3 ubiquitin- ligase HECTD1 isoform X4	4.40E-17
119	comp185954_g1_i2	RNA ligase	1.00E-74
120	comp182068_g1_i6	Hypothetical protein, putative	3.40E-09
121	comp184946_g1_i7	MGAT3_RAT	6.30E-21
122	comp194370_g1_i3	YELD_SCHPO	2.10E-28
123	comp184449_g1_i1	ENPP3_MOUSE	1.90E-14
124	comp191381_g2_i6	OTU family cysteine protease	9.20E-17
125	comp193264_g1_i6	ASO_CUCPM	5.90E-74
126	comp186752_g1_i5	nicotinate-nucleotide adenylyltransferase	1.10E-52
127	comp186529_g1_i8	CHO2_ASPFU	2.00E-101
128	comp170304_g1_i1	tetratricopeptide repeat partial	2.60E-69
129	comp178764_g3_i2	YGRG_SCHPO	1.60E-33
130	comp189758_g2_i8	AARA_DICDI	4.90E-11
131	comp178843_g1_i1	RDMC_STREF	8.50E-27
132	comp179596_g1_i4	signal peptide peptidase-like, aspartyl protease family A22B,	7.90E-38
133	comp182081_g1_i8	I5P2_HUMAN	3.70E-45
134	comp172240_g2_i2	DPOE2_CHICK	6.50E-82
135	comp183367_g1_i3	TNI3K_MOUSE	2.60E-10
136	comp167168_g1_i1	PIRL5_ORYSJ	2.30E-12
137	comp183468_g1_i2	KMO_XENTR	5.80E-77
138	comp189110_g1_i2	NLRC3_HUMAN	1.20E-14
139	comp189876_g4_i5	predicted protein	1.10E-17
140	comp189703_g1_i3	MTAP2_RAT	2.40E-12
141	comp183158_g1_i2	T184C_BOVIN	5.40E-46
142	comp174253_g2_i1	PSTS_RHILO	3.50E-12
143	comp178763_g1_i3	TTD14_DROME	2.40E-34

144	comp178962_g3_i5	Y771_METTH	1.60E-18
145	comp194281_g1_i2	hybrid sensor histidine kinase response regulator	1.00E-20
146	comp190339_g3_i1	pilus assembly	5.50E-85
147	comp187108_g1_i1	NLRC3 isoform X1	3.70E-21
148	comp188155_g3_i3	hypothetical protein MVEG_05746	1.40E-34
149	comp167679_g1_i1	folate pteridine transporter	1.20E-14
150	comp183648_g1_i1	NLRC3_HUMAN	4.30E-36
151	comp174989_g1_i2	probable S-acyltransferase 4	9.90E-44
152	comp183693_g1_i1	lipase family	5.80E-65
153	comp188807_g1_i1	unnamed protein product	5.70E-28
154	comp195081_g2_i6	Dual specificity phosphatase 1, isoform 1	3.10E-17
155	comp193410_g1_i3	MSH1_ARATH	2.90E-85
156	comp180755_g1_i1	secreted partial	4.50E-12
157	comp185831_g2_i5	methylase	8.50E-19
158	comp185399_g1_i2	DNRP_STRPE	2.40E-27
159	comp189137_g2_i11	ZNFX1_MOUSE	5.60E-59
160	comp193347_g3_i3	gonidia-specific KA_k47	1.20E-29
161	comp130265_g1_i2	DNA-directed RNA polymerase II subunit RPB7	3.40E-48
162	comp179131_g1_i4	bile Acid:Na+ symporter family	1.70E-39
163	comp188444_g6_i5	tRNA pseudouridine synthase A, mitochondrial isoform X3	4.90E-35
164	comp173443_g1_i6	CRTY_ARTPT	1.10E-94
165	comp185933_g2_i3	NLRC3_HUMAN	7.20E-27
166	comp190669_g2_i8	DDX51_DANRE	1.60E-79
167	comp167805_g1_i2	KAD_DESMR	2.90E-67
168	comp176797_g1_i1	ELAV 3 isoform X8	4.70E-21
169	comp185999_g4_i10	ILKAP_BOVIN	5.00E-08
170	comp161092_g1_i1	SETD3_MOUSE	2.10E-07
171	comp164409_g1_i1	LPAT1_ARATH	1.10E-57
172	comp194270_g3_i1	ERYA3_SACER	3.50E-40
173	comp191263_g3_i6	predicted protein	1.00E-16
174	comp166774_g1_i5	polypyrimidine tract-binding homolog 2-like	1.20E-08
175	comp189424_g2_i2	PTGR2_PONAB	3.40E-66
176	comp188155_g3_i2	hypothetical protein MVEG_05746	6.30E-35
177	comp98361_g1_i1	IM30_PEA	1.40E-09
178	comp185063_g3_i2	hypothetical protein SARC_00180	9.10E-149

179	comp192177_g1_i10	KAPR_ASPFU	7.40E-08
180	comp179716_g1_i3	CCD39_CHLRE	1.90E-105
181	comp176797_g1_i3	ELAV2_PONAB	2.60E-20
182	comp183254_g1_i1	probable methyltransferase TARBP1	8.80E-10
183	comp189059_g1_i7	STK11_CHICK	9.30E-54
184	comp182673_g2_i1	VAC8_CRYNB	1.00E-09
185	comp188679_g1_i3	vacuolar 8	2.60E-30
186	comp178988_g1_i5	hypothetical protein SPRG_14035	3.90E-19
187	comp186900_g1_i2	NNTM_HUMAN	3.30E-09
188	comp180076_g1_i2	regulator of chromosome condensation RCC1	1.10E-23

(c) List of TF family of part B (3 transcripts of 3 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp194087_g1_i4	C3H	3.90E-16
2	comp184042_g4_i2	GeBP	4.90E-08
3	comp189839_g1_i1	NF-YA	1.30E-08

(d) List of other genes name of part B (34 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp182448_g1_i7	pyruvate kinase	2.10E-123
2	comp195118_g1_i7	E3 ubiquitin- ligase RLIM-like isoform X1	1.30E-15
3	comp193573_g1_i10	hypothetical protein AURANDRAFT_62073	4.00E-15
4	comp192965_g5_i2	hypothetical protein Naga_100042g35	5.60E-08
5	comp192076_g1_i6	breast cancer early onset (ISS)	1.10E-27
6	comp187966_g1_i2	PUB1_SCHPO	2.60E-44
7	comp179825_g1_i4	MUTYH_MOUSE	9.10E-78
8	comp159355_g1_i2	RHBGB_XENLA	6.80E-79
9	comp168035_g1_i5	CY42_TRYBB	1.00E-13
10	comp194154_g1_i2	PDE1A_MOUSE	5.20E-60
11	comp167702_g1_i2	GLAK2_ARATH	1.10E-79
12	comp5784_g1_i1	NAD-dependent alcohol dehydrogenase	1.80E-09
13	comp191053_g1_i3	TRPV6_HUMAN	3.90E-13
14	comp180056_g1_i2	CAS1 domain-containing 1-like	4.10E-124

15	comp182293_g3_i2	kelch domain-containing 3	2.00E-22
16	comp185293_g1_i11	metallophosphoesterase domain-containing 1	4.80E-61
17	comp183252_g2_i3	nucleic acid-binding	1.70E-08
18	comp176517_g2_i1	FLNA_DROME	1.00E-12
19	comp175670_g1_i2	ILVB_SCHPO	7.80E-176
20	comp169667_g1_i2	S20A2_XENLA	7.30E-30
21	comp193064_g1_i3	hypothetical protein	2.00E-58
22	comp195012_g1_i3	p-type atpase (p-atpase) superfamily	1.50E-59
23	comp195653_g1_i3	glutamine-fructose-6-phosphate transaminase (isomerizing)	3.50E-179
24	comp189197_g1_i3	NLRC3_HUMAN	1.30E-07
25	comp192998_g1_i2	PDC6I_XENLA	3.90E-14
26	comp176269_g1_i3	disulfide-isomerase A3	3.70E-09
27	comp190245_g1_i1	OGG1_ARATH	2.00E-59
28	comp184567_g2_i1	SYK_CRIGR	0.00E+00
29	comp176827_g1_i3	muscle-specific 20-like	5.60E-14
30	comp189538_g1_i11	FIXL_RHIME	1.20E-13
31	comp187961_g1_i6	ELAV4_RAT	9.10E-21
32	comp191505_g2_i1	uroporphyrinogen decarboxylase	8.10E-123
33	comp170275_g1_i12	Poly (A) RNA polymerase cid14	3.40E-51
34	comp192586_g5_i2	membrane-associated ,	5.00E-45

Appendix table 10 genes name of area 10

(a) List of TF family (9 transcripts of 8 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193129_g2_i1	B3	1.80E-20
2	comp168285_g1_i1	C3H	4.60E-28
3	comp163951_g2_i2	C3H	2.40E-12
4	comp190130_g2_i6	WRKY	2.00E-78
5	comp192336_g2_i3	YABBY	3.10E-25
6	comp192382_g1_i2	G2-like	5.60E-21
7	comp189870_g3_i1	C2H2	1.00E-12
8	comp184852_g1_i4	Trihelix	2.50E-14
9	comp189433_g1_i16	GeBP	8.20E-12

(b) List of other genes name (87 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp159075_g1_i2	s-adenosylmethionine mitochondrial carrier	1.90E-15
2	comp151721_g1_i1	WD40 repeat-containing ,	7.00E-08
3	comp188940_g1_i1	COPDA_ARATH	1.40E-69
4	comp191100_g1_i4	ACADV_MOUSE	3.80E-19
5	comp193459_g1_i5	hypothetical protein	2.80E-08
6	comp179652_g1_i3	transmembrane 18-like	2.10E-10
7	comp155310_g1_i2	CBS domain containing	1.10E-18
8	comp190752_g2_i1	metallophosphoesterase	2.50E-156
9	comp182658_g1_i6	TKL kinase	2.40E-22
10	comp180180_g1_i1	LDHD_MOUSE	2.80E-93
11	comp194459_g1_i6	CAN7_HUMAN	1.40E-92
12	comp168939_g1_i2	hypothetical protein Ctob_006924	8.90E-24
13	comp192841_g2_i6	hypothetical protein Ctob_010698	9.70E-08
14	comp192134_g1_i1	EGY1_ARATH	1.50E-26
15	comp179513_g1_i1	radical SAM	7.50E-59
16	comp184026_g1_i1	autoinducer 2 sensor kinase phosphatase luxQ	4.80E-17
17	comp164457_g1_i2	DOHH_DROME	1.60E-71
18	comp174363_g3_i1	KCD17_HUMAN	1.90E-08

19	comp170754_g1_i4	SMYD3_MOUSE	2.70E-07
20	comp195793_g1_i5	Soluble guanylate cyclase gcy-31	1.20E-09
21	comp154819_g1_i1	Zinc finger Ran-binding domain-containing 2	1.70E-09
22	comp178116_g1_i3	SAM-dependent methyltransferase	1.30E-76
23	comp194749_g2_i1	carboxypeptidase Q	8.20E-93
24	comp194338_g1_i9	MFS12_MOUSE	9.60E-46
25	comp192886_g2_i3	SKI2_YEAST	1.10E-27
26	comp179648_g1_i4	peroxisomal membrane family	2.30E-62
27	comp185343_g2_i5	SPNS1_ARATH	2.20E-37
28	comp163062_g1_i4	phosphatase PP2A regulatory subunit B	2.00E-126
29	comp194001_g1_i4	hypothetical protein	9.80E-20
30	comp177589_g1_i2	alpha-type kinase	1.10E-46
31	comp182804_g1_i1	hypothetical protein TTHERM_001054339	9.20E-08
32	comp183531_g3_i6	ADCY8_HUMAN	3.00E-12
33	comp193214_g4_i5	hypothetical protein Ctob_013122	7.70E-54
34	comp195671_g2_i2	FAS2_SCHPO	0.00E+00
35	comp186752_g1_i6	nicotinate-nucleotide adenylyltransferase	9.60E-53
36	comp159950_g1_i2	AK1_DICDI	1.30E-44
37	comp21858_g1_i1	PHB2_DICDI	4.60E-33
38	comp166758_g1_i1	hypothetical protein SDRG_02025	7.60E-15
39	comp162434_g1_i2	HECD2_PONAB	9.50E-113
40	comp194572_g2_i2	OHK1_ORYSJ	6.30E-15
41	comp185788_g1_i17	GDSL-like Lipase Acylhydrolase	1.30E-22
42	comp195435_g1_i10	histidine kinase	3.00E-20
43	comp174499_g2_i2	cell division cycle 123 homolog	1.20E-39
44	comp194992_g1_i2	ankyrin repeat	3.50E-84
45	comp192241_g1_i7	PAS domain-containing sensor histidine kinase	7.40E-22
46	comp167679_g1_i7	folate pteridine transporter	4.70E-16
47	comp186014_g2_i8	6GAL_HYPRU	5.50E-27
48	comp182472_g1_i3	COPDA_ARATH	0.00E+00
49	comp167623_g1_i3	TATA_PROM0	4.40E-10
50	comp159201_g1_i4	KAP0_HUMAN	3.60E-09
51	comp186034_g1_i4	S15 RNA-binding	6.50E-27
52	comp191753_g1_i5	CDSA_SYNY3	7.20E-34
53	comp183382_g1_i6	phosphoglycerate mutase family domain containing	7.80E-16

54	comp187494_g1_i2	BETC_RHIME	6.70E-17
55	comp168632_g1_i3	PFKA_TRYBB	6.00E-143
56	comp195808_g2_i7	mannosylphosphorylation (Mnn4)	7.70E-16
57	comp189516_g1_i4	FIXL_RHIME	1.50E-09
58	comp188539_g1_i2	ACBP4_ARATH	7.90E-15
59	comp181836_g2_i2	PDE1A_MOUSE	4.90E-60
60	comp192486_g1_i9	Proteasome, alpha subunit	1.20E-19
61	comp169328_g1_i7	DRKD_DICDI	1.20E-27
62	comp174011_g2_i12	hydroxyacylglutathione hydrolase	3.50E-112
63	comp180254_g1_i7	COQ4_XENLA	5.30E-64
64	comp174715_g1_i3	beta-type IP39	2.70E-33
65	comp185616_g1_i10	KAP0_BOVIN	8.70E-33
66	comp187593_g3_i2	CTL5_HUMAN	2.70E-33
67	comp188290_g1_i1	hypothetical protein GUITHDRAFT_81225	5.90E-32
68	comp172496_g1_i6	AMMECR1 family	7.50E-41
69	comp184876_g3_i2	Cobyrinic acid a,c-diamide synthase	4.20E-12
70	comp189630_g1_i7	P2C10_ORYSJ	8.70E-41
71	comp181938_g1_i4	ribosome-binding factor A	4.50E-27
72	comp195006_g1_i6	hybrid sensor histidine kinase response regulator	9.90E-21
73	comp170991_g1_i1	glutathione S-transferase theta-1-like	1.00E-28
74	comp189985_g1_i1	unnamed protein product	5.40E-25
75	comp171450_g1_i7	PSBD_EUGGR	1.50E-44
76	comp176899_g1_i3	CSE_ARATH	7.20E-43
77	comp194745_g2_i4	METH_BOVIN	0.00E+00
78	comp181351_g1_i1	YQ51_CAEEL	6.80E-14
79	comp182889_g1_i2	hypothetical cytosolic	3.20E-35
80	comp189538_g1_i1	PAS domain-containing sensor histidine kinase	2.00E-26
81	comp190963_g1_i4	zinc finger family	9.20E-44
82	comp180109_g2_i5	ankyrin repeat-containing ,	4.90E-24
83	comp185257_g1_i1	ALA5_ARATH	2.30E-83
84	comp194258_g4_i2	ADP-ribosylglycohydrolase superfamily	3.20E-52
85	comp175491_g1_i2	SCMC1_DANRE	1.20E-86
86	comp187345_g1_i2	BRO1 domain-containing BROX isoform X1	2.80E-11
87	comp190086_g1_i1	1A1C_MALDO	1.80E-67

Appendix table 11 genes name of area 11

(a) List of TF family (24 transcripts of 16 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp172655_g1_i1	CAMTA	2.30E-20
2	comp180854_g1_i5	FAR1	5.70E-140
3	comp193452_g1_i21	MYB_related	4.40E-58
4	comp191755_g1_i4	NAC	1.10E-69
5	comp193779_g2_i8	MYB_related	6.30E-36
6	comp194509_g2_i1	bHLH	9.90E-85
7	comp188725_g2_i3	B3	1.10E-14
8	comp192725_g1_i1	EIL	5.00E-10
9	comp160570_g1_i3	G2-like	4.40E-09
10	comp178049_g1_i4	HB-other	1.70E-13
11	comp195160_g2_i7	MYB	1.20E-11
12	comp193139_g1_i8	MYB_related	1.20E-28
13	comp194688_g1_i3	Nin-like	2.80E-33
14	comp192838_g2_i5	Nin-like	1.30E-46
15	comp182502_g1_i4	YABBY	3.20E-43
16	comp183451_g2_i2	YABBY	4.00E-31
17	comp185900_g1_i12	bHLH	7.00E-09
18	comp195070_g2_i6	bZIP	1.30E-13
19	comp172950_g1_i1	C3H	3.40E-10
20	comp187592_g3_i6	GRAS	1.70E-18
21	comp188918_g1_i1	MYB_related	5.30E-08
22	comp186290_g1_i8	WRKY	6.50E-115
23	comp188710_g2_i5	WRKY	1.10E-37
24	comp194278_g3_i3	WRKY	7.80E-10

(b) List of other genes name (107 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp155077_g1_i5	hypothetical protein CHLNCDRAFT_145050	8.30E-59
2	comp179043_g1_i1	HERC1_HUMAN	3.80E-09
3	comp185293_g1_i9	MPPD1_HUMAN	4.80E-55

4	comp189792_g1_i4	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit trmt61a-like	1.90E-63
5	comp190609_g1_i3	hypothetical protein PBRA_002490	2.80E-48
6	comp163597_g1_i2	Y1017_SYNY3	3.50E-12
7	comp185416_g1_i3	YC09_KLEPN	7.60E-10
8	comp183926_g2_i1	dynamin family	9.80E-49
9	comp166049_g1_i4	50S ribosomal L21	9.10E-76
10	comp167409_g1_i2	YKX5_SCHPO	1.60E-11
11	comp178350_g2_i5	tetratricopeptide repeat family	4.40E-12
12	comp189170_g1_i3	hypothetical protein PTSG_08670	7.00E-32
13	comp184647_g2_i1	IF4E2_MOUSE	7.40E-15
14	comp171450_g1_i19	PSBC_EUGGR	0.00E+00
15	comp188699_g1_i2	calcium sensing receptor, chloroplastic	6.90E-21
16	comp190381_g7_i5	arylsulfatase A family	1.20E-67
17	comp167633_g1_i1	bacterial leucyl aminopeptidase	1.50E-07
18	comp146646_g1_i4	Uncharacterized conserved	5.40E-19
19	comp184602_g1_i5	CRTI_NARPS	0.00E+00
20	comp127814_g1_i1	COX2_LASSP	1.50E-77
21	comp172899_g2_i7	folate pteridine transporter	3.20E-15
22	comp185224_g1_i3	ubiquitin-activating enzyme E1 1	0.00E+00
23	comp194723_g1_i1	PABP5_ARATH	2.00E-17
24	comp159124_g1_i2	AOX4_ARATH	1.30E-64
25	comp182507_g1_i11	DDL_RALME	2.50E-09
26	comp176130_g1_i2	RH30_ARATH	1.80E-30
27	comp189059_g1_i6	STK11_CHICK	4.30E-53
28	comp173150_g1_i3	hypothetical protein SARC_00571	1.40E-16
29	comp152590_g1_i3	ASAH1_MACFA	1.40E-30
30	comp194572_g1_i1	OHK1_ORYSI	6.00E-15
31	comp192395_g1_i3	QOR_PSEAE	3.10E-92
32	comp188199_g3_i2	receptor-type adenylate cyclase GRESAG 4,	4.20E-47
33	comp180132_g1_i5	OMH4_SCHPO	7.70E-14
34	comp143978_g1_i1	OTU domain-containing 7B	3.50E-30
35	comp187105_g1_i5	POLH_ARATH	3.00E-101
36	comp183799_g6_i2	RENT1_NEUCR	2.40E-10
37	comp167227_g3_i1	PEAR1_HUMAN	4.60E-11
38	comp190776_g1_i1	alpha beta-hydrolase	3.80E-31

39	comp179756_g1_i2	ELOV4_MOUSE	2.80E-09
40	comp171485_g1_i5	ATPB_EUGGR	0.00E+00
41	comp175257_g1_i2	CAAX prenyl protease 2	6.40E-11
42	comp239682_g1_i1	COX3_DROME	4.80E-43
43	comp184137_g1_i4	GGPPS_MUCCL	2.80E-35
44	comp190501_g1_i13	chloroplast fructose-1,6-bisphosphatase	1.20E-175
45	comp195361_g2_i2	LHTL6_ARATH	1.80E-11
46	comp179159_g1_i6	ATP-dependent RNA helicase DED1	0.00E+00
47	comp190776_g1_i2	alpha beta-hydrolase	3.90E-31
48	comp173543_g2_i3	hypothetical protein EMIHUDRAFT_251565	3.10E-09
49	comp184095_g1_i9	DNRP_STRPE	2.50E-27
50	comp180295_g1_i6	manganese transporter	3.30E-145
51	comp192939_g1_i5	PCYAA_EUGGR	9.30E-82
52	comp191132_g2_i1	pantothenate kinase 1	4.20E-52
53	comp172432_g1_i4	glutaminyl-peptide cyclotransferase isoform X2	1.60E-61
54	comp177214_g1_i3	NLRC3_HUMAN	1.40E-28
55	comp189792_g1_i1	TRM61_HUMAN	1.50E-49
56	comp191463_g1_i5	VMAT1_MOUSE	1.80E-17
57	comp193720_g2_i1	TPC11_MOUSE	8.50E-15
58	comp182810_g1_i4	Hypothetical protein, putative	3.00E-09
59	comp192977_g1_i1	APC amino acid permease	3.00E-74
60	comp170807_g2_i4	secreted predicted protein, partial	4.40E-12
61	comp192413_g1_i2	ALA2_ARATH	1.90E-153
62	comp192661_g1_i4	L2HDH_CAEBR	5.70E-100
63	comp163029_g1_i5	CIMA_GEOSL	3.70E-149
64	comp188878_g3_i2	EXPR_XANCP	2.10E-72
65	comp153254_g1_i2	DDX20_HUMAN	4.00E-92
66	comp193243_g2_i2	kinase domain	3.40E-07
67	comp195956_g1_i3	predicted protein, partial	3.00E-20
68	comp163029_g1_i3	citramalate synthase	1.40E-174
69	comp190336_g1_i3	ALG12_ARATH	2.70E-77
70	comp184541_g1_i2	RLUD_NITEU	6.40E-28
71	comp189186_g1_i2	TRA1_SCHPO	4.70E-48
72	comp192295_g1_i1	AARA_DICDI	4.10E-11
73	comp192163_g1_i4	Y1796_ARATH	4.90E-81

74	comp191055_g1_i13	TPR domain-containing	8.40E-12
75	comp156345_g1_i2	NLRC3_HUMAN	7.50E-25
76	comp167232_g1_i3	LIM domain containing	1.20E-13
77	comp194923_g1_i2	pentafunctional AROM polypeptide	0.00E+00
78	comp175659_g1_i4	CC151_MOUSE	3.70E-27
79	comp193577_g1_i5	SPXS1_DICDI	4.90E-45
80	comp195808_g2_i8	mannosylphosphorylation (Mnn4),	1.70E-15
81	comp188859_g1_i5	hypothetical protein H310_00261	1.40E-13
82	comp191890_g1_i2	RDMC_STREF	7.90E-13
83	comp195430_g1_i1	telomerase reverse transcriptase catalytic subunit	2.60E-21
84	comp189671_g1_i2	Y2006_MYCTO	1.60E-105
85	comp176648_g2_i1	NEK7_HUMAN	3.20E-12
86	comp192544_g1_i2	Mucolipidin and related s (TRML subfamily of transient receptor potential s)	2.20E-23
87	comp190482_g3_i2	GUX2_ARATH	1.40E-10
88	comp181548_g1_i1	PP2C_LEICH	2.20E-69
89	comp188564_g1_i2	uracil-DNA glycosylase	4.30E-08
90	comp165267_g1_i2	cyclopropane-fatty-acyl-phospholipid synthase	9.70E-81
91	comp169846_g1_i1	TRXB_DICDI	5.00E-133
92	comp186394_g1_i2	ornithine decarboxylase	2.30E-74
93	comp190804_g1_i7	Transcription factor	1.20E-57
94	comp193264_g1_i5	ASO_CUCPM	2.90E-74
95	comp182507_g1_i12	DDL_RALME	2.60E-09
96	comp196225_g1_i5	adenylate cyclase,	1.10E-29
97	comp176312_g1_i1	YH010_HUMAN	3.40E-16
98	comp186602_g1_i1	FLU_ARATH	7.00E-17
99	comp188686_g1_i8	serine threonine kinase	1.30E-68
100	comp186690_g1_i2	kif3a predicted protein, partial	3.10E-18
101	comp184856_g1_i2	amino acid permease	3.40E-80
102	comp189352_g2_i2	hypothetical protein	1.20E-27
103	comp169645_g1_i3	predicted protein, partial	7.40E-59
104	comp190368_g1_i2	KAPR_YARLI	1.40E-12
105	comp191315_g2_i1	cyclic nucleotide-binding domain	2.30E-18
106	comp142258_g1_i1	DLRB2_MOUSE	5.70E-39
107	comp194897_g1_i3	ankyrin repeat domain-containing 29 isoform X2	5.90E-09

Appendix table 12 genes name of area 12

(a) List of TF family (12 transcripts of 11 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193232_g1_i2	ERF	1.30E-18
2	comp182026_g1_i1	FAR1	7.20E-141
3	comp162887_g1_i3	NF-YB	4.80E-58
4	comp179590_g1_i4	CPP	4.90E-08
5	comp191498_g1_i2	FAR1	7.30E-18
6	comp187693_g1_i4	LBD	1.00E-145
7	comp190010_g1_i1	WRKY	5.90E-36
8	comp161199_g1_i1	C3H	4.90E-07
9	comp195082_g1_i1	Nin-like	7.40E-14
10	comp189435_g1_i2	bHLH	2.00E-09
11	comp192755_g1_i1	TCP	4.20E-71
12	comp193129_g1_i1	B3	1.80E-20

(b) List of other genes name (55 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp187022_g1_i2	ABCA1_MOUSE	2.40E-10
2	comp189361_g1_i2	CCA12_ARATH	6.40E-10
3	comp183265_g1_i2	CSD_SYNY3	7.60E-167
4	comp170644_g1_i1	RBSK_BACSU	5.30E-17
5	comp176985_g1_i5	predicted protein, partial	5.60E-26
6	comp188159_g1_i3	UBC39_ARATH	3.20E-36
7	comp180059_g2_i1	HPGDS_CHICK	5.70E-20
8	comp190760_g1_i2	BRISC and BRCA1-A complex member 1-like	4.10E-14
9	comp195723_g2_i4	CYA1_RHIME	4.80E-24
10	comp181202_g1_i2	FKBP3_MOUSE	6.10E-29
11	comp186161_g2_i2	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like isoform X5	3.90E-70
12	comp181836_g2_i4	PDE1A_MOUSE	2.00E-60
13	comp178988_g1_i1	FYVE-type zinc finger-containing	1.20E-18
14	comp168488_g1_i3	NLRC3_HUMAN	5.90E-36
15	comp172559_g1_i2	NLRC3 isoform X1	2.70E-68

16	comp146107_g1_i3	dihydrolipoamide dehydrogenase	6.10E-52
17	comp187561_g1_i1	RNM5_CLOAB	5.30E-09
18	comp184083_g1_i1	mediator of RNA polymerase II transcription subunit 14-like	8.00E-14
19	comp172246_g2_i2	Adenoviral fiber	1.30E-08
20	comp187022_g1_i5	ABCA1_MOUSE	3.80E-08
21	comp189939_g1_i4	E3 ubiquitin- ligase Siah1	1.80E-09
22	comp174212_g1_i2	HTSF1_HUMAN	3.00E-44
23	comp184547_g1_i5	TBL1X_MACFA	3.00E-151
24	comp194441_g1_i3	permease	2.20E-135
25	comp172871_g1_i2	hypothetical protein H257_10789	6.20E-19
26	comp174867_g1_i2	thylakoid lumenal 29 kDa , chloroplastic	2.10E-26
27	comp184027_g1_i7	NLRC3_HUMAN	2.70E-54
28	comp186061_g1_i2	cytochrome P450	1.60E-19
29	comp194338_g1_i3	MFS12_MOUSE	8.90E-46
30	comp178298_g1_i1	GELA_DICDI	1.50E-08
31	comp187423_g2_i3	unnamed product	6.20E-17
32	comp182672_g1_i2	CAP2_HUMAN	2.20E-22
33	comp180927_g1_i8	ESIB_ECOL6	2.30E-18
34	comp190852_g3_i2	hypothetical protein AMJ87_05245	3.20E-08
35	comp177197_g1_i4	ELF1_SCHPO	0.00E+00
36	comp172411_g1_i2	PAG15_HUMAN	5.80E-68
37	comp162101_g1_i1	hypothetical protein	2.00E-12
38	comp182881_g1_i2	Probable carboxylesterase	1.40E-56
39	comp180536_g1_i2	vacuolar amino acid transporter 1-like	4.50E-38
40	comp195742_g1_i3	hypothetical protein GUITHDRAFT_89280	6.70E-12
41	comp193043_g1_i3	GPDM_MOUSE	1.60E-140
42	comp189863_g1_i1	COBS_LEPCP	3.20E-21
43	comp183097_g1_i6	IDH_AZOVI	0.00E+00
44	comp185832_g1_i4	Hypothetical protein, putative	4.80E-81
45	comp189074_g4_i3	TTLL4_MOUSE	9.20E-44
46	comp166919_g1_i2	signal recognition particle receptor subunit alpha-like	3.60E-122
47	comp182569_g3_i3	NUCL_MACFA	1.40E-14
48	comp176024_g1_i1	calcium-dependent cytoplasmic cysteine ase, papain	2.10E-98
49	comp182848_g1_i3	SNARE associated Golgi	2.60E-47
50	comp191888_g1_i7	Cation-transporting P-type ATPase	1.30E-08

51	comp192710_g1_i7	sestrin-1 isoform X1	4.50E-53
52	comp187441_g1_i1	dual specificity phosphatase 4	2.40E-09
53	comp191183_g1_i7	HPXO_KLEP7	3.00E-11
54	comp188264_g1_i3	MPCP_YEAST	2.80E-18
55	comp192830_g1_i8	MNN4_YEAST	4.50E-09

Appendix table 13 genes name of area 13

(a) List of TF family (25 transcripts of 16 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp190791_g1_i5	B3	1.00E-07
2	comp193139_g1_i4	MYB_related	1.30E-28
3	comp178914_g1_i1	C3H	9.30E-28
4	comp192674_g1_i3	NAC	4.00E-11
5	comp193842_g1_i4	Nin-like	1.10E-61
6	comp188479_g1_i2	WRKY	5.60E-32
7	comp187030_g2_i2	bHLH	9.80E-19
8	comp192711_g1_i6	BBR-BPC	1.70E-27
9	comp187143_g1_i3	C2H2	1.10E-39
10	comp177544_g1_i5	C3H	2.20E-10
11	comp193320_g1_i5	EIL	2.80E-20
12	comp189778_g1_i1	ERF	2.90E-82
13	comp194763_g1_i2	MYB	7.70E-12
14	comp193452_g1_i12	MYB_related	1.00E-57
15	comp196085_g1_i1	Nin-like	1.50E-46
16	comp187849_g1_i2	VOZ	1.00E-28
17	comp190674_g2_i2	B3	2.50E-41
18	comp181157_g2_i2	FAR1	5.70E-09
19	comp177566_g1_i1	C3H	9.70E-15
20	comp192098_g2_i2	LBD	5.90E-15
21	comp192674_g1_i4	NAC	4.30E-11
22	comp186653_g2_i1	YABBY	4.90E-18
23	comp183451_g2_i1	YABBY	2.70E-31
24	comp187030_g2_i1	bHLH	1.50E-18
25	comp193771_g2_i4	bHLH	2.30E-16

(b) List of other genes name (48 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp192682_g3_i4	glycosyltransferase aer61	5.60E-49
2	comp194861_g2_i1	of CLR family	2.50E-41

3	comp181704_g1_i2	unusual kinase	8.90E-15
4	comp193564_g1_i6	NLRC3_HUMAN	4.80E-41
5	comp179995_g1_i4	Y1181_ARATH	2.10E-95
6	comp164747_g1_i2	hypothetical protein COCSUDRAFT_66025	1.20E-09
7	comp190016_g1_i5	FBX11_RAT	5.10E-46
8	comp195932_g1_i3	membrane-bound adenylyl cyclase	3.10E-107
9	comp194224_g1_i4	beta-1,4-N-acetylglucosaminyltransferase	5.50E-37
10	comp189677_g1_i11	alpha subunit of photoactivated adenylyl cyclase	1.10E-161
11	comp179788_g1_i1	methyltransferase	2.70E-11
12	comp155697_g1_i1	YCX91_PHAAO	1.40E-18
13	comp177997_g2_i4	SDP1L_ARATH	1.30E-62
14	comp184575_g2_i1	membrane-associated protein, putative	3.70E-08
15	comp192874_g2_i6	orofacial cleft 1 candidate gene 1	2.50E-28
16	comp190179_g1_i2	cyclic nucleotide-binding	9.40E-16
17	comp182083_g2_i10	PSR2_YEAST	8.70E-14
18	comp188822_g1_i4	NADB_PSEAE	4.90E-79
19	comp177532_g1_i5	WRP73_HUMAN	5.30E-83
20	comp157569_g1_i1	kelch repeat ,	2.30E-25
21	comp192051_g1_i1	pre translocase subunit	2.30E-23
22	comp170972_g1_i4	CBS domain containing	1.70E-21
23	comp176295_g1_i4	transmembrane protease serine 9	2.20E-22
24	comp188499_g3_i2	regulator of nonsense transcripts UPF2-like	2.10E-64
25	comp178816_g8_i4	GDAP2 homolog isoform X1	7.10E-92
26	comp189991_g3_i1	YCF45_ODOSI	2.90E-31
27	comp193009_g1_i1	branched-chain amino acid aminotransferase	1.80E-129
28	comp193410_g1_i1	MSH1_ARATH	2.40E-85
29	comp170829_g2_i2	permease	1.00E-14
30	comp194897_g1_i5	ANR29_HUMAN	9.10E-08
31	comp172738_g1_i2	NCPP_THEKO	4.30E-16
32	comp171672_g1_i2	RH46_ARATH	1.00E-73
33	comp186019_g1_i3	Cysteine-rich secretory family	6.00E-23
34	comp187294_g3_i2	CFA44_CHLRE	6.10E-46
35	comp174958_g1_i5	cystatin-A	2.40E-19
36	comp177197_g1_i3	elongation factor 3 ABCF transporter family	0.00E+00
37	comp179867_g1_i1	HECD3_HUMAN	7.30E-65

38	comp144519_g1_i1	YCX1_EUGGR	1.50E-169
39	comp169246_g1_i3	predicted protein, partial	4.30E-78
40	comp191053_g2_i6	TRPV6_HUMAN	4.00E-13
41	comp190883_g1_i2	molecular chaperone	4.60E-65
42	comp188842_g2_i2	von Willebrand type A	1.30E-66
43	comp193215_g2_i3	Armadillo betacatenin-like repeat domain containing	6.00E-08
44	comp188085_g1_i5	BYR2_SCHPO	2.00E-40
45	comp184910_g4_i12	predicted protein	6.20E-08
46	comp178417_g1_i5	predicted protein, partial	4.20E-25
47	comp175561_g1_i1	hypothetical protein	1.70E-27
48	comp185922_g4_i2	glutamate carboxypeptidase,	3.30E-53

Appendix table 14 genes name of area 14

(a) List of TF family of part A (61 transcripts of 25 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp192585_g1_i1	GRAS	8.00E-10
2	comp193715_g1_i3	NAC	2.30E-61
3	comp180717_g1_i3	C3H	1.10E-12
4	comp161809_g1_i6	HD-ZIP	8.60E-10
5	comp188326_g1_i1	C2H2	2.70E-29
6	comp188686_g1_i9	YABBY	2.80E-28
7	comp190924_g1_i1	B3	1.00E-35
8	comp185753_g1_i5	E2F_DP	2.10E-07
9	comp195103_g1_i4	G2-like	4.00E-08
10	comp195043_g2_i7	MYB	8.70E-11
11	comp178886_g1_i5	CPP	4.60E-08
12	comp180238_g1_i2	NF-YA	1.30E-39
13	comp188479_g1_i3	WRKY	1.90E-31
14	comp187530_g1_i6	C2H2	2.80E-29
15	comp179606_g1_i3	C3H	9.90E-32
16	comp192674_g1_i5	NAC	4.10E-11
17	comp194296_g1_i2	FAR1	1.40E-18
18	comp188523_g1_i4	MYB	2.60E-12
19	comp188009_g1_i4	YABBY	3.10E-32
20	comp186117_g1_i3	bHLH	1.70E-08
21	comp192714_g1_i7	bZIP	1.80E-13
22	comp191702_g3_i3	HB-other	5.30E-10
23	comp191995_g1_i1	NAC	8.90E-11
24	comp187530_g1_i8	C2H2	2.60E-29
25	comp185968_g1_i1	bHLH	3.00E-08
26	comp192674_g1_i6	NAC	4.10E-11
27	comp186653_g3_i5	YABBY	3.70E-18
28	comp179692_g1_i1	C3H	3.70E-47
29	comp185793_g1_i1	E2F_DP	9.80E-17
30	comp180238_g1_i3	NF-YA	1.40E-39
31	comp187198_g1_i2	bHLH	8.60E-28
32	comp188086_g1_i1	C2H2	1.70E-14

33	comp178578_g5_i2	MYB_related	2.30E-12
34	comp195160_g2_i5	MYB	9.40E-12
35	comp188710_g2_i1	WRKY	1.10E-37
36	comp194428_g2_i1	bZIP	3.20E-18
37	comp188131_g1_i1	C2H2	1.20E-38
38	comp194084_g1_i2	bHLH	3.90E-08
39	comp172263_g1_i1	GeBP	3.10E-07
40	comp179351_g2_i2	MYB_related	3.50E-24
41	comp187198_g1_i4	bHLH	7.10E-28
42	comp195219_g2_i7	LBD	3.60E-22
43	comp185204_g1_i2	bHLH	3.50E-12
44	comp188093_g1_i2	C3H	2.00E-07
45	comp184177_g1_i3	ERF	8.10E-78
46	comp192838_g2_i2	Nin-like	1.30E-46
47	comp182302_g2_i2	YABBY	4.60E-15
48	comp186940_g1_i2	B3	3.70E-26
49	comp170769_g1_i16	GATA	4.20E-07
50	comp180402_g2_i2	FAR1	3.60E-16
51	comp190761_g1_i5	NAC	6.90E-22
52	comp192380_g2_i2	EIL	1.50E-28
53	comp187476_g2_i1	MYB_related	9.00E-42
54	comp186008_g2_i8	WRKY	1.40E-73
55	comp187293_g1_i4	G2-like	3.70E-28
56	comp191841_g1_i3	AP2	1.40E-46
57	comp186095_g1_i4	MYB_related	4.60E-14
58	comp170769_g1_i15	GATA	7.50E-07
59	comp190761_g1_i2	NAC	7.10E-22
60	comp186008_g2_i7	WRKY	1.50E-73
61	comp187665_g1_i1	C3H	5.40E-12

(b) List of other genes name of part A (221 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp173360_g1_i2	lysine decarboxylase domain-containing	2.00E-100
2	comp182973_g1_i1	predicted protein, partial	8.70E-43

3	comp176712_g1_i3	SPNS1_ARATH	5.20E-17
4	comp189908_g1_i4	UDP-glucuronic acid UDP-N-acetylgalactosamine transporter	1.00E-67
5	comp182711_g1_i2	Trk1 potassium transporter	2.70E-42
6	comp183059_g1_i2	ankyrin repeat domain-containing 27-like	5.40E-11
7	comp175868_g1_i3	TBC1 domain family member 20	1.20E-41
8	comp178860_g1_i1	poly(ADP-ribose) glycohydrolase-like	2.10E-50
9	comp178026_g3_i1	ADP-ribosyltransferase 1 precursor	2.20E-11
10	comp178799_g1_i7	NADB_PSEAE	4.80E-75
11	comp195507_g1_i3	MELK_XENLA	9.80E-11
12	comp185788_g1_i4	GDSL-like Lipase Acylhydrolase	1.40E-22
13	comp161573_g1_i1	HIR3_ARATH	5.80E-76
14	comp178812_g1_i2	aaap amino acid permease	9.80E-50
15	comp169145_g1_i1	GLD2_DANRE	1.00E-13
16	comp161708_g1_i2	branch point binding	5.20E-30
17	comp192355_g1_i1	WECH_DROME	1.70E-07
18	comp190712_g1_i9	hypothetical protein COCSUDRAFT_31865	1.40E-19
19	comp187961_g3_i5	sex-lethal homolog isoform X1	2.40E-23
20	comp193684_g1_i1	GEFF_DICDI	7.40E-15
21	comp155211_g1_i2	KLHL4_HUMAN	1.30E-27
22	comp182185_g5_i5	conserved unknown protein	2.30E-24
23	comp187330_g2_i2	hypothetical protein PBCVCvsA1_025R	4.40E-19
24	comp152118_g1_i3	YAS2_SCHPO	7.10E-15
25	comp185144_g1_i3	PKWA_THECU	1.70E-28
26	comp175038_g1_i12	Tetratricopeptide-like helical	3.10E-46
27	comp190162_g1_i3	YOUD_CAEEL	7.70E-08
28	comp182956_g1_i1	METL4_MOUSE	3.30E-20
29	comp192377_g1_i2	USP_ORYSI	8.80E-125
30	comp178361_g1_i2	hypothetical protein G7K_4091-t1	2.50E-08
31	comp186111_g2_i10	STPG1_DANRE	8.00E-10
32	comp189953_g2_i4	membrane	3.50E-11
33	comp155310_g1_i1	CBS domain containing	1.30E-18
34	comp191711_g2_i1	flavincontaining amine	1.40E-162
35	comp176301_g1_i2	VMAT1_RAT	2.90E-09
36	comp188033_g1_i8	SAMC2_ARATH	3.30E-22
37	comp166867_g1_i3	ATP-dependent RNA helicase DBP5	2.00E-48

38	comp172681_g1_i1	pyruvate dehydrogenase E1 component alpha subunit	0.00E+00
39	comp163319_g1_i2	psbP domain-containing 6, chloroplastic	2.70E-61
40	comp176106_g2_i2	hypothetical protein DFA_11718	1.40E-39
41	comp170841_g1_i3	TRAM LAG1 CLN8 homology domain	1.40E-17
42	comp180530_g1_i1	SPOT_ECO57	2.80E-76
43	comp178763_g2_i5	TTD14_DROME	2.40E-34
44	comp156508_g1_i3	YCX8_EUGGR	1.10E-53
45	comp163062_g1_i1	2ABD_CHICK	8.60E-116
46	comp184727_g1_i2	GDIR_ARATH	5.40E-08
47	comp183424_g1_i2	hypothetical protein	2.70E-69
48	comp186812_g1_i3	Hypothetical protein, putative	1.90E-12
49	comp184195_g1_i1	SPHK1_ARATH	2.90E-32
50	comp189758_g2_i6	AARA_DICDI	5.80E-11
51	comp187179_g2_i4	CNOT4_MOUSE	1.70E-22
52	comp184314_g1_i2	IQ calmodulin-binding motif family	2.20E-133
53	comp194898_g1_i5	hypothetical protein ACA1_171730	9.40E-11
54	comp176130_g1_i3	RH30_ARATH	1.10E-30
55	comp185971_g1_i3	C-myc promoter-binding -like	9.60E-41
56	comp174011_g2_i2	hydroxyacylglutathione hydrolase	6.60E-112
57	comp192508_g2_i1	ANK3_MOUSE	7.00E-08
58	comp184610_g1_i2	WD-40 repeat	3.30E-29
59	comp191059_g1_i1	TRPV6_HUMAN	2.30E-25
60	comp179865_g1_i1	KAPR_APLCA	1.00E-22
61	comp162559_g1_i3	PR2E2_ORYSJ	9.90E-33
62	comp167300_g1_i4	KAPR_DICDI	1.40E-74
63	comp169420_g1_i2	YMO3_PANSE	2.90E-09
64	comp180755_g1_i2	secreted partial	4.80E-12
65	comp184895_g2_i3	COBT_PSEF5	2.80E-30
66	comp171071_g1_i1	DJ1B_ARATH	7.20E-30
67	comp162429_g1_i2	NIMA-related kinase	6.60E-10
68	comp194459_g1_i4	CAN7_HUMAN	1.40E-92
69	comp179111_g1_i1	C-lectin, putative	7.90E-08
70	comp173974_g2_i2	enoyl- hydratase	1.40E-60
71	comp171352_g1_i1	misato-like isoform X1	3.60E-12
72	comp175716_g1_i1	Tetratricopeptide-like helical	1.40E-46

73	comp177079_g1_i2	exostosin family	4.30E-33
74	comp182908_g1_i3	unnamed product	4.20E-45
75	comp172899_g2_i10	FBT2_ARATH	7.30E-08
76	comp195867_g2_i31	poly(ADP-ribose) polymerase	3.20E-10
77	comp195480_g1_i5	parp domain-containing	3.60E-11
78	comp194688_g1_i9	ATP-binding cassette transporter, subfamily G, member 15, group PDR 15	2.50E-147
79	comp194186_g1_i7	vitamin K-dependent gamma-carboxylase	1.10E-16
80	comp182857_g1_i3	Sphingosine kinase 2 (Fragment)	2.30E-38
81	comp185440_g2_i4	VTI11_ARATH	2.20E-17
82	comp189611_g2_i1	udp-n-acetylglucosamine--peptide n-acetylglucosaminyltransferase	2.10E-17
83	comp189120_g1_i1	DYH10_HUMAN	8.70E-08
84	comp190418_g2_i4	CYA1_RHIME	1.80E-18
85	comp176177_g1_i3	zinc finger FYVE domain-containing 16	8.90E-10
86	comp179162_g1_i9	TSR3_YEAST	1.00E-22
87	comp180339_g1_i1	tRNA-specific adenosine deaminase 1 isoform X1	7.70E-46
88	comp184685_g1_i6	YIS7_YEAST	8.70E-78
89	comp194710_g1_i9	MTFP1_CAEEL	7.90E-11
90	comp188116_g1_i1	PREDICTED: uncharacterized protein LOC756716	6.90E-28
91	comp175309_g1_i2	Neurotrypsin	3.20E-07
92	comp192939_g1_i2	PCYAA_EUGLO	9.90E-139
93	comp179586_g1_i1	TSGAD_HALVD	3.50E-09
94	comp177981_g1_i5	5 -3 exoribonuclease 3-like	3.70E-83
95	comp194004_g1_i1	ADCYA_HUMAN	3.80E-19
96	comp192683_g1_i3	DNJA6_ARATH	7.30E-101
97	comp189293_g1_i3	Poly(ADP-ribose) polymerase catalytic domain containing	6.70E-89
98	comp193535_g1_i2	ADCYA_RAT	2.80E-37
99	comp195700_g2_i1	DDX19_DICDI	7.40E-29
100	comp193939_g1_i4	serine threonine kinase	6.40E-97
101	comp181796_g1_i1	RSH1C_ARATH	4.40E-57
102	comp180823_g4_i9	A13CB_XENLA	1.50E-10
103	comp192723_g1_i3	KAD2_BOVIN	5.00E-41
104	comp194897_g1_i8	ANR29_HUMAN	8.50E-08
105	comp190305_g1_i4	hypothetical protein ZOSMA_452G00060	7.10E-10
106	comp170807_g2_i1	secreted partial	5.00E-12
107	comp186063_g1_i1	predicted protein	1.00E-12

108	comp191932_g1_i9	F213A_XENTR	1.30E-23
109	comp172899_g2_i14	folate pteridine transporter	1.10E-15
110	comp190621_g4_i1	kinase domain partial	3.80E-08
111	comp188065_g1_i1	SLBP2_XENLA	1.10E-08
112	comp182682_g4_i1	ADCA_DICDI	1.00E-09
113	comp161603_g1_i1	RHBGB_XENLA	8.80E-79
114	comp186533_g1_i10	membrane-bound adenylyl cyclase	2.00E-31
115	comp190949_g1_i6	E3 ubiquitin- ligase RLIM-like isoform X1	1.30E-14
116	comp185744_g2_i3	IP6K2_MOUSE	2.20E-19
117	comp193185_g1_i1	ef hand family	4.30E-32
118	comp183895_g1_i4	rieske [2Fe-2S]	1.60E-48
119	comp181188_g1_i4	TIL_ARATH	3.60E-10
120	comp190284_g1_i4	predicted protein	1.50E-24
121	comp181811_g1_i4	SEC14_SCHPO	2.30E-29
122	comp184205_g1_i1	AK1_DICDI	2.20E-20
123	comp195932_g1_i1	membrane-bound adenylyl cyclase	3.10E-107
124	comp191912_g1_i12	AARA_DICDI	1.20E-10
125	comp189916_g1_i3	fructose-1,6-bisphosphatase, chloroplastic-like	1.30E-88
126	comp176465_g1_i1	ribose-phosphate pyrophosphokinase	6.80E-113
127	comp193086_g1_i3	EF-Hand calcium-binding site	0.00E+00
128	comp170391_g1_i2	MSRA_RHOCS	2.00E-83
129	comp194647_g1_i1	ADCY4_MOUSE	5.30E-14
130	comp188259_g2_i1	mitochondrial transcription termination factor family	5.00E-10
131	comp178213_g1_i4	GALM_ACICA	7.50E-102
132	comp180895_g1_i2	WD repeat-containing WRAP73	6.80E-85
133	comp190392_g1_i2	F26L_CHICK	3.70E-89
134	comp172640_g1_i2	threonine dehydratase	2.00E-158
135	comp179619_g1_i3	NPC2_YARLI	8.00E-13
136	comp170754_g2_i5	SMYD3_MOUSE	2.30E-07
137	comp56453_g1_i1	RLA1_CHLRE	1.60E-17
138	comp186752_g1_i4	nicotinate-nucleotide adenylyltransferase	1.10E-52
139	comp185023_g6_i4	transporter ,	2.80E-80
140	comp176974_g1_i6	VDE_LACSA	5.40E-52
141	comp171652_g1_i3	metal tolerance C4-like	2.30E-42
142	comp190947_g2_i3	S12A4_HUMAN	9.40E-120

143	comp189760_g1_i2	DYH1_HUMAN	2.40E-25
144	comp182004_g1_i2	GLOD4_MOUSE	8.60E-63
145	comp191473_g1_i6	WDR90_HUMAN	3.80E-40
146	comp190839_g1_i2	hypothetical protein	2.80E-27
147	comp181907_g1_i1	CDC73_ARATH	1.70E-24
148	comp184170_g1_i1	MA1A1_MOUSE	7.40E-50
149	comp185498_g1_i2	EF2K_RAT	1.30E-20
150	comp178945_g1_i2	Surfeit locus 1 Shy1	8.50E-29
151	comp178819_g1_i2	membrane-associated protein, putative	1.40E-86
152	comp192346_g2_i1	cyclic nucleotide-binding	1.40E-24
153	comp179026_g1_i3	YOS3_SCHPO	4.30E-49
154	comp195372_g2_i5	calpain-7 isoform X2	8.40E-103
155	comp191136_g1_i7	hypothetical protein ACA1_171730	2.60E-23
156	comp160179_g1_i2	procollagen galactosyltransferase 1	4.30E-15
157	comp185795_g1_i3	expressed unknown protein	1.90E-48
158	comp183078_g1_i2	hypothetical protein H310_03776	1.80E-15
159	comp191840_g1_i9	predicted protein, partial	2.70E-120
160	comp193901_g2_i9	Carbohydrate-binding and sugar hydrolysis	8.00E-09
161	comp181865_g1_i2	hypothetical protein	6.10E-10
162	comp183147_g1_i3	cation:proton antiporter	2.60E-50
163	comp190407_g2_i1	mannosylphosphorylation MNN4	2.10E-11
164	comp189074_g2_i3	TTLL4_MOUSE	3.30E-43
165	comp182313_g2_i4	glycosyltransferase aer61	1.00E-23
166	comp178805_g3_i2	Fe-S oxidoreductase	5.50E-168
167	comp185440_g1_i6	VTI11_ARATH	5.90E-17
168	comp181717_g1_i5	glycosyl transferase	1.10E-40
169	comp186814_g1_i4	ABC transporter ATP-binding	9.80E-71
170	comp186236_g1_i4	POLH_ARATH	2.80E-101
171	comp184738_g2_i3	TKL kinase	2.60E-20
172	comp166655_g1_i3	TT39C_MOUSE	9.00E-42
173	comp186806_g1_i3	NIA_VOLCA	4.70E-40
174	comp166745_g2_i2	WDR53_DICDI	2.50E-18
175	comp195932_g1_i5	membrane-bound adenylyl cyclase	4.10E-103
176	comp194293_g1_i9	IP6K3_MOUSE	3.10E-30
177	comp181434_g1_i3	tyrosine biosynthesis bifunctional enzyme	6.00E-155

178	comp183103_g1_i3	Plastid lipid-associated fibrillin conserved domain	1.80E-32
179	comp188676_g1_i4	SPAG1_DANRE	9.00E-19
180	comp194736_g1_i2	Holliday junction resolvase	6.90E-23
181	comp191943_g4_i5	spermidine synthase	2.80E-16
182	comp182569_g3_i7	NUCL_HUMAN	1.50E-12
183	comp182446_g1_i11	MPPD1_HUMAN	3.90E-55
184	comp178741_g1_i3	PA1B3_BOVIN	1.90E-21
185	comp189226_g1_i4	CYAA_TRYEQ	2.90E-53
186	comp182369_g4_i3	feline leukemia virus subgroup C receptor-related 2	4.30E-29
187	comp195028_g1_i2	TOR_DICDI	0.00E+00
188	comp145907_g1_i1	ILV5_SCHPO	7.90E-100
189	comp164768_g1_i1	NUCL_XENLA	2.30E-11
190	comp164157_g1_i6	SGT1 homolog A	4.20E-42
191	comp138431_g1_i3	PSBB_EUGGR	9.30E-135
192	comp176007_g1_i2	predicted protein	2.60E-08
193	comp172013_g1_i2	nucleotidyltransferase DNA polymerase	4.50E-22
194	comp187325_g1_i12	glycosyl transferase	1.30E-17
195	comp175153_g1_i4	SPNS3_ARATH	2.90E-10
196	comp189244_g1_i5	hypothetical protein ACA1_171730	3.50E-23
197	comp188423_g1_i7	-glutamine gamma-glutamyltransferase 4	7.70E-73
198	comp194100_g1_i4	HPT1_ARATH	9.10E-38
199	comp180196_g2_i3	CE164_MOUSE	1.50E-25
200	comp175517_g2_i2	DPH2_NEMVE	4.50E-37
201	comp184180_g3_i2	C-type lectin domain family 2 member D-like	4.20E-09
202	comp185272_g1_i2	Y7588_DICDI	2.10E-53
203	comp184818_g1_i4	M2K2_ARATH	3.50E-41
204	comp169503_g1_i4	hypothetical protein	1.40E-13
205	comp185639_g1_i3	endonuclease exonuclease phosphatase family	1.40E-11
206	comp189378_g1_i1	ZUPT_CHLP8	1.10E-37
207	comp192939_g1_i1	alpha subunit of photoactivated adenylyl cyclase	5.50E-82
208	comp186675_g1_i1	transmembrane protein, putative	5.60E-08
209	comp154886_g1_i5	#NAME?	5.00E-125
210	comp187309_g1_i2	RNA ligase	3.20E-76
211	comp194700_g2_i1	IDE_HUMAN	1.80E-85
212	comp192723_g1_i4	KAD2_BOVIN	5.70E-41

213	comp193750_g2_i2	Ribonuclease H1	5.90E-08
214	comp164481_g2_i2	Appr-1-p processing	8.30E-21
215	comp194250_g1_i6	PFLB_STAEQ	0.00E+00
216	comp172151_g1_i2	ANKK1_HUMAN	2.30E-10
217	comp176173_g2_i1	Y1586_HAEINRecName: Full=Uncharacterized protein HI_1586	3.00E-31
218	comp189030_g1_i4	EGY1_ARATH	1.90E-26
219	comp186164_g4_i1	Armadillo betacatenin-like repeat domain containing	1.40E-24
220	comp185769_g1_i3	SLBP1_XENLA	2.10E-13
221	comp178414_g2_i5	MAP kinase phosphatase 6	5.50E-29

(c) List of TF family of part B (3 transcripts of 3 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp143386_g1_i1	NAC	9.70E-22
2	comp104672_g1_i1	C3H	1.30E-17
3	comp144919_g1_i1	FAR1	1.80E-31

(d) List of other genes name of part B (34 transcripts)

No.	Seqname of Euglena	Annotation	Evalue
1	comp170994_g1_i4	hybrid sensor histidine kinase response regulator	8.10E-52
2	comp163883_g1_i4	thioredoxin domain-containing 9-like	1.90E-45
3	comp143235_g1_i1	TRAF6_HUMAN	2.60E-15
4	comp192841_g2_i2	hypothetical protein Ctob_010698	8.70E-08
5	comp182251_g2_i3	BACH_MOUSE	3.30E-22
6	comp187138_g1_i6	UBA1Y_MOUSE	8.50E-175
7	comp190120_g1_i1	rab9 effector with kelch motifs-like	2.20E-30
8	comp192812_g2_i3	Y1733_ARATH	1.20E-61
9	comp191567_g2_i1	TBC31_HUMAN	5.00E-28
10	comp182366_g1_i5	bile Acid:Na+ symporter family	1.90E-39
11	comp192589_g1_i6	phosphatidylinositol-4-phosphate 5-kinase	3.20E-19
12	comp183339_g1_i1	EXOS9_RAT	3.90E-61
13	comp172494_g2_i1	IDLC_STRPU	1.10E-62
14	comp185483_g1_i1	CRNS1_CHICK	2.60E-18
15	comp195154_g2_i4	ENDOV_MOUSE	1.60E-25

16	comp171107_g1_i1	ATP-binding Cassette (ABC) superfamily	7.80E-14
17	comp189106_g1_i6	hypothetical protein H310_00261	5.40E-14
18	comp192554_g1_i5	EGY1_ARATH	7.50E-25
19	comp183739_g4_i2	TSGAD_HALVD	2.50E-28
20	comp188844_g2_i5	centrosomal of 104 kDapredicted protein, partial	7.20E-14
21	comp184084_g3_i4	kinase domain	1.40E-12
22	comp180579_g1_i12	D-alanine--D-alanine ligase	1.10E-132
23	comp184393_g3_i2	D Chain D	3.40E-09
24	comp179397_g1_i3	hypothetical protein UT61_C0050G0009	2.50E-08
25	comp175031_g1_i2	NSR1_YEAST	1.90E-11
26	comp184495_g1_i1	hypothetical protein COCSUDRAFT_43152	1.40E-11
27	comp187209_g2_i3	predicted protein	1.80E-34
28	comp176652_g3_i3	PSTS_RHILO	4.70E-16
29	comp194630_g1_i1	TCB2_CAEBR	1.20E-07
30	comp162236_g1_i2	hypothetical protein GUITHDRAFT_149743	3.50E-20
31	comp188621_g1_i14	copper transporter family	2.40E-13
32	comp153383_g1_i1	rubrerythrin	5.20E-27
33	comp189490_g1_i22	SET domain-containing	2.60E-29
34	comp156488_g1_i3	YAS2_SCHPO	7.30E-15

(e) List of other genes name of part C (18 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp188033_g1_i7	SAMC2_ARATH	3.10E-22
2	comp188892_g1_i7	c transferase	1.30E-17
3	comp178364_g2_i7	probable E3 ubiquitin- ligase LOG2	4.60E-10
4	comp195101_g1_i4	family 3 adenylate cyclase	8.70E-46
5	comp192143_g1_i5	DDX3_DROME	1.10E-15
6	comp193633_g1_i3	hypothetical protein, variant	9.20E-47
7	comp189197_g3_i5	NLRC3_HUMAN	1.40E-07
8	comp154015_g1_i1	kinase domain	6.50E-25
9	comp173571_g1_i4	hypothetical protein SARC_00571	8.60E-17
10	comp181894_g1_i2	UBE4B_MOUSE	1.50E-104
11	comp196159_g1_i3	ABC transporter substrate-binding	9.70E-21
12	comp194934_g1_i12	PABP_YEAST	4.40E-13

13	comp183342_g1_i2	SR45A_ARATH	4.00E-10
14	comp184347_g1_i4	SEPR_THESR	1.00E-31
15	comp192704_g3_i14	adenylate guanylate cyclase catalytic domain	3.40E-12
16	comp187250_g2_i4	serine threonine kinase	3.30E-40
17	comp193767_g1_i1	CY42_TRYBB	1.40E-27
18	comp193811_g1_i4	NLRC3_MOUSE	1.40E-17

(f) List of TF family of part D (2 transcripts of 2 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp195859_g1_i1	bHLH	7.00E-54
2	comp181866_g2_i1	C3H	2.50E-34

(g) List of other genes name of part D (7 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp173077_g1_i1	C2D61_ARATH	2.50E-07
2	comp184429_g1_i2	NLRC3_HUMAN	9.60E-14
3	comp195156_g1_i1	calpain-like cysteine peptidase	3.30E-14
4	comp189518_g2_i5	IQ calmodulin-binding	2.20E-13
5	comp159676_g1_i4	ubiquitin ligase	4.90E-16
6	comp181836_g2_i3	PDE1A_MOUSE	3.50E-60
7	comp192378_g1_i7	CYAB_STIAU	3.30E-12

Appendix table 15 genes name of area 15

(a) List of TF family (63 transcripts of 19 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp180043_g1_i2	C2H2	9.90E-07
2	comp175137_g1_i7	GATA	6.90E-07
3	comp196332_g1_i2	GATA	1.10E-16
4	comp193548_g1_i3	HB-other	1.30E-08
5	comp188838_g2_i2	MYB	3.70E-21
6	comp193900_g2_i3	MYB	2.70E-35
7	comp175557_g2_i2	MYB_related	5.20E-10
8	comp191995_g2_i3	NAC	1.00E-10
9	comp191995_g2_i4	NAC	1.00E-10
10	comp192373_g1_i4	NAC	6.10E-11
11	comp192023_g2_i3	CPP	2.20E-15
12	comp186212_g2_i1	C2H2	3.30E-07
13	comp186212_g2_i3	C2H2	3.50E-07
14	comp188286_g1_i2	YABBY	3.30E-32
15	comp177345_g2_i1	C3H	3.00E-13
16	comp190210_g1_i4	B3	8.10E-43
17	comp187654_g1_i3	WRKY	1.40E-19
18	comp187257_g1_i2	WRKY	2.30E-28
19	comp188212_g2_i4	WRKY	1.40E-36
20	comp186996_g1_i2	bHLH	7.70E-09
21	comp193322_g1_i2	bHLH	4.80E-08
22	comp189388_g1_i5	B3	1.20E-26
23	comp183118_g2_i6	B3	2.80E-21
24	comp186591_g1_i1	C2H2	2.40E-34
25	comp175411_g1_i2	C3H	3.10E-07
26	comp175305_g1_i2	C3H	5.10E-11
27	comp175529_g1_i1	C3H	1.60E-12
28	comp185753_g1_i1	E2F_DP	2.10E-07
29	comp176305_g1_i2	FAR1	3.30E-08
30	comp193281_g1_i9	G2-like	1.20E-20
31	comp186554_g1_i2	GATA	9.60E-12
32	comp193899_g3_i8	GATA	9.60E-12

33	comp167835_g1_i2	HB-other	9.90E-23
34	comp161240_g1_i3	HD-ZIP	9.80E-10
35	comp184981_g1_i6	MYB	2.80E-21
36	comp192683_g1_i10	MYB	2.30E-34
37	comp175557_g4_i2	MYB_related	8.30E-10
38	comp191995_g1_i7	NAC	1.30E-10
39	comp164604_g1_i1	Trihelix	2.10E-15
40	comp187654_g2_i1	WRKY	1.40E-19
41	comp187970_g2_i6	WRKY	1.50E-26
42	comp188085_g1_i1	WRKY	1.00E-30
43	comp181025_g1_i3	WRKY	2.60E-93
44	comp194098_g2_i3	YABBY	6.10E-25
45	comp182044_g1_i6	bHLH	1.10E-12
46	comp193204_g1_i1	bHLH	1.80E-47
47	comp193135_g1_i4	bZIP	2.90E-07
48	comp189388_g1_i4	B3	1.30E-26
49	comp185753_g1_i3	E2F_DP	1.90E-07
50	comp175137_g1_i2	GATA	6.00E-07
51	comp193991_g1_i1	HB-other	1.30E-08
52	comp190353_g1_i1	MYB	1.80E-15
53	comp177478_g1_i3	C3H	9.70E-30
54	comp191995_g1_i4	NAC	8.60E-11
55	comp192674_g1_i1	NAC	4.20E-11
56	comp192373_g1_i6	NAC	3.80E-10
57	comp175390_g1_i6	NF-YA	2.90E-36
58	comp192055_g1_i1	NF-YA	1.40E-08
59	comp185491_g1_i9	NAC	1.00E-22
60	comp186502_g1_i3	bHLH	3.60E-07
61	comp186677_g1_i3	bHLH	1.90E-10
62	comp186392_g1_i3	bHLH	5.80E-18
63	comp190925_g1_i2	bZIP	5.10E-13

(b) List of other genes name (243 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue

1	comp176727_g1_i3	IM30_ORYSJ	1.40E-53
2	comp190563_g1_i5	AK1_DICDI	2.50E-44
3	comp193576_g2_i5	hypothetical protein SPRG_18819	1.00E-18
4	comp181004_g1_i1	NLRC3_HUMAN	5.20E-31
5	comp192124_g1_i2	PLDA1_RICCO	7.10E-148
6	comp194932_g2_i1	tetrapyrrole uroporphyrinogen III synthase	2.20E-10
7	comp184448_g2_i6	NLRC3_HUMAN	4.20E-22
8	comp164336_g1_i2	RING finger 10	3.40E-29
9	comp188026_g1_i7	MTERF isoform 2	6.00E-15
10	comp152886_g1_i1	mitochondrial tryptophan rich sensory 1	8.30E-09
11	comp184316_g1_i6	COPDA_ARATH	0.00E+00
12	comp180480_g3_i3	caffeoylshikimate esterase-like	4.30E-41
13	comp167885_g1_i2	stress regulated isoform 3	2.50E-89
14	comp178026_g3_i4	ADP-ribosyltransferase 1 precursor	8.40E-12
15	comp170415_g1_i2	LIPV_MYCTU	2.70E-07
16	comp192577_g2_i7	probable palmitoyltransferase ZDHHC1 isoform X1	2.10E-39
17	comp191847_g1_i5	CYAA_TRYEQ	2.70E-26
18	comp195238_g1_i4	OHK3_ORYSI	1.30E-28
19	comp179179_g2_i1	RIB2_YEAST	1.80E-16
20	comp173909_g2_i5	UN104_CAEEL	1.30E-10
21	comp179316_g1_i3	TTL_BOVIN	1.50E-45
22	comp183521_g1_i5	MNTH_RHILO	8.30E-130
23	comp175579_g1_i2	Bcy1p	2.20E-14
24	comp187600_g1_i2	#NAME?	3.10E-65
25	comp173811_g1_i1	dual specificity phosphatase DSP8	1.70E-13
26	comp163805_g1_i1	ENPP1_MOUSE	3.20E-09
27	comp188490_g1_i5	PABP_COCIM	5.20E-20
28	comp184803_g1_i2	alcohol acetyltransferase	2.20E-32
29	comp176030_g2_i2	BACH_RAT	6.10E-43
30	comp191621_g2_i10	hypothetical protein CHLNCDRAFT_138448	3.90E-08
31	comp150070_g1_i2	KAP0_HUMAN	2.10E-09
32	comp195557_g1_i5	hypothetical protein H257_09746	3.80E-08
33	comp189055_g1_i3	DPOL_DHBV1	3.90E-10
34	comp195443_g3_i3	hypothetical protein Ctob_006104	3.70E-08
35	comp191242_g1_i16	hypothetical protein AURANDRAFT_70781	4.40E-34

36	comp264207_g1_i1	PEB3_DROME	1.10E-16
37	comp191249_g3_i4	IF4G2_ORYSJ	6.70E-12
38	comp177847_g1_i1	GT16_ARATH	3.10E-08
39	comp146874_g1_i2	CCL11_ORYSJ	8.00E-32
40	comp109800_g1_i1	RS24_DICDI	6.30E-25
41	comp178888_g1_i1	IF4E_RAT	2.10E-45
42	comp167840_g1_i1	sphingomyelin synthase-related 1	2.80E-18
43	comp194738_g2_i9	4-demethylwyosine synthase-like	2.10E-108
44	comp191964_g2_i6	predicted protein	1.40E-09
45	comp176906_g1_i2	lycopene beta cyclase	2.90E-81
46	comp162631_g1_i1	alpha-(1,6)-fucosyltransferase	2.20E-18
47	comp190716_g1_i10	AK1_DICDI	2.60E-40
48	comp189102_g1_i2	conserved unknown protein	1.00E-06
49	comp127935_g1_i3	PSBA_EUGGR	1.60E-31
50	comp184205_g1_i2	transient receptor potential cation channel subfamily M member 6	2.30E-21
51	comp192841_g2_i8	hypothetical protein Ctob_010698	8.30E-08
52	comp163784_g1_i3	protein deltex-2	7.90E-94
53	comp165040_g1_i3	cytochrome b5	1.90E-17
54	comp170245_g1_i1	MYCF_MICGR	8.50E-19
55	comp185637_g1_i2	SAM-dependent methyltransferase	2.50E-09
56	comp165297_g2_i1	DDX4_BOVIN	1.50E-69
57	comp172763_g1_i2	Hypothetical protein, putative	4.00E-23
58	comp158577_g1_i5	zinc finger with UFM1-specific peptidase domain	7.40E-53
59	comp170807_g2_i7	secreted partial	4.50E-12
60	comp149582_g1_i2	RIBX_MIMIV	1.60E-47
61	comp192979_g1_i7	TRAP1_MOUSE	0.00E+00
62	comp188625_g1_i6	predicted protein, partial	1.60E-54
63	comp186457_g2_i4	MBTP1_HUMAN	2.50E-133
64	comp178763_g2_i8	TTD14_DROME	2.20E-34
65	comp184170_g1_i4	MA1A1_RABIT	6.20E-112
66	comp192919_g2_i3	karyopherin beta	8.20E-104
67	comp184116_g1_i2	NLRC3_HUMAN	1.10E-27
68	comp161920_g1_i2	glycosyl family 43	2.40E-50
69	comp185679_g1_i3	nicotinate-nucleotide adenylyltransferase	5.80E-52

70	comp191850_g1_i1	MENA_SYNY3	5.40E-49
71	comp159493_g1_i5	COX1_CHOFU	6.90E-126
72	comp172016_g4_i2	DPOE2_CHICK	6.00E-81
73	comp188707_g1_i2	unnamed protein product	5.90E-28
74	comp193521_g1_i7	Cna B domain	1.30E-11
75	comp189912_g2_i1	Dihydroorotate dehydrogenase (quinone), mitochondrial	0.00E+00
76	comp184138_g1_i1	progesterone-induced-blocking factor 1	7.30E-21
77	comp174840_g1_i4	RING finger	8.20E-38
78	comp190284_g1_i3	predicted protein	1.00E-22
79	comp194573_g7_i4	CY42_TRYBB	1.40E-31
80	comp192586_g1_i2	CRLA_DICDI	1.50E-20
81	comp185125_g1_i7	PAS domain-containing sensor histidine kinase	2.90E-14
82	comp193413_g1_i1	YJJW_ECOLI	9.60E-132
83	comp177683_g1_i1	NLRC3_HUMAN	6.80E-10
84	comp188045_g2_i3	predicted protein	7.30E-19
85	comp192852_g1_i1	RBBP6_HUMAN	4.90E-22
86	comp188869_g1_i4	probable palmitoyltransferase ZDHHC1 isoform X1	1.90E-39
87	comp174015_g1_i3	rhodanese	4.20E-15
88	comp195154_g2_i7	endonuclease V isoform X1	2.70E-25
89	comp190928_g3_i4	phosphohydrolase	3.60E-48
90	comp192353_g1_i4	MCES_SCHPO	1.90E-09
91	comp189302_g1_i2	YGRG_SCHPO	3.60E-33
92	comp162703_g2_i1	AP4S_ARATH	2.00E-59
93	comp166774_g1_i3	PTBP3_ARATH	7.20E-08
94	comp157182_g1_i1	hypothetical protein PBRA_008009	6.40E-18
95	comp186460_g1_i2	PTBP2_HUMAN	7.30E-68
96	comp185224_g1_i1	UBA1_HUMAN	0.00E+00
97	comp195191_g1_i5	magnesium chelatase	0.00E+00
98	comp196286_g3_i5	TCPR1_CHICK	5.40E-07
99	comp175031_g1_i3	NSR1_YEAST	1.90E-11
100	comp174836_g1_i5	Hypothetical protein, putative	1.10E-11
101	comp188940_g1_i4	COPDA_ARATH	3.20E-61
102	comp189471_g1_i3	KLHDB_DROSI	1.50E-18
103	comp186653_g3_i14	KPK2_ARATH	2.80E-34
104	comp194535_g1_i7	chromosome partitioning	2.20E-09

105	comp175936_g1_i1	IF4E_CANGA	1.00E-14
106	comp173788_g1_i2	hypothetical protein GUITHDRAFT_100284	1.40E-10
107	comp183704_g1_i3	unnamed product	1.40E-23
108	comp190100_g1_i4	membrane-associated protein, putative	1.30E-08
109	comp183544_g1_i2	AT3g61870 F21F14_40	1.10E-54
110	comp177998_g1_i3	hypothetical protein	3.10E-08
111	comp194215_g1_i5	hypothetical protein AURANDRAFT_67549	4.10E-13
112	comp176177_g1_i5	zinc finger FYVE domain-containing 16	9.10E-10
113	comp159709_g1_i1	exostosin	1.30E-23
114	comp195010_g2_i2	RWD2B_MOUSE	8.20E-08
115	comp183852_g1_i6	RBN_CROS8	7.20E-55
116	comp188956_g1_i5	VPS16_DICDI	2.20E-141
117	comp188064_g1_i5	MHCKB_DICDI	7.20E-47
118	comp166254_g1_i2	lipopolysaccharide biosynthesis	1.90E-28
119	comp174958_g1_i2	cystatin-A	2.30E-19
120	comp189162_g1_i2	CRTI_SOYBN	0.00E+00
121	comp181202_g5_i3	peptidyl-prolyl cis-trans isomerase FKBP3	4.40E-29
122	comp194688_g1_i5	AB31G_ARATH	5.30E-141
123	comp180864_g2_i1	PAR14_HUMAN	1.50E-10
124	comp195514_g1_i4	OTU domain-containing 7B isoform X3	4.30E-24
125	comp186290_g1_i1	MMK2_MEDSA	4.90E-123
126	comp196294_g1_i4	YCF45_PORPURecName: Full=Uncharacterized protein	2.30E-20
127	comp192602_g1_i6	CYAA_LEIDO	8.80E-48
128	comp176350_g2_i3	Sjoegren syndrome nuclear autoantigen 1	8.50E-19
129	comp182624_g1_i8	Y1621_SYNY3	1.10E-93
130	comp192340_g1_i5	MCES_SCHPO	2.00E-09
131	comp185229_g3_i2	PR38B_RAT	3.60E-13
132	comp188956_g1_i1	VPS16_DICDI	2.50E-141
133	comp183895_g1_i6	Rieske [2Fe-2S] domain,	2.00E-49
134	comp152441_g2_i1	NAD-dependent alcohol dehydrogenase	1.70E-09
135	comp186690_g1_i3	kif3a predicted protein, partial	3.40E-18
136	comp194698_g1_i5	PRMS_MAIZE	2.10E-11
137	comp183098_g1_i3	Os03g0259700predicted protein, partial	1.70E-40
138	comp188155_g5_i1	NLRC3_HUMAN	1.80E-16
139	comp196283_g1_i1	cytoplasmic dynein 2 heavy chain 1	0.00E+00

140	comp186828_g1_i6	predicted protein	3.40E-43
141	comp195444_g1_i5	SWET1_MOUSE	1.60E-09
142	comp180837_g1_i1	metallophosphoesterase	3.50E-46
143	comp177210_g1_i4	BIRC8_PANTR	3.90E-08
144	comp181142_g1_i2	predicted protein	1.30E-26
145	comp184473_g1_i2	ZRAN1_XENTR	2.30E-30
146	comp193263_g1_i11	FZR2_ARATH	1.00E-88
147	comp193751_g1_i2	TMM43_RAT	2.70E-22
148	comp176603_g1_i4	zinc permease family	2.40E-22
149	comp196225_g1_i1	CYAB_STIAU	5.90E-12
150	comp164300_g1_i2	CYAB_STIAU	8.20E-08
151	comp192701_g2_i3	SWET1_MOUSE	1.50E-09
152	comp189904_g1_i7	peptide methionine sulfoxide reductase msrB	3.60E-63
153	comp195470_g1_i15	NEK4_MOUSE	1.20E-52
154	comp188052_g2_i2	unnamed protein product	1.70E-21
155	comp184394_g1_i4	RTCA_DESK1	2.40E-28
156	comp192586_g1_i7	CRLA_DICDI	1.80E-20
157	comp162540_g1_i5	pap fibrillin family	9.00E-16
158	comp185654_g4_i7	ubiquitin carboxyl-terminal hydrolase 49	1.80E-43
159	comp186188_g1_i1	predicted protein	2.10E-18
160	comp176606_g2_i9	NLRC3_HUMAN	1.20E-40
161	comp187484_g3_i6	hypothetical protein PTSG_01283	4.00E-127
162	comp184920_g1_i2	ANKR1_CHICK	1.40E-10
163	comp180647_g1_i2	lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial	1.20E-73
164	comp196371_g1_i1	TYCC_BREPA	7.40E-40
165	comp187682_g1_i3	GLF_MYCTU	1.00E-73
166	comp164390_g1_i2	S-adenosyl-L-methionine-dependent methyltransferases superfamily isoform 1	2.50E-34
167	comp185922_g4_i4	glutamate carboxypeptidase,	4.50E-53
168	comp185293_g1_i4	MPPD1_HUMAN	1.40E-56
169	comp185832_g1_i1	Hypothetical protein, putative	8.10E-81
170	comp188065_g1_i5	SLBP2_XENLA	1.10E-08
171	comp182105_g1_i4	OTU family cysteine protease	1.60E-35
172	comp189220_g3_i1	mitochondrial transcription termination factor family	6.90E-13

173	comp189959_g1_i8	TMCO3_HUMAN	3.30E-74
174	comp187915_g1_i3	MYCE_MICGR	4.70E-12
175	comp184953_g1_i3	DSF1_YEAST	7.50E-108
176	comp186238_g1_i6	mitochondrial carrier	9.90E-43
177	comp189538_g1_i8	FIXL_RHIME	1.10E-13
178	comp142981_g1_i1	mitochondrial carrier family	8.50E-37
179	comp190355_g1_i3	FIXL_RHIME	6.90E-11
180	comp190652_g1_i1	S16C6_BOVIN	7.30E-14
181	comp190716_g1_i5	alpha-type kinase	2.60E-46
182	comp187514_g3_i10	23S rRNA (adenine(2503)-C(2))-methyltransferase	5.00E-173
183	comp174766_g1_i4	seleno M-like	2.90E-32
184	comp194417_g1_i10	poly(ADP-ribose) polymerase	3.20E-10
185	comp174370_g1_i1	SHQ1_BOVIN	3.70E-31
186	comp177764_g1_i1	FUT8_XENTR	1.40E-18
187	comp192936_g2_i7	DDX11_MOUSE	1.80E-163
188	comp190368_g1_i3	cyclic nucleotide-binding	1.80E-19
189	comp194707_g1_i1	RN213_MOUSE	1.10E-13
190	comp191347_g1_i1	CORA_THEMA	1.30E-14
191	comp172809_g1_i1	tyrosine phosphatase	1.70E-45
192	comp185744_g2_i1	IP6K2_MOUSE	2.20E-19
193	comp190305_g1_i2	hypothetical protein ZOSMA_452G00060	7.40E-10
194	comp172635_g1_i3	UDP-galactose:fucoside alpha-3-galactosyltransferase	6.90E-11
195	comp182885_g2_i2	methyltransferase	4.90E-23
196	comp182973_g1_i7	enhancer of polycomb group	5.00E-43
197	comp161776_g1_i1	N2B_HAEIR	4.30E-75
198	comp195163_g1_i3	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.00E-07
199	comp181182_g1_i2	1A12_CUCMA	7.60E-71
200	comp195767_g2_i7	OTU domain-containing 7B isoform X3	7.20E-24
201	comp195430_g1_i5	telomerase reverse transcriptase catalytic subunit	2.60E-21
202	comp171450_g1_i8	YCX1_EUGGR	1.50E-156
203	comp159621_g1_i1	HTSF1_HUMAN	3.00E-44
204	comp183059_g1_i3	ankyrin repeat domain-containing 27-like	4.40E-11
205	comp173150_g1_i2	hypothetical protein SARC_00571	8.80E-17
206	comp188315_g2_i4	regulator of telomere elongation helicase 1 rtel1,	0.00E+00

207	comp191882_g1_i4	RYR1_HUMAN	2.40E-40
208	comp179244_g1_i2	ESIB_ECOL6	2.00E-18
209	comp190070_g2_i9	major facilitator superfamily (MFS)	1.60E-47
210	comp188382_g1_i1	TYDP2_XENLA	1.60E-37
211	comp169910_g1_i2	hypothetical protein ACA1_171730	5.00E-18
212	comp185399_g1_i6	DNRP_STRPE	2.50E-27
213	comp165938_g1_i1	Hermansky-Pudlak syndrome 1 isoform X1	7.10E-17
214	comp191344_g1_i4	Lmbr1-like motif	9.40E-20
215	comp195476_g1_i5	phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity phosphatase PTEN isoform X2	3.50E-62
216	comp194165_g1_i12	CYA1_RHIME	5.30E-24
217	comp187147_g1_i2	YKP9_SCHPO	5.70E-17
218	comp152293_g1_i2	MMP10_RAT	1.10E-16
219	comp181188_g1_i5	TIL_ARATH	4.00E-10
220	comp191372_g1_i2	CC108_MOUSE	2.40E-132
221	comp195081_g2_i7	SSH_XENLA	1.10E-15
222	comp162489_g1_i1	conserved unknown protein	3.50E-14
223	comp183455_g1_i5	TYTR_TRYBB	0.00E+00
224	comp188065_g1_i3	SLBP2_XENLA	1.10E-08
225	comp195634_g1_i4	guanylate cyclase, putative	1.90E-24
226	comp190612_g1_i3	NUDT5_HUMAN	2.60E-35
227	comp167336_g1_i1	leucine rich repeat domain	3.00E-07
228	comp178558_g1_i3	glycosyltransferase family 1	3.50E-87
229	comp182536_g1_i2	PHOT2_ARATH	3.30E-22
230	comp191679_g1_i10	F213A_XENTR	1.10E-22
231	comp195047_g1_i8	membrane-associated ,	3.30E-07
232	comp171176_g1_i3	XRN4_ARATH	1.90E-73
233	comp187000_g1_i14	kelch domain-containing 3-like	4.30E-13
234	comp193446_g1_i4	lipase domain	5.00E-30
235	comp191353_g1_i3	PREDICTED: uncharacterized protein LOC106175552	1.30E-41
236	comp176976_g1_i9	duf6 domain	4.60E-16
237	comp193325_g1_i5	DEAD DEAH box helicase	9.80E-14
238	comp195263_g1_i1	hybrid sensor histidine kinase response regulator	3.80E-18
239	comp180978_g1_i7	CRNS1_CHICK	2.60E-30
240	comp177684_g1_i5	PREDICTED: uncharacterized protein C17orf53 homolog	1.60E-09

241	comp188651_g1_i6	sulfoquinovosyl transferase SQD2-like	6.00E-33
242	comp189023_g1_i2	PDE11_DROME	5.10E-74
243	comp190203_g1_i1	MTP9_ARATH	1.10E-37

Appendix table 16 genes name of area 16

(a) List of TF family (10 transcripts of 9 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp185793_g1_i5	E2F_DP	8.90E-17
2	comp185793_g1_i7	E2F_DP	8.50E-17
3	comp191474_g1_i5	NF-YC	2.10E-12
4	comp190790_g1_i1	WRKY	6.30E-14
5	comp192206_g1_i2	GRAS	1.10E-13
6	comp188424_g2_i6	bHLH	1.80E-26
7	comp189652_g1_i1	C2H2	3.00E-28
8	comp179973_g1_i4	MYB_related	1.00E-12
9	comp192755_g1_i11	TCP	3.30E-71
10	comp172263_g1_i2	GeBP	3.10E-07

(b) List of other genes name (64 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp195665_g1_i4	uracil-DNA glycosylase	4.30E-16
2	comp193275_g2_i2	transcriptional regulator	1.00E-12
3	comp188515_g1_i2	FRO4_ARATH	2.60E-22
4	comp191128_g1_i2	EFHC2_HUMAN	2.10E-31
5	comp186079_g1_i1	hypothetical protein PTSG_08670	5.20E-32
6	comp167586_g1_i1	Y4967_ARATH	8.50E-51
7	comp182448_g1_i2	pyruvate kinase	2.90E-123
8	comp167787_g1_i1	hypothetical protein VOLCADRAFT_87598	4.90E-50
9	comp186164_g4_i6	AARA_DICDI	3.40E-17
10	comp192436_g1_i11	predicted protein	1.50E-16
11	comp193627_g2_i11	EF-hand calcium-binding domain-containing 6	6.40E-12
12	comp170991_g1_i3	GSTT1_ARATH	9.10E-21
13	comp183962_g1_i11	UDP-galactose:fucoside alpha-3-galactosyltransferase	1.80E-10
14	comp195760_g1_i2	transformation transcription domain-associated	5.30E-57
15	comp162141_g2_i2	AGAL1_PEDPE	4.30E-111
16	comp185525_g1_i4	TRI56_HUMAN	8.10E-09
17	comp195507_g1_i4	MELK_XENLA	2.00E-10

18	comp181985_g1_i3	serine threonine kinase	4.10E-40
19	comp179846_g1_i5	tankyrase isoform X2	2.60E-12
20	comp185343_g2_i2	major facilitator superfamily (MFS)	1.40E-47
21	comp184245_g1_i5	PDRP_LACS1	1.20E-44
22	comp157059_g1_i2	PUR7_BACFN	1.20E-62
23	comp186813_g2_i1	glycosyltransferase aer61	6.10E-26
24	comp176320_g1_i1	PAQR4_MOUSE	1.00E-20
25	comp190245_g1_i3	OGG1_ARATH	2.30E-58
26	comp171374_g1_i2	Interferon-induced 44 family	4.20E-19
27	comp183867_g1_i1	PSM4B_DANRE	5.80E-09
28	comp181743_g2_i1	apical membrane	2.10E-09
29	comp150581_g1_i1	beta-type IP39	9.20E-20
30	comp189196_g2_i1	methyl transferase	4.20E-29
31	comp148315_g1_i1	NLRC3_HUMAN	2.50E-27
32	comp157268_g1_i2	RMGL_PICST	7.40E-22
33	comp190162_g1_i1	YOUD_CAEEL	6.20E-08
34	comp194957_g1_i4	RECQ4_MOUSE	9.50E-76
35	comp175114_g1_i3	AIG2 family	2.10E-18
36	comp177598_g1_i1	UNG_ECOLI	5.70E-76
37	comp190793_g1_i3	glycoside hydrolase	5.70E-93
38	comp170688_g1_i2	Alpha-1,3-mannosyl-glyco 4-beta-N-acetylglucosaminyltransferase, family GT54	3.60E-10
39	comp190219_g1_i6	CYAA_YEAST	9.90E-32
40	comp189696_g2_i3	domain	3.00E-78
41	comp160043_g1_i4	hypothetical protein F443_00246	9.90E-09
42	comp180080_g1_i1	T9SS C-terminal target domain-containing	8.20E-17
43	comp177900_g1_i2	LCYD1_ORYSJ	3.90E-42
44	comp175327_g2_i1	PPIL3_NEUCR	1.90E-10
45	comp167607_g1_i1	glycine cleavage system R	1.60E-15
46	comp164129_g1_i3	ZRAB2_CHICK	2.10E-11
47	comp185930_g4_i1	eukaryotic translation initiation factor 4E	1.40E-09
48	comp178432_g1_i3	OMH4_SCHPO	4.00E-13
49	comp187397_g3_i2	CYA1_RHIME	6.30E-12
50	comp190793_g1_i2	glycoside hydrolase	4.50E-93
51	comp186187_g1_i6	CHO2_ASPFU	3.10E-101
52	comp183344_g1_i2	GCR1_ARATH	2.40E-22

53	comp159493_g1_i4	COX1_CHOFU	1.30E-137
54	comp179060_g1_i5	MGDP1_MOUSE	2.70E-12
55	comp186732_g1_i5	mitochondrial carrier	1.80E-42
56	comp195443_g3_i4	hypothetical protein Ctob_006104	3.70E-08
57	comp184351_g3_i1	hypothetical protein SBOR_8785	9.10E-12
58	comp179701_g1_i4	duf676 domain containing hydrolase	1.80E-21
59	comp189352_g2_i3	hypothetical protein	1.60E-27
60	comp193970_g2_i1	TILS_ENTFA	2.80E-20
61	comp185400_g2_i1	AMT12_SOLLC	1.00E-18
62	comp173457_g4_i1	MOV10_MOUSE	9.50E-60
63	comp185242_g1_i5	TRIM1_HUMAN	3.60E-17
64	comp192103_g2_i2	probable Xaa-Pro aminopeptidase P	0.00E+00

Appendix table 17 genes name of area 17

(a) List of TF family of part A (38 transcripts of 20 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp181355_g1_i1	ARR-B	4.40E-11
2	comp190474_g1_i4	YABBY	4.00E-34
3	comp181355_g1_i3	ARR-B	4.10E-11
4	comp190525_g2_i3	YABBY	4.40E-34
5	comp183220_g1_i1	E2F_DP	3.80E-11
6	comp104672_g1_i2	C3H	2.90E-24
7	comp171796_g1_i2	NF-YA	1.00E-36
8	comp189365_g1_i6	YABBY	1.00E-31
9	comp141811_g1_i1	C2H2	2.00E-16
10	comp175353_g1_i2	WOX	5.60E-33
11	comp149649_g1_i1	MYB	1.30E-75
12	comp160199_g1_i1	NAC	9.50E-22
13	comp146196_g1_i1	NAC	2.60E-113
14	comp157071_g1_i2	bHLH	1.30E-08
15	comp156701_g1_i1	MYB_related	4.40E-59
16	comp182505_g1_i5	E2F_DP	4.00E-11
17	comp179351_g2_i3	MYB_related	3.30E-25
18	comp192674_g1_i7	NAC	4.20E-11
19	comp192585_g1_i5	GRAS	3.00E-09
20	comp193807_g3_i3	NAC	8.00E-11
21	comp179479_g1_i1	HD-ZIP	2.10E-98
22	comp190932_g1_i3	B3	8.60E-39
23	comp192755_g1_i2	TCP	4.60E-71
24	comp190318_g2_i1	bHLH	3.50E-21
25	comp189433_g1_i4	GeBP	9.80E-12
26	comp183220_g1_i5	E2F_DP	4.40E-11
27	comp193012_g1_i1	bHLH	1.90E-22
28	comp185900_g1_i10	bHLH	3.70E-09
29	comp191755_g1_i10	NAC	2.90E-70
30	comp174474_g1_i2	GATA	1.10E-22
31	comp186106_g1_i4	WRKY	6.60E-23
32	comp187431_g2_i1	B3	3.70E-17

33	comp182106_g1_i1	HB-other	1.20E-08
34	comp188771_g1_i6	MYB_related	8.10E-153
35	comp191207_g1_i10	NAC	3.60E-69
36	comp185685_g1_i2	bHLH	2.00E-19
37	comp170694_g1_i1	NF-YB	7.40E-20
38	comp174026_g1_i4	C3H	1.30E-15

(b) List of other genes name of part A (188 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp174929_g1_i8	RFA1_HUMAN	2.90E-91
2	comp175752_g1_i2	RNA-binding family isoform 1	3.70E-09
3	comp189794_g1_i3	AB2D_ARATH	3.80E-112
4	comp193607_g1_i3	unnamed product	7.90E-62
5	comp182634_g1_i5	carnitine O-acetyltransferase	3.20E-136
6	comp158945_g1_i3	mitochondrial transcription termination factor family	7.40E-10
7	comp175194_g1_i3	DIOXL_ARATH	3.50E-42
8	comp170699_g1_i4	non-canonical poly(A) RNA polymerase PAPD7-like	1.60E-38
9	comp194489_g2_i6	Y663_METTH	6.50E-76
10	comp171241_g1_i2	integrator complex subunit 11	0.00E+00
11	comp184027_g1_i13	NLRC3_HUMAN	7.70E-54
12	comp175540_g1_i2	IP3KA_RAT	1.10E-40
13	comp176156_g1_i1	sodium- and chloride-dependent GABA transporter 2-like	5.60E-14
14	comp188870_g1_i4	snf1-related kinase catalytic subunit alpha kin10	2.50E-136
15	comp191011_g1_i1	tetratricopeptide repeat domain containing	3.30E-135
16	comp185979_g1_i2	coproporphyrinogen III oxidase	4.30E-78
17	comp164077_g1_i1	chloroplast envelope membrane	5.20E-42
18	comp187387_g1_i6	multiple C2 and transmembrane domain-containing 1	2.40E-09
19	comp195079_g5_i3	MET22_HUMAN	4.20E-12
20	comp176906_g1_i1	LCYB_ARATH	3.00E-67
21	comp194976_g2_i5	exocyst complex component 1 isoform X2	4.90E-12
22	comp193741_g1_i1	ADIA_ECOLI	1.80E-36
23	comp187558_g2_i3	predicted protein	8.80E-10
24	comp194683_g1_i2	adenylate cyclase type 10	2.90E-18
25	comp184682_g2_i6	methylenetetrahydrofolate reductase	3.90E-16

26	comp187682_g2_i2	UDP-galactopyranose mutase	8.20E-85
27	comp178369_g1_i3	YRHG_BACSU	3.00E-12
28	comp194509_g1_i2	HELLS_HUMAN	6.20E-154
29	comp196206_g3_i2	receptor-type adenylate cyclase GRESAG 4,	2.00E-57
30	comp187551_g1_i4	expressed unknown protein	1.30E-39
31	comp187755_g2_i2	hypothetical protein CHLREDRAFT_206193	5.80E-07
32	comp189904_g1_i5	MSRB5_ARATH	2.10E-40
33	comp183732_g2_i2	CYAA_LEIDO	4.20E-31
34	comp173359_g1_i1	predicted protein	2.30E-17
35	comp182446_g1_i9	metallophosphoesterase domain-containing 1	3.50E-61
36	comp193872_g2_i5	nicotinate-nucleotide adenylyltransferase	1.50E-123
37	comp179168_g1_i6	PTR36_ARATH	1.10E-32
38	comp175490_g1_i2	ODA1_CHLRE	7.10E-43
39	comp196073_g1_i4	adenylate cyclase	1.20E-22
40	comp177184_g1_i1	CJ011_MOUSE	4.60E-15
41	comp184392_g1_i2	cilia- and flagella-associated 69	3.10E-62
42	comp195922_g1_i6	CYA1_RHIME	1.40E-24
43	comp193711_g1_i2	MDR1_HUMAN	1.40E-130
44	comp186653_g2_i8	SGK2_MOUSE	1.60E-34
45	comp183860_g1_i2	FLU_ARATH	5.40E-17
46	comp190848_g2_i2	PHIP_MOUSE	2.40E-23
47	comp178417_g1_i1	STRG_STRGR	8.60E-12
48	comp169146_g1_i2	transporter permease	1.20E-17
49	comp158183_g1_i3	ODC_DICDI	3.00E-65
50	comp155077_g1_i4	hypothetical protein CHLNCDRAFT_145050	5.90E-59
51	comp173193_g1_i1	tetratricopeptide repeat	1.20E-45
52	comp174296_g1_i3	TCPR1_RAT	3.40E-07
53	comp143128_g1_i4	unnamed protein product	2.30E-08
54	comp192335_g4_i2	YR831_MIMIV	2.20E-11
55	comp161419_g1_i1	ferredoxin-NADP+ reductase	2.20E-94
56	comp189051_g1_i1	macrocin O-methyltransferase	8.70E-35
57	comp176373_g1_i1	cdpalcohol phosphatidyltransferase	3.40E-110
58	comp193906_g1_i1	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07
59	comp185560_g1_i2	PEK GCN2 kinase	9.50E-21
60	comp181985_g1_i7	serine threonine kinase	6.10E-33

61	comp195371_g1_i9	TARB1_HUMAN	4.10E-19
62	comp176060_g1_i8	glycosyl transferase	4.50E-18
63	comp191132_g2_i3	pantothenate kinase 1	4.60E-52
64	comp182012_g1_i8	type III domain and Immunoglobulin-like fold domain-containing	5.10E-08
65	comp191839_g1_i3	ANR50_HUMAN	1.70E-19
66	comp160438_g1_i3	AZOR_DECAR	2.20E-67
67	comp182957_g1_i4	PANK_YEAST	9.10E-15
68	comp185375_g1_i6	IF4E2_MOUSE	1.50E-18
69	comp181990_g1_i2	Hypothetical protein, putative	5.90E-16
70	comp167577_g2_i2	ankyrin repeat	8.60E-57
71	comp194110_g1_i7	hypothetical protein	4.10E-20
72	comp188879_g1_i7	cAMP specific phosphodiesterase	7.20E-172
73	comp170807_g2_i3	secreted partial	4.80E-12
74	comp154614_g1_i1	RPGR_CANFA	9.00E-19
75	comp194250_g1_i7	PFL_CLOPA	0.00E+00
76	comp194047_g2_i2	AMP-dependent synthetase	6.10E-20
77	comp150691_g1_i2	riken cdna 2700097o09	5.00E-55
78	comp181036_g1_i1	unnamed product	5.10E-16
79	comp195043_g2_i3	serine threonine kinase	5.80E-10
80	comp180227_g2_i2	ELF1_SCHPO	0.00E+00
81	comp189232_g1_i2	hypothetical protein O9G_000036	5.40E-100
82	comp175449_g1_i1	NSR1_YEAST	1.90E-11
83	comp174430_g2_i1	dual specificity phosphatase 22-like	2.30E-08
84	comp55141_g1_i1	unnamed product	7.10E-12
85	comp163197_g1_i2	RBMX2_RAT	3.00E-47
86	comp171334_g2_i2	kinase-like domain	1.50E-11
87	comp176037_g1_i1	inositol 1,4,5-trisphosphate receptor type 3	4.50E-15
88	comp192163_g1_i5	Y1796_ARATH	1.40E-82
89	comp182507_g1_i16	DDL_RALME	3.20E-09
90	comp192307_g1_i6	predicted protein	5.40E-17
91	comp187147_g1_i3	cellular retinaldehyde-binding triple function domain-containing	7.30E-24
92	comp174227_g1_i2	COXZ_RECAM	7.30E-37
93	comp188008_g1_i4	RING finger	1.50E-36
94	comp186313_g2_i3	uracil-DNA glycosylase	3.20E-36
95	comp160660_g1_i1	DB10_NICSY	6.60E-53

96	comp181220_g1_i2	DPOA2_MOUSE	3.20E-26
97	comp194553_g1_i5	AL3B1_HUMAN	5.40E-121
98	comp192685_g3_i2	KAPR_BLAEM	7.50E-09
99	comp182512_g1_i6	YPL1_DROME	1.30E-09
100	comp195265_g1_i12	PUB15_ARATH	4.60E-08
101	comp174212_g1_i1	Splicing factor U2AF-associated 2	6.00E-37
102	comp189059_g1_i10	serine threonine- kinase STK11 isoform X1	1.50E-64
103	comp189165_g1_i2	histidine kinase	5.10E-13
104	comp188008_g1_i2	RN216_MOUSE	6.80E-15
105	comp182433_g3_i2	hypothetical protein GUITHDRAFT_131487	9.90E-22
106	comp186626_g1_i3	RFC1_HUMAN	3.20E-89
107	comp187028_g2_i2	phospho phosphatase	2.60E-12
108	comp186649_g1_i3	2-heptaprenyl-1,4-naphthoquinone methyltransferase	2.80E-10
109	comp190512_g1_i2	Y8236_DICDI	2.50E-99
110	comp184954_g2_i6	TRI18_HUMAN	3.10E-21
111	comp179584_g1_i2	Y777_CHLL2	3.20E-11
112	comp192838_g1_i2	PDR3_TOBAC	1.30E-44
113	comp177716_g1_i2	formate transporter	8.50E-114
114	comp163041_g1_i1	YCX91_PHAAO	2.10E-18
115	comp194840_g2_i3	WD40 repeat-containing ,	3.20E-40
116	comp187136_g2_i5	PROP_SALTY	7.50E-28
117	comp178805_g3_i1	Fe-S oxidoreductase	6.90E-168
118	comp191677_g3_i1	mannosylphosphorylation (Mnn4),	4.40E-11
119	comp185271_g1_i1	LIP_THELA	2.80E-10
120	comp180843_g3_i2	ankyrin domain predicted protein, partial	2.10E-15
121	comp154783_g1_i1	OGFR_MOUSE	8.80E-36
122	comp186214_g1_i2	TTD14_DROME	3.40E-74
123	comp166867_g1_i2	DBP5_LODEL	1.50E-48
124	comp165323_g2_i3	PSBA_EUGGR	0.00E+00
125	comp180530_g1_i3	SPOT_ECO57	1.30E-75
126	comp183277_g1_i5	TKL IRAK kinase	8.10E-29
127	comp190716_g1_i1	AK1_DICDI	8.70E-41
128	comp186187_g1_i10	phosphatidylethanolaminen-methyltransferase-lik e	1.10E-133
129	comp184647_g2_i2	IF4E2_MOUSE	7.00E-15
130	comp189912_g2_i2	PYRD_RAT	2.10E-121

131	comp187334_g2_i2	AK1_DICDI	6.00E-34
132	comp174015_g1_i4	rhodanese	4.00E-15
133	comp192508_g1_i4	ANK3_MOUSE	7.00E-08
134	comp189812_g1_i10	PREDICTED: uncharacterized protein LOC106806144	2.40E-20
135	comp184300_g1_i2	DNA repair endonuclease XPF-like	6.60E-73
136	comp184549_g2_i2	CYA1_RHIME	4.50E-14
137	comp192886_g2_i2	SKI2_YEAST	9.10E-28
138	comp166886_g1_i2	EF2K_RAT	4.00E-53
139	comp149282_g1_i1	Y108_SYNY3	4.00E-10
140	comp184418_g1_i1	Proteophosphoglycan ppg4	5.00E-35
141	comp170185_g1_i2	hypothetical protein PPERSA_09341	2.30E-13
142	comp184689_g2_i5	hypothetical protein DQ04_09241000	2.60E-43
143	comp193410_g1_i4	DNA mismatch repair	6.40E-119
144	comp186928_g4_i2	MCATL_BOVIN	1.80E-24
145	comp195863_g1_i9	membrane-associated partial	3.10E-09
146	comp181303_g1_i6	EF-Hand calcium-binding site	1.30E-86
147	comp189391_g1_i4	EMAL6_HUMAN	2.60E-83
148	comp164091_g1_i1	lateral signaling target 2 homolog	5.40E-07
149	comp193275_g2_i8	Y663_CHLTE	3.00E-11
150	comp186814_g1_i1	ABC transporter ATP-binding	1.80E-70
151	comp194007_g1_i4	PABP6_ARATH	1.40E-13
152	comp185385_g1_i3	CACP_RAT	1.80E-87
153	comp194154_g2_i3	PDE1A_MOUSE	1.00E-61
154	comp185293_g1_i10	MPPD1_HUMAN	4.60E-55
155	comp194364_g1_i1	E3 ubiquitin- ligase PRT1-like	3.10E-44
156	comp174929_g1_i7	RFA1_HUMAN	3.20E-91
157	comp188471_g1_i2	SYG_LEPBJ	2.40E-140
158	comp176049_g1_i1	YADS_AERCA	5.30E-11
159	comp193694_g1_i4	cold-shock protein	1.10E-07
160	comp180732_g5_i2	hypothetical protein THAOC_06282predicted protein, partial	9.40E-26
161	comp163175_g1_i2	TAR3_ARATH	1.60E-70
162	comp181437_g1_i3	fatty acyl- reductase	0.00E+00
163	comp181997_g1_i2	dynein heavy chain axonemal	6.90E-10
164	comp178851_g2_i3	ribulose-1,5 bisphosphate carboxylase oxygenase large subunit N-methyltransferase, chloroplastic-like	7.80E-42

165	comp193535_g1_i3	ADCYA_RAT	3.10E-37
166	comp189244_g1_i11	hypothetical protein ACA1_171730	3.30E-23
167	comp168871_g1_i5	YKX5_SCHPO	1.30E-11
168	comp185224_g1_i2	UBA1_HUMAN	0.00E+00
169	comp185561_g1_i4	predicted protein, partial	0.00E+00
170	comp162540_g1_i4	pap fibrillin family	8.80E-16
171	comp177551_g1_i1	ASHR1_ARATH	7.00E-27
172	comp192089_g2_i3	GPR107-like	2.60E-80
173	comp177141_g2_i2	hypothetical protein	3.20E-08
174	comp190830_g1_i4	MKT1_YEAST	3.00E-11
175	comp184855_g1_i3	embryogenesis-associated EMB8	2.20E-71
176	comp154198_g1_i1	Hypothetical protein, putative	1.40E-09
177	comp185437_g3_i1	predicted protein	3.90E-09
178	comp196197_g4_i1	nfx1-type zinc finger-containing 1	1.00E-17
179	comp192832_g1_i10	lipase domain	3.90E-08
180	comp183285_g1_i2	ECERIFERUM 3	2.70E-175
181	comp180933_g1_i5	PSAB_EUGGR	2.90E-78
182	comp170762_g1_i1	COQ9_DROME	5.60E-36
183	comp194634_g2_i3	DYHC_DICDI	5.10E-85
184	comp180507_g1_i1	hypothetical protein GUITHDRAFT_121297	1.50E-30
185	comp175891_g2_i4	sodium-coupled neutral amino acid transporter 11	1.30E-55
186	comp188898_g1_i5	hypothetical protein MVEG_10442	1.30E-26
187	comp173764_g1_i2	NIPSNAP-related (ISS)	2.70E-27
188	comp194716_g1_i3	xylanase	1.80E-128

(c) List of TF family of part B (2 transcripts of 2 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp187971_g1_i1	C3H	5.00E-09
2	comp188366_g1_i3	G2-like	2.20E-08

(d) List of other genes name of part B (28 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp176712_g1_i1	SPNS1_ARATH	5.00E-17

2	comp189367_g1_i6	KHA1_YEAST	9.20E-22
3	comp185179_g1_i3	rossmann-fold nad -binding domain-containing	1.00E-50
4	comp190947_g1_i4	cation-chloride cotransporter family	0.00E+00
5	comp147297_g1_i1	fungal specific transcription	5.30E-13
6	comp171957_g1_i4	mitochondrial import inner membrane translocase subunit Tim17-B	7.50E-31
7	comp188085_g1_i7	kinase,	2.90E-49
8	comp188214_g1_i8	S38A5_MOUSE	4.30E-18
9	comp173619_g1_i2	CI171_DANRERecName: Full=Uncharacterized protein C9orf171 homolog	8.80E-09
10	comp182507_g1_i13	D-alanine--D-alanine ligase	1.60E-141
11	comp77291_g1_i2	titin-like isoform X1	2.80E-15
12	comp191601_g3_i1	PR40A_ARATH	1.40E-49
13	comp194250_g1_i2	formate acetyltransferase	0.00E+00
14	comp186281_g1_i3	serine threonine- kinase STN7, chloroplastic	7.80E-95
15	comp190973_g2_i14	histone-lysine N-methyltransferase SETMAR-like	2.90E-30
16	comp174929_g1_i9	RFA1_HUMAN	8.80E-91
17	comp179941_g1_i1	radical SAM domain containing	4.00E-93
18	comp192340_g1_i3	mRNA-capping enzyme subunit alpha	1.40E-13
19	comp184027_g1_i1	NLRC3_HUMAN	3.20E-54
20	comp184358_g1_i5	CNRN_DICDI	1.00E-48
21	comp176712_g1_i2	SPNS1_ARATH	4.90E-17
22	comp188507_g1_i3	carbamoyltransferase	6.60E-46
23	comp185776_g1_i3	KLHDB_CULQU	6.70E-30
24	comp167820_g1_i2	hypothetical protein M569_08619	7.40E-11
25	comp180056_g2_i4	CAS1 domain-containing 1-like	5.10E-124
26	comp178171_g1_i5	Neurotrypsin	3.20E-07
27	comp189089_g1_i2	something that sticks like isoform B	1.30E-15
28	comp179878_g2_i4	PI42C_DANRE	1.10E-09

Appendix table 18 genes name of area 18

(a) List of TF family (82 transcripts of 24 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp193807_g3_i12	NAC	8.10E-11
2	comp185793_g1_i2	E2F_DP	1.20E-16
3	comp185793_g1_i8	E2F_DP	9.20E-17
4	comp189067_g1_i6	G2-like	1.80E-08
5	comp189590_g1_i1	GRAS	7.20E-08
6	comp192585_g1_i11	GRAS	1.80E-09
7	comp161809_g1_i1	HD-ZIP	9.00E-10
8	comp190387_g2_i3	FAR1	3.00E-43
9	comp193452_g1_i5	MYB_related	4.30E-58
10	comp193452_g1_i7	MYB_related	4.90E-58
11	comp193263_g1_i1	MYB_related	6.10E-58
12	comp180471_g1_i2	C3H	5.00E-56
13	comp194020_g1_i5	NAC	3.70E-49
14	comp192870_g1_i4	NAC	2.30E-09
15	comp182293_g3_i1	NF-YA	1.70E-13
16	comp191415_g1_i1	NF-YA	1.00E-35
17	comp162887_g1_i1	NF-YB	3.40E-55
18	comp194746_g1_i1	Nin-like	8.90E-14
19	comp194688_g1_i4	Nin-like	2.80E-33
20	comp191325_g2_i5	FAR1	3.20E-27
21	comp186653_g3_i8	YABBY	3.00E-18
22	comp187333_g1_i4	bHLH	3.20E-17
23	comp189433_g1_i1	GeBP	9.00E-12
24	comp189654_g2_i2	WRKY	1.40E-13
25	comp192336_g2_i1	YABBY	3.30E-25
26	comp187762_g1_i5	bHLH	6.10E-09
27	comp190348_g1_i4	bHLH	6.70E-28
28	comp194693_g1_i2	bHLH	3.40E-09
29	comp193326_g3_i3	B3	1.30E-75
30	comp191633_g1_i1	B3	1.70E-59
31	comp191902_g1_i3	B3	6.90E-08
32	comp189056_g1_i5	C2H2	3.30E-19

33	comp189090_g1_i6	C2H2	6.30E-09
34	comp169387_g1_i2	C3H	4.40E-07
35	comp194667_g1_i1	EIL	1.30E-09
36	comp194667_g1_i7	EIL	8.40E-10
37	comp193232_g1_i1	ERF	1.20E-18
38	comp190377_g2_i1	FAR1	1.10E-08
39	comp190377_g2_i3	FAR1	1.10E-08
40	comp182026_g1_i4	FAR1	9.80E-141
41	comp181870_g2_i2	FAR1	3.60E-19
42	comp191274_g2_i4	G2-like	6.00E-10
43	comp189001_g3_i5	GRAS	1.80E-25
44	comp192107_g1_i1	LBD	5.60E-30
45	comp195200_g1_i10	MYB	5.60E-10
46	comp179973_g1_i1	MYB_related	3.00E-17
47	comp179973_g1_i3	MYB_related	9.20E-13
48	comp193139_g1_i9	MYB_related	1.20E-28
49	comp190879_g1_i1	NF-YA	1.20E-39
50	comp196174_g1_i3	Nin-like	2.40E-46
51	comp189823_g1_i2	WRKY	1.20E-22
52	comp189174_g1_i3	WRKY	4.00E-16
53	comp189059_g1_i1	WRKY	1.30E-18
54	comp188792_g1_i3	WRKY	5.90E-71
55	comp187439_g1_i2	bHLH	1.00E-128
56	comp187208_g2_i6	bHLH	4.20E-26
57	comp193740_g1_i2	BBR-BPC	7.60E-22
58	comp188326_g1_i10	C2H2	2.40E-29
59	comp169488_g1_i3	C3H	4.90E-17
60	comp189090_g1_i8	C2H2	6.30E-09
61	comp190327_g1_i4	FAR1	4.70E-47
62	comp195103_g2_i6	G2-like	1.10E-07
63	comp192381_g2_i2	HD-ZIP	1.00E-17
64	comp181961_g1_i1	FAR1	2.00E-11
65	comp179973_g1_i2	MYB_related	9.20E-13
66	comp194020_g1_i2	NAC	4.40E-49
67	comp194020_g1_i6	NAC	1.50E-51

68	comp195082_g1_i2	Nin-like	9.00E-14
69	comp190932_g1_i5	B3	7.90E-39
70	comp170041_g1_i1	C3H	1.40E-15
71	comp187849_g1_i4	VOZ	1.00E-28
72	comp188286_g1_i3	YABBY	3.20E-32
73	comp183451_g2_i4	YABBY	2.90E-31
74	comp183451_g2_i7	YABBY	3.80E-31
75	comp179592_g1_i2	MYB_related	2.60E-45
76	comp184679_g1_i3	Trihelix	2.90E-07
77	comp192065_g1_i2	B3	3.70E-32
78	comp189859_g1_i3	VOZ	6.40E-23
79	comp190790_g2_i2	WRKY	6.10E-14
80	comp194125_g1_i3	bHLH	3.90E-14
81	comp194553_g1_i2	bHLH	2.70E-14
82	comp188424_g2_i1	bHLH	1.80E-26

(b) List of other genes name (558 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp163952_g1_i5	DOHH_DROME	1.40E-71
2	comp196212_g1_i5	hypothetical protein PBRA_002490	3.50E-99
3	comp187494_g1_i5	BETC_RHIME	2.00E-17
4	comp172528_g1_i1	histone lysine methyltransferase, SET	4.90E-18
5	comp190816_g2_i3	HTRB_BACSU	1.40E-07
6	comp187161_g1_i2	hypothetical protein GUITHDRAFT_105708	5.80E-10
7	comp178069_g2_i6	enoyl- hydratase	1.50E-77
8	comp195350_g1_i3	zeta partial	2.60E-35
9	comp193198_g2_i5	hypothetical protein MNEG_6228	2.10E-08
10	comp194189_g1_i3	5NTD_LUTLO	3.30E-32
11	comp184037_g1_i3	collagen alpha-6(VI) chain-like	2.70E-09
12	comp164108_g1_i3	spermatogenesis-associated 17	8.30E-20
13	comp171241_g1_i1	integrator complex subunit 11	2.40E-170
14	comp192584_g1_i1	hypothetical protein GM43_2235	3.10E-13
15	comp182507_g1_i7	DDL_RALME	4.20E-09
16	comp190991_g1_i2	AC010924_18 ESTs gb	9.40E-34

17	comp195163_g1_i11	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.00E-07
18	comp190031_g1_i7	PREDICTED: uncharacterized protein LOC106806144	5.50E-20
19	comp194733_g1_i5	SPNS1_ARATH	2.80E-19
20	comp195414_g1_i10	PABP6_ARATH	1.50E-13
21	comp189292_g1_i3	hypothetical protein COCSUDRAFT_55703	1.50E-10
22	comp196040_g1_i3	PAS domain-containing sensor histidine kinase	7.10E-22
23	comp193577_g1_i10	SPXS1_DICDI	4.30E-45
24	comp186672_g2_i13	MCE_MIMIV	9.10E-22
25	comp193223_g1_i3	FIXL_BRADU	6.00E-09
26	comp186779_g2_i4	tyrosine phosphatase	3.50E-67
27	comp181589_g1_i13	PDPK1_RAT	1.70E-74
28	comp171107_g1_i5	ATP-binding Cassette (ABC) superfamily	8.00E-14
29	comp195858_g1_i4	proteasome activator complex subunit 4	1.40E-42
30	comp191770_g2_i2	AP-4 complex subunit mu-like	7.80E-39
31	comp134569_g1_i1	Y1181_ARATH	1.10E-31
32	comp182185_g5_i1	conserved unknown protein	2.50E-24
33	comp183642_g1_i4	nuclear inhibitor of phosphatase 1	3.10E-16
34	comp175453_g1_i3	e3 ubiquitin- ligase	1.50E-13
35	comp167949_g1_i1	SNF8_XENTR	2.70E-51
36	comp175448_g1_i5	hypothetical protein CHLNCDRAFT_138448	9.50E-10
37	comp192508_g1_i1	ANK3_MOUSE	3.70E-08
38	comp176930_g1_i2	RAD53_YEAST	4.20E-09
39	comp182957_g1_i1	PANK_YEAST	1.90E-20
40	comp98148_g1_i1	COX2_LASSP	2.20E-76
41	comp181655_g2_i1	TBL11_ARATH	2.70E-11
42	comp179000_g4_i1	adenylate cyclase	1.30E-19
43	comp188962_g1_i3	class II glutamine amidotransferase	1.30E-95
44	comp192568_g1_i9	ADP-ribosyltransferase 1 precursor	1.90E-33
45	comp182889_g2_i3	unknown protein	9.30E-18
46	comp187624_g1_i1	hypothetical protein COCSUDRAFT_45614	1.50E-16
47	comp181223_g1_i2	PREDICTED: uncharacterized protein LOC103046511 isoform X1	9.40E-81
48	comp180251_g1_i6	MHCKA_DICDI	2.10E-41
49	comp163780_g1_i2	hypothetical protein GUITHDRAFT_158222	2.30E-14
50	comp181464_g1_i3	PARP2_ARATH	5.70E-10
51	comp191449_g1_i3	hypothetical protein CHLNCDRAFT_138448	6.30E-09

52	comp176024_g1_i7	CAN15_MOUSE	3.30E-41
53	comp193202_g2_i2	ADCY8_HUMAN	5.00E-15
54	comp194060_g1_i1	CAN12_HUMAN	2.50E-20
55	comp176177_g1_i2	ZFY16_HUMAN	1.70E-09
56	comp182174_g1_i1	maleylacetoacetate isomerase isoform X1	7.80E-56
57	comp196261_g3_i2	IF4G2_ARATH	8.70E-19
58	comp195163_g1_i1	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.00E-07
59	comp195632_g1_i1	Tbc2 translation chloroplastic	4.10E-39
60	comp183474_g2_i2	SWET1_XENTR	1.80E-24
61	comp178442_g1_i5	CRTY_ARTPT	6.70E-95
62	comp189490_g1_i3	SET domain-containing	2.50E-28
63	comp187184_g1_i1	SWET1_XENTR	6.00E-24
64	comp141217_g1_i1	uv radiation resistance-associated	6.00E-16
65	comp183649_g1_i5	glycosyl transferase	2.00E-18
66	comp196045_g3_i5	p-type ATPase superfamily	8.40E-167
67	comp176106_g2_i1	MESA_EMENI	9.00E-14
68	comp178860_g1_i3	PARG_DROME	4.60E-47
69	comp175279_g1_i2	RBL18_ARATH	1.10E-24
70	comp194154_g2_i2	3 ,5 -cyclic-nucleotide phosphodiesterase	4.80E-123
71	comp185096_g9_i2	CAPTC_DICDI	1.40E-28
72	comp194379_g1_i1	guanylate cyclase	5.50E-64
73	comp192517_g1_i3	FIXL_RHIME	5.50E-12
74	comp193353_g2_i5	NOA1_ARATH	1.10E-42
75	comp191850_g1_i3	MENA_SYNY3	4.00E-49
76	comp171450_g3_i1	PSBC_EUGGR	0.00E+00
77	comp170734_g1_i5	calpain-like cysteine peptidase	9.40E-17
78	comp187437_g2_i2	exostosin	3.50E-15
79	comp181987_g2_i1	Hypothetical protein, putative	2.60E-11
80	comp195863_g1_i1	membrane-associated ,	1.60E-12
81	comp183455_g1_i6	TYTR_TRYBB	0.00E+00
82	comp192830_g1_i3	mannosylphosphorylation (Mnn4),	1.10E-28
83	comp169745_g1_i2	MFS transporter, sugar porter family	1.30E-73
84	comp155796_g1_i4	Mitochondrial transcription termination factor family	6.30E-09
85	comp192442_g1_i1	adenylate cyclase type 10	7.00E-39
86	comp181585_g1_i3	kelch 5	7.00E-11

87	comp162865_g1_i3	2ABD_CHICK	3.80E-118
88	comp170754_g1_i3	isoform a	9.10E-21
89	comp179132_g1_i2	vacuolar ATP synthase	1.90E-26
90	comp190494_g1_i4	S38A5_MOUSE	1.10E-17
91	comp190563_g1_i3	AK1_DICDI	2.10E-44
92	comp191912_g1_i10	AARA_DICDI	9.80E-11
93	comp178432_g1_i1	glycosyltransferase family 15	1.10E-32
94	comp192793_g3_i1	EDRF1_PONAB	9.30E-36
95	comp193359_g1_i2	CNRB_DICDI	4.00E-101
96	comp191055_g1_i16	TPR domain-containing	1.10E-11
97	comp185389_g3_i3	CY43_TRYBB	2.00E-15
98	comp193603_g1_i3	TMCA_NITHN	6.50E-50
99	comp182313_g2_i2	glycosyltransferase aer61	1.30E-31
100	comp179782_g1_i1	probable vesicular acetylcholine transporter-B	1.10E-23
101	comp191277_g1_i1	integral membrane	6.00E-25
102	comp194775_g2_i1	STT7_CHLRE	2.30E-100
103	comp190428_g4_i3	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1 isoform X2	1.80E-84
104	comp167300_g1_i1	KAPR_DICDI	1.20E-74
105	comp183533_g1_i6	hypothetical protein SAMD00019534_007780predicted protein, partial	8.30E-15
106	comp186058_g1_i2	family methyltransferase	1.90E-16
107	comp194283_g1_i1	QKY_ARATH	2.80E-16
108	comp177126_g2_i2	glucosephosphate-mutase GPM2	2.90E-48
109	comp191344_g1_i2	Lmbr1-like motif	6.60E-48
110	comp192632_g1_i3	hypothetical protein GUITHDRAFT_139529	4.30E-71
111	comp183383_g1_i3	NLRC3_HUMAN	1.90E-45
112	comp185518_g1_i3	PPDEX_ARATH	4.80E-23
113	comp191706_g1_i5	peroxidase	3.00E-88
114	comp193186_g1_i1	MORN repeat-containing 5	2.00E-23
115	comp193641_g1_i2	ABC transporter D family	0.00E+00
116	comp185411_g2_i2	FER_GLEJA	1.30E-10
117	comp190238_g1_i2	kinase domain	8.30E-07
118	comp189588_g2_i3	beta-glucan elicitor receptor	0.00E+00
119	comp192932_g2_i5	HECD1_HUMAN	2.00E-09
120	comp192829_g1_i2	Inositol-trisphosphate 3-kinase Apredicted protein, partial	9.90E-48
121	comp185517_g1_i3	conserved hypothetical protein	9.70E-40

122	comp195268_g1_i7	hypothetical protein	2.20E-23
123	comp192964_g1_i2	Tbc2 translation chloroplastic	2.70E-30
124	comp176566_g1_i5	DSP8_ARATH	6.80E-18
125	comp185469_g1_i1	major facilitator superfamily	1.10E-67
126	comp191010_g1_i3	VTC4_SCHPO	7.50E-63
127	comp167863_g1_i8	RL21_CALS4	2.50E-25
128	comp158867_g3_i3	F26_ARATH	1.30E-66
129	comp189338_g1_i12	membrane-bound adenylyl cyclase	2.30E-39
130	comp190515_g2_i4	LON peptidase N-terminal domain and RING finger 1-like	1.90E-12
131	comp188131_g2_i1	TIC32_PEA	1.30E-52
132	comp194807_g1_i7	pseudouridine synthase	1.80E-07
133	comp162429_g1_i3	NIMA-related kinase	6.60E-10
134	comp189197_g3_i1	NLRC3_HUMAN	1.30E-07
135	comp186168_g1_i2	Alpha beta-Hydrolases superfamily isoform 1	2.80E-09
136	comp176060_g1_i1	glycosyl transferase	2.00E-17
137	comp189487_g2_i3	alpha beta-hydrolase	2.40E-20
138	comp185023_g6_i6	transporter, major facilitator family	1.40E-81
139	comp187503_g1_i2	EDRF1_MOUSE	2.40E-32
140	comp187997_g1_i7	WDR35_HUMAN	4.40E-22
141	comp193823_g1_i3	BPRX_DICNO	4.50E-72
142	comp192775_g2_i4	coproporphyrinogen III oxidase	2.00E-64
143	comp170251_g1_i2	IF4E_CANGA	7.30E-15
144	comp194361_g1_i4	hypothetical protein EMIHUDRAFT_237418	1.70E-09
145	comp179184_g2_i7	KDSR_HUMAN	1.40E-53
146	comp174870_g1_i1	BTB POZ domain-containing	2.30E-17
147	comp184450_g1_i2	FAEB_PIREQ	2.30E-10
148	comp189726_g1_i3	predicted protein	1.50E-47
149	comp193901_g2_i3	Carbohydrate-binding and sugar hydrolysis	7.40E-09
150	comp182810_g1_i5	Hypothetical protein, putative	3.80E-09
151	comp168035_g1_i1	receptor-type adenylate cyclase a	1.90E-14
152	comp183592_g1_i4	LON peptidase N-terminal domain and RING finger 2predicted protein, partial	2.50E-46
153	comp191795_g1_i18	CDO1_DANRE	3.00E-14
154	comp190830_g1_i1	XPG I-region	5.20E-69
155	comp186813_g1_i1	glycosyltransferase aer61	1.00E-26
156	comp191641_g1_i1	hypothetical protein COCSUDRAFT_45614	1.50E-14

157	comp183323_g1_i4	Coatomer subunit beta -3	1.00E-24
158	comp160629_g1_i2	ubiquitin family	2.20E-07
159	comp176212_g1_i1	TBC20_BOVIN	4.30E-47
160	comp195665_g1_i8	uracil-DNA glycosylase	4.30E-16
161	comp186228_g1_i6	M3K2_ARATH	1.70E-77
162	comp195984_g1_i2	nxn protein	0.00E+00
163	comp195420_g1_i7	IF4E_DICDI	1.70E-11
164	comp172028_g1_i3	NLRC3_HUMAN	1.30E-40
165	comp183741_g1_i2	calmodulin mutant syncam9	8.40E-09
166	comp178031_g1_i2	carbon catabolite repressor 4-like 4	1.50E-46
167	comp190280_g1_i3	DNA-directed primase polymerase	3.00E-33
168	comp166967_g1_i4	GP157_MOUSE	2.00E-08
169	comp187981_g2_i1	WD-40 repeat	1.70E-18
170	comp187784_g1_i2	NSUN2_CHICK	1.80E-111
171	comp179808_g1_i4	UGGG_DROME	8.10E-168
172	comp184385_g1_i6	IF4E3_ORYSJ	2.60E-10
173	comp188869_g1_i6	probable palmitoyltransferase ZDHHC1 isoform X1	1.70E-39
174	comp189241_g1_i5	CYAA_NEUCR	5.30E-09
175	comp184803_g1_i4	alcohol acetyltransferase	8.30E-38
176	comp160278_g1_i1	SIN3A_MOUSE	1.10E-35
177	comp163237_g1_i2	conserved unknown protein	2.80E-14
178	comp174011_g2_i3	hydroxyacylglutathione hydrolase	7.70E-112
179	comp192778_g1_i7	nicotinate-nucleotide adenylyltransferase	1.80E-123
180	comp181997_g1_i1	Dynein heavy chain axonemal	7.10E-16
181	comp193924_g1_i1	dynein heavy chain 5, axonemal	0.00E+00
182	comp189220_g1_i5	mitochondrial transcription termination factor family	7.10E-13
183	comp190368_g1_i7	cyclic nucleotide-binding	1.90E-19
184	comp175729_g1_i6	LIM zinc-binding	2.30E-19
185	comp173250_g1_i2	GMT_PICST	2.60E-07
186	comp166348_g1_i2	hypothetical protein DICPUDRAFT_99500	1.40E-08
187	comp172639_g1_i4	SAMC1_ARATH	1.10E-10
188	comp181159_g3_i2	TP4AA_DICDI	9.70E-41
189	comp188328_g2_i1	MSS51 homolog, mitochondrial	1.20E-10
190	comp188583_g1_i2	CY43_TRYBB	1.60E-20
191	comp195441_g1_i1	Nuclear pore complex Nup205	1.40E-11

192	comp195456_g1_i2	kinase domain	1.00E-27
193	comp187012_g1_i12	peptidase M50	2.60E-37
194	comp188033_g2_i14	SAMC2_ARATH	1.60E-22
195	comp189611_g2_i2	udp-n-acetylglucosamine--peptide n-acetylglucosaminyltransferase	2.10E-17
196	comp194269_g1_i4	MSH1_ARATH	4.10E-85
197	comp184642_g2_i17	membrane-bound adenylyl cyclase	8.00E-40
198	comp186058_g1_i6	family methyltransferase	3.10E-16
199	comp195035_g1_i1	ankyrin repeat	3.40E-35
200	comp190486_g1_i3	E3 ubiquitin- ligase SIAH1-like	8.90E-13
201	comp253188_g1_i1	NU1M_ANOGA	3.20E-25
202	comp185604_g1_i2	PUMP4_ARATH	1.70E-56
203	comp175039_g2_i2	BMT5_SCHPO	1.00E-30
204	comp187537_g1_i5	Tha4 Hcf106	2.50E-17
205	comp175223_g2_i1	DCA13_XENLA	4.30E-113
206	comp181202_g1_i5	FKBP3_MOUSE	9.60E-29
207	comp193751_g1_i3	thiol:disulfide interchange	7.80E-33
208	comp178094_g1_i3	glycosyl transferase	2.50E-18
209	comp193157_g1_i3	FKB62_ARATH	5.30E-19
210	comp182012_g1_i7	type III domain and Immunoglobulin-like fold domain-containing	1.40E-07
211	comp187283_g6_i9	predicted protein, partial	1.10E-25
212	comp193197_g1_i4	guanylate cyclase	1.10E-64
213	comp187022_g1_i6	ABCA1_MOUSE	2.80E-10
214	comp183741_g1_i8	calmodulin mutant syncam9	8.00E-09
215	comp194344_g1_i3	histidine kinase	3.00E-09
216	comp194659_g1_i1	peroxidase	8.70E-93
217	comp175966_g1_i1	hypothetical protein SDRG_02977	1.00E-40
218	comp184347_g1_i7	SEPR_THESR	1.60E-31
219	comp179878_g2_i1	PI5L1_MOUSE	1.40E-09
220	comp193054_g2_i1	na+ solute symporter	1.20E-62
221	comp193142_g1_i3	PLCG1_MOUSE	7.70E-10
222	comp176686_g1_i3	proline dehydrogenase 1, mitochondrial-like	7.90E-140
223	comp187484_g3_i10	LUC7L_DICDI	3.10E-24
224	comp184717_g1_i5	PSP1_YEAST	2.40E-22
225	comp184804_g1_i1	hypothetical protein G7K_4091-t1	2.70E-08
226	comp184195_g2_i9	SPHK2_MOUSE	1.40E-09

227	comp171334_g2_i1	kinase-like domain	1.50E-11
228	comp193875_g2_i9	CCNA2_MESAU	1.10E-10
229	comp192442_g1_i3	adenylate cyclase type 10	7.00E-39
230	comp167281_g1_i2	CP18B_ARATH	5.30E-16
231	comp190862_g1_i4	predicted protein	4.40E-26
232	comp180662_g1_i2	hypothetical protein SPRG_14866	2.50E-07
233	comp193627_g2_i6	hypothetical protein SDRG_04080	3.50E-12
234	comp168933_g1_i8	ZDH1_STAAW	2.20E-19
235	comp194921_g1_i2	predicted protein	3.10E-17
236	comp195127_g1_i1	probable beta-1,3-galactosyltransferase 19	2.30E-16
237	comp195184_g1_i6	CYAA_TRYEQ	1.60E-27
238	comp164778_g1_i4	DNJ10_ARATH	1.10E-34
239	comp183741_g1_i9	calmodulin mutant syncam9	6.90E-09
240	comp192057_g1_i3	FBX11_HUMAN	6.10E-10
241	comp155993_g1_i5	predicted protein	1.40E-76
242	comp191381_g2_i7	OTU family cysteine protease	9.00E-17
243	comp195047_g1_i4	membrane-associated protein, putative	3.50E-07
244	comp185804_g1_i1	pbs lyase heat-like repeat	2.40E-10
245	comp166051_g1_i3	DLDH_SCHPO	2.00E-39
246	comp186997_g2_i1	EGY1_ARATH	2.10E-24
247	comp184118_g1_i3	hypothetical protein SAMD00019534_103110	1.60E-10
248	comp183918_g1_i2	hybrid sensor histidine kinase response regulator	1.50E-30
249	comp184225_g1_i3	cyclic nucleotide-binding domain	2.90E-19
250	comp180235_g1_i2	PTA14_ARATH	8.60E-09
251	comp183357_g2_i2	RH39_ORYSJ	6.00E-36
252	comp171802_g1_i2	Major Facilitator Superfamily (MFS)	4.70E-31
253	comp193125_g2_i6	PCYAA_EUGLO	1.60E-173
254	comp192165_g1_i1	COG7_HUMAN	4.30E-22
255	comp189140_g2_i1	NRF6_CAEEL	3.10E-28
256	comp183398_g1_i3	ALG6_DROME	6.40E-90
257	comp186653_g3_i15	SGK2_MOUSE	2.00E-34
258	comp191077_g1_i4	hypothetical protein PPTG_15242	1.40E-08
259	comp193459_g1_i7	hypothetical protein	2.80E-08
260	comp193627_g1_i10	EF-hand calcium-binding domain-containing 6	5.20E-12
261	comp191051_g1_i9	adenylate cyclase type 8 isoform X1	7.80E-15

262	comp181407_g1_i3	RAD18_ASHGO	9.60E-09
263	comp182956_g1_i3	METL4_MOUSE	4.80E-20
264	comp177583_g1_i9	N-terminal kinase	6.40E-105
265	comp194462_g1_i4	Transcription factor	1.50E-57
266	comp195435_g1_i3	SLN1_YEAST	9.20E-14
267	comp155556_g1_i5	NLRC3_HUMAN	2.60E-45
268	comp193407_g2_i4	ATC1_DICDI	1.70E-37
269	comp156404_g1_i2	YKT61_ARATH	1.10E-44
270	comp164733_g1_i3	PTPS_DROME	3.10E-29
271	comp188655_g3_i6	of CLR family	7.00E-14
272	comp188822_g1_i6	L-aspartate oxidase	2.20E-85
273	comp179970_g2_i1	Endonuclease exonuclease phosphatase	8.10E-20
274	comp179393_g1_i1	YMDB_ENTCC	1.90E-33
275	comp182908_g1_i1	unnamed product	5.30E-45
276	comp186793_g2_i1	CYA1_RHIME	4.90E-07
277	comp193300_g3_i3	PTR36_ARATH	4.80E-25
278	comp187279_g3_i6	uridine kinase	3.80E-69
279	comp188566_g1_i5	MYCF_MICGR	4.50E-34
280	comp170994_g1_i3	hybrid sensor histidine kinase response regulator	6.80E-52
281	comp182068_g1_i5	Hypothetical protein, putative	3.10E-09
282	comp182725_g1_i5	PABP5_ARATH	4.00E-15
283	comp193009_g1_i4	BCAT_DICDI	8.60E-132
284	comp190340_g2_i1	ELAV 3 isoform X8	2.30E-21
285	comp190961_g1_i6	LPAT2_BRANA	1.20E-30
286	comp168760_g1_i2	TBL32_ARATH	3.00E-16
287	comp188898_g1_i9	hypothetical protein MVEG_10442	1.30E-26
288	comp191673_g1_i1	ITPR3_BOVIN	1.60E-19
289	comp148809_g1_i2	DNA endonuclease RBBP8	7.40E-13
290	comp195975_g2_i2	HERC4_MOUSE	3.30E-14
291	comp195944_g1_i1	polycystin-2-like isoform X2	9.90E-31
292	comp179258_g1_i1	FAS2_LACKL	1.60E-180
293	comp173359_g1_i6	predicted protein	5.10E-17
294	comp185971_g1_i4	C-myc promoter-binding -like	9.70E-41
295	comp178741_g1_i1	SGNH hydrolase	1.10E-29
296	comp187615_g1_i2	PFLA_CLOPA	8.00E-32

297	comp175844_g1_i3	Fumarate reductase	0.00E+00
298	comp181497_g1_i8	ubiquinone biosynthesis methyltransferase	1.80E-14
299	comp165936_g1_i2	SMYD3_HUMAN	2.70E-10
300	comp195714_g2_i6	uracil-DNA glycosylase	3.60E-16
301	comp175449_g1_i3	NSR1_YEAST	1.80E-11
302	comp195469_g1_i9	expression site-associated gene 4 (ESAG4) ,	0.00E+00
303	comp196083_g1_i3	sensory histidine kinase	9.90E-43
304	comp195129_g1_i2	AFK_PHYPO	3.00E-15
305	comp193275_g2_i4	Y663_CHLTE	3.70E-11
306	comp185260_g1_i4	KAP0_BOVIN	3.70E-33
307	comp190899_g1_i14	Y663_CHLTE	1.20E-11
308	comp187814_g2_i4	PCAT2_MOUSE	9.50E-40
309	comp179782_g1_i3	vesicular acetylcholine transporter	1.40E-17
310	comp195445_g1_i4	DUG3_YEAST	6.80E-64
311	comp195445_g1_i2	Beta-Ala-His dipeptidase	1.10E-92
312	comp182421_g1_i6	Squamous cell carcinoma antigen recognized by T-cells 3	4.00E-38
313	comp175716_g1_i6	Tetratricopeptide-like helical	1.20E-46
314	comp192963_g1_i3	NSUN2_MOUSE	8.70E-103
315	comp187997_g1_i6	WDR35_HUMAN	7.40E-22
316	comp189792_g2_i2	TRM61_HUMAN	2.80E-49
317	comp184765_g1_i4	GUAD_BACSU	1.00E-09
318	comp155782_g1_i3	FIS1_DEBHA	2.20E-08
319	comp176295_g1_i3	transmembrane protease serine 9	2.10E-22
320	comp115089_g1_i1	KAT_DICDI	1.60E-81
321	comp150091_g1_i1	CYSG_NITEC	8.10E-94
322	comp181743_g3_i2	predicted protein	8.20E-48
323	comp192964_g1_i3	Tbc2 translation chloroplastic	2.10E-30
324	comp177607_g1_i4	kinase domain	2.90E-12
325	comp176165_g1_i4	AK1_DICDI	2.70E-18
326	comp179709_g1_i2	KAD2_BOVIN	5.70E-41
327	comp180923_g1_i1	PAP12_ARATH	5.40E-10
328	comp190216_g2_i3	predicted protein	3.10E-18
329	comp186592_g1_i5	mRNA capping enzyme	2.90E-119
330	comp191116_g1_i2	mitochondrial phosphate carrier	3.50E-38
331	comp196030_g1_i2	EXPR_XANCP	9.40E-67

332	comp184539_g1_i1	RNA binding	1.00E-89
333	comp193019_g1_i10	Extracellular ligandbinding receptor,	3.60E-07
334	comp181835_g1_i4	GDIR_ARATH	6.70E-08
335	comp186653_g2_i3	KPK2_ARATH	2.40E-34
336	comp190899_g1_i2	transcriptional regulator	5.60E-11
337	comp175716_g1_i8	Tetratricopeptide-like helical	2.40E-46
338	comp185857_g1_i1	Lipase, class 3	6.30E-48
339	comp192736_g2_i4	BARA_ECO57	3.70E-37
340	comp170275_g1_i9	Poly (A) RNA polymerase cid14	4.20E-51
341	comp190793_g1_i7	glycoside hydrolase	5.80E-93
342	comp174475_g2_i2	hypothetical protein PBRA_004479	3.10E-17
343	comp185965_g1_i1	cyclic nucleotide-binding	8.90E-135
344	comp186155_g1_i3	PREDICTED: uncharacterized protein LOC105847846	1.30E-31
345	comp93600_g1_i2	NLRC3_HUMAN	1.10E-35
346	comp187755_g1_i7	hypothetical protein CHLREDRAFT_206193	2.00E-07
347	comp172902_g1_i4	exostosin family	5.80E-61
348	comp189916_g1_i6	fructose-1,6-bisphosphatase, chloroplastic-like	5.10E-89
349	comp189308_g1_i4	Mir1p	4.60E-34
350	comp152869_g1_i4	TRX1_YEAST	5.70E-26
351	comp185526_g1_i8	MCPAL_ARTBC	6.20E-29
352	comp148135_g1_i1	RENT1_CAEEL	3.60E-25
353	comp186313_g2_i4	uracil-DNA glycosylase	1.60E-33
354	comp194861_g1_i1	hypothetical protein Ctob_007064	5.50E-39
355	comp175055_g1_i1	LCYD1_ORYSJ	3.60E-42
356	comp176377_g1_i3	FCA1_TRYRA	9.70E-20
357	comp190240_g1_i3	DEAD DEAH box helicase	2.40E-16
358	comp170544_g1_i3	SGNH hydrolase	8.90E-30
359	comp166574_g1_i3	MGDP1_HUMAN	4.10E-40
360	comp183431_g1_i3	Hypothetical protein, putative	2.80E-18
361	comp176985_g1_i3	YDJX_ECOLI	4.40E-07
362	comp185194_g1_i1	E3 ubiquitin- ligase Hakai-like	8.10E-39
363	comp187103_g2_i1	predicted protein	3.80E-20
364	comp174805_g1_i3	hypothetical protein	7.30E-07
365	comp185922_g1_i2	PPIL2_EMENI	5.70E-78
366	comp7396_g1_i1	RL39_CAEEL	3.20E-08

367	comp187834_g1_i10	RP6L3_ARATH	5.70E-09
368	comp168870_g1_i4	receptor-type adenylate cyclase a	8.70E-14
369	comp148285_g1_i1	RNZ1_HUMAN	1.30E-56
370	comp191941_g1_i3	VWA8_MOUSE	0.00E+00
371	comp195147_g2_i1	KDIS_DANRE	8.70E-22
372	comp188576_g1_i5	Zinc (Zn2)-Iron (Fe2) Permease (ZIP) Family	5.60E-35
373	comp196105_g2_i1	beta subunit of photoactivated adenylyl cyclase	7.10E-16
374	comp143492_g1_i1	hypothetical protein	4.10E-16
375	comp188981_g4_i5	ITS3_SCHPO	5.00E-54
376	comp180692_g1_i1	kinase domain	9.40E-16
377	comp181992_g2_i1	MHCKA_DICDI	7.50E-23
378	comp147265_g1_i1	beta-type IP39	7.90E-20
379	comp187922_g1_i4	hypothetical protein CAPTEDRAFT_187274	5.30E-22
380	comp177842_g1_i2	CEEH1_CAEEL	1.30E-08
381	comp186502_g2_i4	phosphatidylinositol-4-phosphate 5-kinase	4.20E-55
382	comp193127_g1_i4	T9SS C-terminal target domain-containing	1.20E-21
383	comp196152_g1_i2	PAS domain-containing sensor histidine kinase	8.20E-22
384	comp194224_g1_i1	beta-1,4-mannosyl-glyco beta-1,4-N-acetylglucosaminyltransferase	8.50E-37
385	comp181497_g1_i4	ubiquinone biosynthesis methyltransferase	3.10E-14
386	comp161932_g2_i1	hypothetical protein SELMODRAFT_54741predicted protein, partial	1.50E-20
387	comp165338_g1_i4	hypothetical protein	1.50E-30
388	comp164252_g1_i6	TRAM LAG1 CLN8 homology domain	2.10E-17
389	comp186188_g1_i10	predicted protein	2.00E-12
390	comp195426_g3_i2	Y2006_MYCTO	1.50E-105
391	comp182512_g1_i7	YPL1_DROME	1.40E-09
392	comp195184_g1_i4	receptor-type adenylate cyclase a	2.00E-80
393	comp193479_g3_i2	TRMA_AZOSB	7.90E-13
394	comp178350_g1_i4	tetratricopeptide repeat family	2.50E-12
395	comp185440_g2_i1	VTI11_ARATH	2.70E-17
396	comp191752_g2_i2	histidine kinase	5.30E-12
397	comp189630_g1_i2	probable phosphatase 2C 59 isoform X1	1.90E-43
398	comp178833_g1_i3	Tetratricopeptide repeat domain	3.90E-19
399	comp179825_g1_i2	MUTYH_MOUSE	7.50E-77
400	comp191381_g2_i2	OTU family cysteine protease	2.50E-15
401	comp185323_g2_i1	PLOD3_RAT	3.60E-08

402	comp190889_g2_i3	histidine kinase	3.80E-58
403	comp181588_g1_i3	DNJC7_DICDI	2.50E-60
404	comp183830_g1_i4	hypothetical protein THAOC_27631	5.90E-18
405	comp188625_g1_i4	NLRC3_MOUSE	1.90E-37
406	comp188621_g1_i10	copper transporter family	2.00E-13
407	comp193684_g1_i6	Leucine-zipper-like transcriptional regulator 1	1.40E-14
408	comp183568_g1_i3	TAGD_DICDI	1.50E-30
409	comp185788_g1_i16	GDSL-like Lipase Acylhydrolase	1.30E-22
410	comp188116_g1_i2	PREDICTED: uncharacterized protein LOC756716	5.80E-28
411	comp190804_g1_i3	Transcription factor	1.00E-57
412	comp185023_g6_i5	transporter, major facilitator family	1.40E-81
413	comp193325_g1_i7	DEAD DEAH box helicase	1.00E-13
414	comp140387_g1_i2	WD40 repeat-containing	4.90E-08
415	comp184895_g2_i10	COBT_PSEF5	2.70E-30
416	comp192811_g1_i1	acyltransferase, WS DGAT MGAT	8.20E-43
417	comp193729_g1_i7	dicer-like protein	1.40E-09
418	comp195504_g5_i5	leucine-rich repeat and guanylate kinase domain-containing isoform X1	4.20E-18
419	comp191957_g1_i5	CY42_TRYBB	5.20E-16
420	comp183592_g1_i5	LON peptidase N-terminal domain and RING finger 2predicted protein, partial	3.70E-46
421	comp182313_g2_i5	glycosyltransferase aer61	6.10E-42
422	comp195470_g1_i2	NEK4_MOUSE	3.60E-39
423	comp193533_g1_i2	S phase cyclin A-associated in the endoplasmic reticulum-like	6.40E-22
424	comp187961_g3_i12	ELAV4_RAT	7.70E-21
425	comp196069_g1_i6	hypothetical protein GUITHDRAFT_142802	1.30E-17
426	comp176950_g1_i1	IPYR_ZYGBA	2.90E-62
427	comp186828_g1_i4	predicted protein	2.40E-42
428	comp162429_g1_i4	NIMA-related kinase	6.80E-10
429	comp192253_g2_i1	leucine-rich repeat-containing C10orf11 homolog	2.70E-19
430	comp195344_g1_i5	e3 ubiquitin- ligase	9.30E-24
431	comp182973_g1_i5	predicted protein, partial	2.80E-43
432	comp167037_g1_i3	RMGL_PICST	8.10E-22
433	comp190606_g2_i8	PARG_BOVIN	8.40E-55
434	comp187182_g1_i5	ABCGM_DICDI	2.00E-08
435	comp194148_g1_i3	CID4_ARATH	2.70E-08
436	comp185293_g1_i2	metallophosphoesterase domain-containing 1	4.10E-61

437	comp178106_g1_i2	Palmitoyltransferase ZDHHC2predicted protein, partial	1.50E-43
438	comp182956_g1_i2	methyltransferase 4	6.30E-21
439	comp183911_g1_i1	mitochondrial carrier family	2.70E-72
440	comp163029_g1_i11	CIMA_GEOSL	3.30E-149
441	comp188052_g2_i8	predicted protein, partial	1.40E-23
442	comp162177_g1_i4	LPAT4_ARATH	1.90E-39
443	comp183955_g1_i1	RFWD2_MOUSE	1.50E-07
444	comp161211_g1_i2	MCE_FOWPN	9.30E-13
445	comp185634_g3_i1	TAGC_DICDI	1.90E-22
446	comp189150_g2_i3	IMCL1_ORYSJ	5.20E-58
447	comp174994_g3_i1	PUF68_DROME	4.10E-14
448	comp195514_g1_i1	OTU domain-containing 7B isoform X3	4.50E-24
449	comp187447_g2_i2	CYAA_STIAU	4.80E-40
450	comp194978_g1_i3	PUB35_ARATH	2.90E-10
451	comp176872_g1_i1	SAC1_CHLRE	4.00E-17
452	comp186300_g2_i5	ELAV4_RAT	1.50E-16
453	comp194535_g1_i5	chromosome partitioning	2.60E-10
454	comp186645_g1_i1	HCP_GEOMG	1.90E-139
455	comp175886_g1_i3	Ig domain-containing	1.10E-14
456	comp163029_g1_i4	CIMA_GEOSL	9.90E-149
457	comp194662_g1_i6	von Willebrand factor type A domain containing	2.70E-47
458	comp183035_g2_i7	hypothetical protein	2.10E-10
459	comp192856_g2_i2	transcriptional coactivator yorkie	1.10E-10
460	comp193216_g3_i1	hypothetical protein GUITHDRAFT_162439	6.60E-11
461	comp178442_g1_i4	phytoene synthase	1.90E-109
462	comp194245_g1_i2	conserved unknown protein	1.10E-16
463	comp186453_g1_i1	DAP4_PSEMX	1.50E-122
464	comp158793_g2_i2	hypothetical protein CAOG_03935	1.40E-15
465	comp185640_g1_i7	ADP-ribosyltransferase 1 precursor	1.60E-36
466	comp187994_g1_i5	HCNC_PSEAE	2.20E-19
467	comp190238_g2_i2	kinase domain	6.00E-07
468	comp195270_g5_i2	signal transduction histidine kinase	2.50E-16
469	comp192028_g1_i9	unnamed protein product	7.40E-24
470	comp171295_g1_i2	NLRC3_HUMAN	1.20E-15
471	comp191799_g1_i6	CBS domain containing	1.20E-10

472	comp180583_g1_i1	VPS45_RAT	4.50E-152
473	comp183974_g2_i3	kinase domain	2.00E-10
474	comp186039_g1_i3	OOPDA_ARATH	4.10E-128
475	comp182446_g1_i5	MPPD1_HUMAN	4.50E-55
476	comp193892_g1_i3	RBM45_MOUSE	1.80E-14
477	comp183035_g2_i2	hypothetical protein	2.00E-10
478	comp195357_g1_i1	parp domain-containing	3.40E-11
479	comp164102_g1_i2	transmembrane 42	4.00E-09
480	comp178350_g2_i3	tetratricopeptide repeat family	2.70E-12
481	comp190975_g2_i6	predicted protein, partial	1.60E-90
482	comp194648_g1_i4	mitochondrial fission process 1	1.50E-11
483	comp193413_g1_i3	YJJW_ECOLI	1.30E-131
484	comp161345_g1_i1	PCS3_LOTJA	1.50E-67
485	comp178799_g1_i2	NADB_PSEAE	9.70E-75
486	comp195636_g1_i1	TRPV5_HUMAN	6.00E-25
487	comp159340_g1_i2	hydroxyacylglutathione hydrolase	1.80E-52
488	comp184826_g1_i3	predicted protein	2.10E-18
489	comp195850_g1_i1	Proteophosphoglycan ppg4	2.70E-36
490	comp188065_g1_i4	SLBP2_XENLA	1.10E-08
491	comp127935_g1_i2	PSBA_EUGGR	0.00E+00
492	comp185125_g1_i4	PAS domain-containing sensor histidine kinase	2.10E-14
493	comp174697_g1_i9	amino acid adenylation	2.50E-15
494	comp195673_g1_i1	ADP-ribosyltransferase 1 precursor	8.80E-11
495	comp185389_g4_i8	CY43_TRYBB	2.10E-14
496	comp191813_g2_i2	CLPC1_ARATH	0.00E+00
497	comp192682_g3_i6	glycosyltransferase aer61	1.70E-43
498	comp185483_g1_i11	CRNS1_CHICK	3.10E-18
499	comp161697_g1_i1	APM1_ARATH	1.40E-120
500	comp190114_g2_i5	EXPR_XANCP	2.10E-72
501	comp175153_g1_i5	SPNS3_ARATH	2.60E-10
502	comp194576_g1_i8	myosin heavy chain kinase	5.60E-52
503	comp165149_g1_i4	TM222_HUMAN	6.10E-30
504	comp189758_g2_i9	AARA_DICDI	1.00E-10
505	comp181717_g1_i1	glycosyl transferase	1.00E-40
506	comp195299_g1_i3	stage II sporulation E	2.60E-10

507	comp193820_g1_i6	FPRA1_CLOAB	8.50E-79
508	comp171334_g1_i2	kinase-like domain	1.40E-11
509	comp182081_g1_i1	type II inositol 1,4,5-trisphosphate 5-phosphatase	4.40E-53
510	comp183538_g1_i1	hypothetical protein GUITHDRAFT_164479	2.00E-30
511	comp171999_g2_i1	receptor-type adenylate	9.20E-10
512	comp185788_g1_i8	GDSL-like Lipase Acylhydrolase	1.50E-22
513	comp179808_g1_i3	UGGG_DROME	1.40E-167
514	comp192784_g1_i3	sporulation	6.80E-09
515	comp180295_g1_i2	manganese transporter	4.00E-145
516	comp192804_g4_i1	ARM repeat-containing	3.70E-65
517	comp172016_g4_i11	DPOE2_CHICK	9.50E-82
518	comp193197_g1_i1	PDE10_HUMAN	1.60E-22
519	comp179132_g1_i1	VATL_AEDAE	1.50E-22
520	comp174547_g1_i1	glyoxalase-like domain	4.00E-68
521	comp122456_g2_i2	PREDICTED: uncharacterized protein LOC18444466	1.20E-07
522	comp196079_g5_i3	endonuclease V	1.90E-25
523	comp184170_g1_i3	mannosyl-oligosaccharide 1,2-alpha-mannosidase	3.20E-128
524	comp178580_g1_i1	GELA_DICDI	1.50E-08
525	comp190688_g1_i5	Aspartic peptidase	4.10E-21
526	comp195471_g1_i5	TSGAD_HALVD	2.90E-29
527	comp181438_g2_i1	Hypothetical protein, putative	6.30E-64
528	comp189557_g1_i5	CYAA_PODAS	1.40E-25
529	comp195270_g5_i14	signal transduction histidine kinase	2.20E-16
530	comp160438_g1_i2	AZOR_AROAE	2.40E-23
531	comp144297_g1_i1	ILV5_SCHPO	4.60E-99
532	comp170245_g1_i3	NOVP_STRNV	1.50E-42
533	comp185293_g1_i3	metallophosphoesterase domain-containing 1	4.20E-61
534	comp165748_g1_i3	Endonuclease exonuclease phosphatase	8.30E-30
535	comp189519_g1_i3	YFZ3_SCHPO	4.80E-12
536	comp190120_g1_i5	TEA1_SCHPO	1.90E-20
537	comp161280_g1_i1	hypothetical protein F443_00243	3.00E-11
538	comp182887_g2_i4	hypothetical protein ABB37_09572	2.00E-29
539	comp191298_g4_i4	S38AB_BOVIN	2.30E-20
540	comp183073_g2_i3	lipase YOR059C isoform X2	1.00E-57
541	comp181985_g1_i1	TIGR02452 family	5.20E-52

542	comp185604_g1_i7	PUMP4_ARATH	1.10E-56
543	comp193633_g1_i1	hypothetical protein, variant	6.30E-45
544	comp190712_g1_i6	hypothetical protein COCSUDRAFT_31865	1.50E-19
545	comp173925_g1_i7	STX16_HUMAN	4.60E-13
546	comp180333_g1_i1	transporter, major facilitator subfamily	8.40E-36
547	comp195957_g1_i4	INT2_DICDI	4.90E-09
548	comp192775_g2_i2	coproporphyrinogen III oxidase	2.50E-64
549	comp188913_g2_i1	WDHD1_XENLA	2.70E-12
550	comp187624_g1_i4	hypothetical protein COCSUDRAFT_45614	4.10E-14
551	comp182658_g1_i2	NLRC3_HUMAN	1.10E-16
552	comp164778_g1_i2	chaperone dnaJ 10-like isoform X1	6.70E-56
553	comp163285_g1_i1	MTCA2_MYCTO	3.60E-37
554	comp164262_g1_i1	TLC1_CHLTR	6.80E-11
555	comp177924_g1_i1	cdpalcohol phosphatidyltransferase	1.20E-108
556	comp193694_g1_i7	cold-shock protein	2.00E-07
557	comp177383_g1_i2	glycosyl transferase family 2	9.90E-29
558	comp184225_g1_i4	kinase domain	1.10E-18

Appendix table 19 genes name of area 19

(a) List of TF family (56 transcripts of 22 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp175780_g1_i2	B3	1.70E-09
2	comp175780_g1_i4	B3	1.70E-09
3	comp181747_g2_i2	BBR-BPC	1.90E-21
4	comp188354_g1_i4	C3H	3.30E-08
5	comp188354_g1_i7	C3H	3.50E-08
6	comp178174_g1_i1	MIKC	3.00E-35
7	comp193263_g1_i6	MYB_related	6.30E-58
8	comp176398_g1_i5	NAC	7.00E-34
9	comp174424_g2_i3	NAC	5.30E-08
10	comp178548_g1_i3	NAC	2.30E-22
11	comp177560_g1_i4	YABBY	2.30E-35
12	comp180626_g1_i14	MYB	1.90E-135
13	comp192254_g1_i2	FAR1	2.90E-45
14	comp180981_g1_i2	STAT	4.70E-15
15	comp174861_g1_i6	B3	4.10E-52
16	comp162969_g1_i7	GeBP	1.50E-25
17	comp177778_g1_i4	bHLH	8.30E-50
18	comp182592_g1_i5	bZIP	4.30E-09
19	comp185545_g1_i2	bZIP	2.80E-11
20	comp175007_g1_i1	C2H2	7.30E-08
21	comp188529_g2_i1	C3H	1.10E-07
22	comp169524_g2_i3	EIL	3.50E-56
23	comp184829_g1_i3	EIL	4.20E-21
24	comp193468_g1_i5	FAR1	3.90E-44
25	comp170831_g3_i2	FAR1	3.90E-34
26	comp170244_g1_i3	FAR1	5.20E-19
27	comp166760_g1_i1	GRAS	1.60E-13
28	comp190700_g2_i3	GeBP	2.50E-09
29	comp185219_g1_i2	LBD	8.90E-10
30	comp187750_g1_i4	M-type	4.50E-23
31	comp178949_g1_i3	MYB	3.10E-45
32	comp180626_g1_i12	MYB	4.70E-136

33	comp158180_g1_i6	MYB_related	2.60E-59
34	comp178168_g1_i2	NAC	3.90E-77
35	comp187312_g1_i1	Nin-like	1.60E-79
36	comp187312_g1_i5	Nin-like	3.70E-78
37	comp175353_g1_i13	WOX	8.30E-33
38	comp178399_g2_i3	WRKY	3.70E-28
39	comp178399_g2_i4	WRKY	2.30E-28
40	comp193377_g1_i1	WRKY	1.30E-46
41	comp178274_g1_i3	WRKY	5.70E-13
42	comp177560_g1_i2	YABBY	2.50E-35
43	comp176811_g1_i3	bHLH	1.60E-41
44	comp188354_g1_i6	C3H	3.20E-08
45	comp188939_g1_i1	C3H	7.70E-07
46	comp188939_g1_i2	C3H	5.80E-13
47	comp188939_g1_i5	C3H	5.30E-13
48	comp168941_g1_i2	ERF	8.50E-08
49	comp181802_g1_i5	LBD	1.70E-44
50	comp178454_g1_i2	NAC	1.40E-14
51	comp176175_g2_i1	YABBY	2.70E-42
52	comp177750_g1_i1	STAT	1.70E-19
53	comp193820_g1_i2	WRKY	1.10E-07
54	comp178399_g2_i1	WRKY	1.70E-28
55	comp177991_g2_i3	bHLH	1.40E-41
56	comp176811_g1_i2	bHLH	1.50E-41

(b) List of other genes name (209 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp190703_g1_i5	hypothetical protein SAMD00019534_055790	2.20E-16
2	comp182284_g1_i1	PCT2B_MOUSE	5.90E-43
3	comp190869_g1_i3	RNF12_MOUSE	3.80E-08
4	comp183101_g1_i7	ALG8_ASPOR	1.20E-63
5	comp181987_g1_i5	Hypothetical protein, putative	4.80E-12
6	comp195741_g1_i1	GACS_PSESY	6.20E-27
7	comp164942_g3_i2	5 -AMP-activated kinase subunit gamma-1	8.00E-44

8	comp183568_g1_i2	TAGD_DICDI	7.80E-31
9	comp187113_g10_i2	NLRC3_MOUSE	1.60E-35
10	comp190515_g1_i1	LON peptidase N-terminal domain and RING finger 1-like	2.70E-12
11	comp99586_g1_i3	PGML1_ARATH	6.80E-09
12	comp170860_g1_i1	AOX_CRYNH	3.20E-53
13	comp182126_g1_i3	DNA repair RAD51 homolog 2	2.00E-54
14	comp193985_g1_i2	YQ77_SCHPO	2.00E-22
15	comp168871_g1_i4	YKX5_SCHPO	1.80E-11
16	comp192204_g1_i5	DEK1_ORYSJ	1.40E-41
17	comp178042_g1_i4	MMAB_MOUSE	2.40E-45
18	comp183048_g2_i1	WECE_ECOLI	6.20E-126
19	comp174505_g2_i2	Poly(ADP-ribose) polymerase catalytic domain containing	4.30E-89
20	comp189878_g1_i8	hypothetical protein GUITHDRAFT_121297	1.40E-28
21	comp187660_g1_i8	C2 domain-containing	5.90E-14
22	comp166919_g1_i5	SRPR_DROME	1.30E-118
23	comp186503_g2_i3	formin 5	4.90E-26
24	comp168979_g1_i2	ELMO domain-containing A isoform X2	7.50E-26
25	comp140961_g2_i1	ATP synthase F0 subunit 6 (mitochondrion)	1.80E-12
26	comp194643_g1_i2	calpain-like cysteine peptidase	1.60E-08
27	comp178587_g3_i1	hypothetical protein Ctob_008097	9.10E-15
28	comp178514_g1_i3	GRDP1_ARATH	5.00E-30
29	comp171450_g1_i6	YCX1_EUGGR	7.30E-155
30	comp192221_g2_i2	ABC transporter, ATP-binding predicted protein, partial	2.70E-54
31	comp188609_g4_i6	macrocin O-methyltransferase	7.30E-50
32	comp175453_g1_i4	ubiquitin-ribosomal 60S subunit L40B fusion	7.60E-23
33	comp181360_g1_i6	CAMK CAMKL AMPK kinase	2.50E-132
34	comp187595_g1_i1	extended synaptotagmin-1	5.90E-10
35	comp195101_g1_i2	family 3 adenylate cyclase	4.60E-46
36	comp168478_g1_i2	CYSK_SYNE7	2.20E-124
37	comp170571_g1_i2	TI17A_DROME	4.10E-30
38	comp187988_g1_i10	expressed unknown protein	8.40E-46
39	comp186814_g1_i6	ABCD4_MOUSE	4.30E-67
40	comp196150_g1_i4	CYA1_RHIME	4.40E-37
41	comp180797_g1_i2	PLBL1_RAT	9.40E-98
42	comp187503_g1_i3	EDRF1_MOUSE	2.40E-32

43	comp189241_g1_i1	CYAA_NEUCR	5.30E-09
44	comp192944_g1_i2	unknown	1.10E-50
45	comp189197_g1_i4	kinase domain	1.30E-20
46	comp187790_g1_i11	CAD -like	0.00E+00
47	comp180998_g3_i3	Cysteine dioxygenase type 1	5.20E-41
48	comp185767_g1_i12	outer dense fiber 3B	3.40E-16
49	comp158590_g1_i2	LENG8_XENLA	3.40E-43
50	comp159355_g1_i6	ammonium transporter Rh type A	2.10E-92
51	comp172013_g1_i3	nucleotidyltransferase DNA polymerase	5.30E-22
52	comp181709_g1_i4	UBP49_MOUSE	1.60E-40
53	comp196175_g1_i9	HAMP domain	1.20E-12
54	comp185046_g1_i1	SULT2_CHLRE	8.50E-38
55	comp175614_g1_i3	zinc induced facilitator	8.30E-47
56	comp192461_g1_i2	TRPA1_HUMAN	2.40E-08
57	comp188052_g2_i7	unnamed protein product	1.40E-23
58	comp117008_g1_i1	TRAF6_HUMAN	2.60E-15
59	comp192022_g1_i2	KCNKI_MOUSE	1.50E-11
60	comp185771_g1_i5	kinase domain	9.60E-34
61	comp188155_g5_i2	NLRC3_HUMAN	9.70E-16
62	comp192602_g1_i1	receptor-type adenylate cyclase,	2.00E-51
63	comp187520_g3_i3	membrane-bound adenylyl cyclase	3.50E-78
64	comp157136_g1_i1	transmembrane 128-like	3.00E-08
65	comp181605_g1_i1	Hypothetical protein, putative	5.10E-42
66	comp171053_g1_i1	hypothetical protein	1.70E-07
67	comp189653_g2_i1	TEB_ARATH	1.20E-175
68	comp178789_g6_i2	KAP1_MOUSE	9.50E-13
69	comp188125_g1_i2	CYA1_RHIME	6.30E-11
70	comp174392_g2_i1	DUF4442 domain-containing	7.30E-23
71	comp174743_g1_i2	OTU7B_HUMAN	2.80E-32
72	comp181650_g1_i11	NLRC3_HUMAN	4.80E-20
73	comp187168_g1_i6	cyclic nucleotide-binding	7.10E-134
74	comp186562_g1_i1	Plastid lipid-associated fibrillin conserved domain	3.50E-32
75	comp177204_g1_i3	TSNAX_SCHPO	3.10E-08
76	comp187411_g2_i3	LOW PSII ACCUMULATION 3, chloroplastic	6.60E-41
77	comp184507_g4_i2	CBPC4_HUMAN	4.40E-100

78	comp191645_g2_i5	MFS transporter	2.50E-11
79	comp165323_g2_i1	PSBA_EUGGR	1.20E-31
80	comp189768_g2_i3	RIBC1_HUMAN	4.80E-15
81	comp192403_g1_i3	PBL9F_EUGGR	2.60E-13
82	comp185403_g4_i3	UPF0554 C2orf43 homolog isoform X1	4.20E-33
83	comp185343_g2_i4	major facilitator superfamily (MFS)	2.30E-47
84	comp180616_g1_i7	THI4_SCHPO	1.70E-72
85	comp191941_g2_i3	von Willebrand factor A domain-containing 8	0.00E+00
86	comp186550_g1_i1	exonuclease 3 -5 domain-containing 1-like	1.00E-39
87	comp170589_g1_i2	AXEA1_PRER2	1.10E-18
88	comp177365_g1_i1	UNG_ECOLI	4.30E-76
89	comp169362_g1_i3	YG75_SCHPO	1.50E-37
90	comp153060_g1_i2	#NAME?	2.00E-125
91	comp172315_g1_i7	kinase domain	7.20E-14
92	comp183058_g1_i4	solute carrier family 25 member 34	3.10E-79
93	comp189081_g1_i2	DBP2_ASHGO	9.10E-89
94	comp164252_g1_i5	TRAM LAG1 CLN8 homology domain	2.00E-17
95	comp181182_g2_i2	1A1C_MALDO	3.70E-67
96	comp169714_g1_i3	hypothetical protein GUITHDRAFT_133706	1.60E-36
97	comp173359_g1_i10	predicted protein	1.90E-17
98	comp158867_g2_i1	F26_ARATH	3.60E-66
99	comp194716_g1_i2	xylanase	3.30E-130
100	comp176712_g1_i6	SPNS1_ARATH	5.20E-17
101	comp181541_g2_i1	TWF_DICDI	5.40E-23
102	comp182924_g1_i3	arylsulfate sulfotransferase	1.20E-08
103	comp189402_g2_i3	K0556_XENLARecName: Full=Uncharacterized protein KIAA0556 homolog	3.00E-77
104	comp174787_g1_i2	Y2800_NOSS1	2.20E-14
105	comp186058_g2_i6	family methyltransferase	2.90E-16
106	comp188525_g1_i5	PI5K2_ARATH	2.30E-11
107	comp183357_g3_i1	RH39_ORYSJ	6.00E-36
108	comp181707_g1_i5	Auxin-responsive IAA20, related	5.70E-13
109	comp182283_g1_i1	alpha beta hydrolase	6.60E-67
110	comp187984_g1_i6	GLYT1_ARATH	7.50E-15
111	comp170607_g1_i1	predicted protein	2.00E-19
112	comp187555_g3_i5	AK1_DICDI	2.10E-34

113	comp181770_g1_i2	VASH2_HUMAN	3.40E-31
114	comp188361_g1_i1	EXT1_DROME	6.90E-33
115	comp171559_g1_i4	Hypothetical protein, putative	4.40E-20
116	comp193627_g2_i9	EF-hand calcium-binding domain-containing 6	6.80E-12
117	comp189048_g1_i1	GACP3_ARATH	5.60E-62
118	comp171818_g1_i2	LEO1_DANRE	1.20E-10
119	comp180227_g2_i3	elongation factor 3 ABCF transporter family	0.00E+00
120	comp174697_g3_i1	amino acid adenylation	1.10E-20
121	comp185679_g1_i7	nicotinate-nucleotide adenylyltransferase	6.50E-52
122	comp158810_g1_i2	hydroxyacylglutathione hydrolase	2.70E-52
123	comp169592_g1_i2	hypothetical protein EUTSA_v10004293mg	4.60E-18
124	comp196206_g3_i1	CY42_TRYBB	3.80E-46
125	comp177753_g1_i4	XRP2_XENLA	2.10E-51
126	comp192383_g5_i1	peptidase S15	2.10E-70
127	comp159559_g1_i2	vacuolar sorting-associated 22 homolog 1	2.60E-52
128	comp189383_g1_i2	PCYOX_ARATH	8.10E-57
129	comp184600_g3_i3	S-adenosylmethionine decarboxylase	2.00E-78
130	comp193894_g1_i2	DYH1A_CHLRE	0.00E+00
131	comp186127_g1_i6	casein kinase	1.90E-57
132	comp190768_g1_i1	DNA ligase 1	0.00E+00
133	comp189607_g1_i2	hypothetical protein AURANDRAFT_68787predicted protein, partial	9.00E-15
134	comp192684_g2_i2	PLDB1_ARATH	8.00E-139
135	comp190512_g1_i4	tetratricopeptide repeat domain containing	8.30E-127
136	comp185909_g3_i2	cellular retinaldehyde-binding triple function domain-containing	1.40E-26
137	comp192307_g1_i9	predicted protein	6.70E-19
138	comp175734_g1_i3	NLRC3_HUMAN	3.00E-23
139	comp194069_g1_i1	PUB35_ARATH	2.90E-10
140	comp190793_g1_i1	glycoside hydrolase	2.00E-86
141	comp174958_g1_i9	CPI1_PIG	1.20E-19
142	comp194969_g1_i3	AT133_HUMAN	4.30E-32
143	comp184084_g3_i5	kinase domain	1.40E-12
144	comp191923_g1_i4	STK11_CHICK	3.70E-53
145	comp184138_g1_i3	progesterone-induced-blocking factor 1	1.10E-20
146	comp190383_g2_i7	kinase domain	3.10E-08
147	comp175716_g1_i4	Tetratricopeptide-like helical	3.60E-46

148	comp171697_g4_i2	predicted protein, partial	2.30E-36
149	comp188956_g1_i3	vacuolar sorting vps16	2.30E-137
150	comp188033_g1_i4	SAMC2_ARATH	3.40E-22
151	comp185375_g2_i3	IF4E2_MOUSE	1.40E-18
152	comp163278_g1_i2	PLSP1_ARATH	1.50E-64
153	comp192453_g1_i1	MMP10_RAT	2.00E-16
154	comp167115_g1_i2	hypothetical protein CAOG_06172	2.10E-23
155	comp183296_g1_i3	hypothetical protein Ctob_007573predicted protein, partial	2.90E-27
156	comp195867_g2_i27	TCCD-inducible-PARP-like domain-containing	2.30E-10
157	comp185375_g1_i5	Eukaryotic translation initiation factor 4E-3	8.80E-20
158	comp188049_g1_i1	PHSA_ARATH	7.20E-41
159	comp186005_g1_i3	alpha beta-hydrolase	8.80E-18
160	comp165338_g1_i5	hypothetical protein	3.80E-30
161	comp175189_g1_i2	exostosin	9.90E-32
162	comp161719_g1_i2	CYC6_EUGVI	3.40E-32
163	comp173020_g1_i5	dystrophia myotonica WD repeat-containing -like	3.80E-47
164	comp171295_g1_i3	NLRC3_HUMAN	1.20E-15
165	comp192633_g1_i1	WDFY3_HUMAN	3.40E-105
166	comp180109_g2_i2	LITD_LATTR	7.10E-16
167	comp171450_g1_i13	YCX1_EUGGR	7.40E-158
168	comp182118_g1_i11	AARA_DICDI	3.00E-17
169	comp126696_g1_i2	hypothetical protein T9A_01030	3.70E-47
170	comp195334_g2_i3	glutathione-s-transferase (ISS)	2.10E-31
171	comp190618_g1_i2	pectin acetylesterase 5-like	1.00E-25
172	comp179862_g1_i2	PREDICTED: uncharacterized protein LOC106175552	3.20E-47
173	comp177970_g1_i1	L-allo-threonine aldolase	3.90E-99
174	comp182658_g1_i5	NLRC3_HUMAN	1.20E-16
175	comp166556_g1_i2	Os03g0719500predicted protein, partial	2.20E-158
176	comp181401_g1_i4	RPGR_CANFA	3.20E-17
177	comp191875_g1_i1	receptor-type adenylate cyclase GRESAG 4,	7.40E-63
178	comp190343_g1_i4	hypothetical protein EMIHUDRAFT_422811	1.50E-08
179	comp191136_g1_i4	hypothetical protein ACA1_171730	2.40E-23
180	comp165515_g1_i1	MPK3_ARATH	7.20E-135
181	comp185161_g1_i2	UBP1_NICPL	9.70E-13
182	comp189471_g3_i4	kelch diablo	1.80E-16

183	comp180616_g1_i11	THI4_SCHPO	7.00E-73
184	comp188522_g1_i7	kinase domain	1.80E-15
185	comp180579_g1_i3	D-alanine--D-alanine ligase	1.30E-132
186	comp192233_g1_i7	class II glutamine amidotransferase	3.60E-95
187	comp194330_g1_i5	TRPV6_HUMAN	1.30E-33
188	comp161699_g1_i1	CFA58_CHLRE	1.80E-46
189	comp189853_g3_i2	tuberin isoform X3	1.20E-13
190	comp190114_g2_i2	EXPR_XANCP	2.00E-72
191	comp193382_g1_i1	MKKA_DICDI	7.20E-38
192	comp166048_g1_i1	ERV1_YEAST	3.50E-08
193	comp191861_g1_i7	TUB1_CAEBR	2.30E-13
194	comp178408_g1_i3	CASTO_ORYSJ	5.00E-86
195	comp196139_g1_i12	cyclic nucleotide-binding	1.20E-26
196	comp190839_g2_i6	hypothetical protein	2.80E-27
197	comp192121_g1_i3	flagellar associated	1.10E-13
198	comp177181_g1_i2	hypothetical protein COCSUDRAFT_66769	2.40E-07
199	comp192932_g1_i5	HECD1_HUMAN	3.10E-09
200	comp188052_g2_i10	predicted protein, partial	1.70E-21
201	comp185679_g1_i8	nicotinate-nucleotide adenylyltransferase	4.70E-52
202	comp166647_g1_i3	dual specificity phosphatase 7	1.10E-46
203	comp168220_g1_i1	PTHB1_HUMAN	0.00E+00
204	comp154886_g1_i4	#NAME?	1.20E-125
205	comp193688_g2_i7	XPG I-region	3.30E-68
206	comp163952_g1_i4	DOHH_DROME	1.00E-71
207	comp190746_g1_i2	ZUPT_CHLP8	1.30E-37
208	comp174242_g1_i4	Transcription factor	2.50E-56
209	comp165063_g1_i4	TEA1_SCHPO	1.80E-20

Appendix table 20 genes name of area 20

(a) List of TF family (2 transcripts of 2 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp109547_g1_i1	C3H	4.80E-09
2	comp165138_g1_i1	Nin-like	9.40E-13

(b) List of other genes name (58 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp171450_g1_i10	YCX2_EUGGR	0.00E+00
2	comp184418_g1_i6	Proteophosphoglycan ppg4	5.60E-35
3	comp191376_g1_i1	EMARD_RAT	5.60E-16
4	comp191293_g1_i3	adenylyl cyclase	2.40E-44
5	comp122705_g1_i1	YL477_MIMIV	1.10E-31
6	comp185560_g1_i1	eukaryotic translation initiation factor 2-alpha kinase 4	1.30E-20
7	comp147074_g1_i2	hypothetical protein VOLCADRAFT_100878	6.40E-60
8	comp180445_g4_i2	oxidoreductase	1.70E-99
9	comp165022_g1_i1	PREDICTED: uncharacterized protein LOC105646130	5.80E-19
10	comp170779_g1_i2	UGAL1_ARATH	9.10E-10
11	comp194716_g1_i6	xylanase	1.00E-127
12	comp195147_g2_i2	ankyrin repeat	1.00E-32
13	comp193667_g1_i4	CRYD_GLOVI	1.10E-97
14	comp185096_g9_i3	CAPTC_DICDI	1.70E-28
15	comp175309_g1_i5	Neurotrypsin	3.10E-07
16	comp192188_g2_i1	membrane-associated protein, putative	6.70E-07
17	comp185479_g2_i1	predicted protein	1.30E-32
18	comp163128_g1_i1	SCRN3_BOVIN	1.90E-87
19	comp183097_g1_i3	IDH_AZOVI	5.30E-39
20	comp186242_g1_i5	hypothetical protein L916_17225	2.80E-09
21	comp166556_g1_i4	Os03g0719500predicted protein, partial	4.70E-159
22	comp183693_g1_i2	lipase family	4.20E-65
23	comp193410_g1_i5	MSH1_ARATH	1.50E-85
24	comp173241_g1_i3	conserved hypothetical protein	3.60E-07
25	comp172015_g1_i7	WD repeat-containing 43-like	1.40E-08

26	comp173458_g1_i2	IF2A_SCHPO	6.90E-89
27	comp184402_g1_i2	methyltransferase fkbm family	2.30E-26
28	comp186058_g1_i5	family methyltransferase	1.10E-16
29	comp181191_g1_i1	PTH_SULIA	1.70E-16
30	comp189598_g2_i4	Hypothetical protein, putative	6.60E-15
31	comp188825_g1_i4	receptor-type adenylate cyclase,	1.00E-53
32	comp192905_g2_i7	fructose-bisphosphate aldolase-lysine N-methyltransferase, chloroplastic	4.90E-24
33	comp193577_g1_i2	Phosphate transporter PHO1	2.10E-60
34	comp175354_g1_i4	ARMC2_BOVIN	8.40E-35
35	comp194744_g1_i3	hypothetical protein SAMD00019534_018240predicted protein, partial	5.40E-08
36	comp178860_g1_i2	PARG_DROME	4.80E-47
37	comp170798_g1_i3	UN104_CAEEL	1.50E-10
38	comp164653_g1_i4	Hypothetical protein, putative	7.70E-12
39	comp185578_g4_i3	predicted protein	7.80E-09
40	comp189108_g1_i7	endoplasmic reticulum membrane-associated RNA degradation isoform X1	3.10E-18
41	comp188131_g1_i3	TIC32_PEA	1.10E-52
42	comp182802_g1_i2	KLHL5_HUMAN	2.60E-16
43	comp184541_g1_i5	pseudouridine synthase	1.70E-09
44	comp193688_g2_i8	MKT1_YEAST	2.90E-11
45	comp166049_g1_i2	50S ribosomal L21	1.20E-75
46	comp165023_g2_i1	FDHC_METTF	1.40E-19
47	comp185191_g1_i5	UPF0418 C6orf94	3.30E-12
48	comp193446_g1_i3	PLA16_ARATH	4.30E-14
49	comp190925_g1_i9	predicted protein	8.90E-24
50	comp176029_g1_i1	Tok1p	1.30E-08
51	comp196175_g1_i3	HAMP domain	1.20E-11
52	comp194186_g1_i2	vitamin K-dependent gamma-carboxylase	1.10E-16
53	comp186187_g1_i12	CHO2_ASPFU	3.40E-101
54	comp142531_g2_i1	4-isomerase	3.90E-45
55	comp182908_g1_i2	unnamed product	3.00E-45
56	comp189140_g2_i2	NRF6_CAEEL	5.90E-30
57	comp187161_g1_i4	hypothetical protein GUITHDRAFT_105708	3.30E-10
58	comp188214_g1_i5	S38A5_MOUSE	6.20E-18

Appendix table 21 genes name of area 21

(a) List of TF family (74 transcripts of 26 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp170394_g1_i5	B3	1.70E-09
2	comp170394_g1_i7	B3	1.70E-09
3	comp173004_g1_i1	B3	5.20E-13
4	comp188093_g1_i7	C3H	2.00E-07
5	comp196197_g4_i3	C3H	2.10E-17
6	comp160115_g1_i2	ERF	1.00E-08
7	comp188263_g1_i4	FAR1	7.10E-41
8	comp176706_g1_i2	HB-other	1.30E-07
9	comp170533_g1_i5	MYB	4.20E-13
10	comp170882_g1_i3	NAC	5.80E-13
11	comp170933_g1_i1	NAC	5.00E-16
12	comp177411_g1_i3	WRKY	8.50E-64
13	comp192254_g1_i12	FAR1	3.00E-45
14	comp187743_g1_i2	ARR-B	9.20E-08
15	comp187743_g1_i4	ARR-B	7.30E-08
16	comp173208_g1_i3	TALE	1.80E-37
17	comp172343_g1_i2	B3	1.60E-51
18	comp174861_g1_i1	B3	4.80E-43
19	comp190700_g1_i9	GeBP	3.10E-09
20	comp176018_g1_i1	WRKY	1.40E-22
21	comp192596_g1_i3	WRKY	1.20E-46
22	comp174213_g1_i2	YABBY	1.40E-22
23	comp174213_g1_i4	YABBY	1.50E-22
24	comp181747_g1_i3	BBR-BPC	1.00E-21
25	comp168844_g1_i2	C2H2	8.60E-31
26	comp170065_g1_i1	C2H2	1.70E-22
27	comp188093_g1_i9	C3H	1.90E-07
28	comp188239_g1_i1	C3H	3.50E-08
29	comp188239_g1_i3	C3H	3.30E-08
30	comp159871_g1_i2	ERF	4.20E-12
31	comp169788_g1_i1	FAR1	4.00E-77
32	comp170244_g1_i1	FAR1	5.00E-19

33	comp194799_g1_i2	GRAS	1.40E-38
34	comp190700_g1_i6	GeBP	3.20E-09
35	comp187413_g1_i3	M-type	2.70E-23
36	comp173735_g3_i2	MIKC	2.50E-51
37	comp171889_g2_i1	MYB	4.00E-34
38	comp170533_g1_i7	MYB	4.60E-13
39	comp158180_g1_i4	MYB_related	3.80E-59
40	comp193263_g1_i3	MYB_related	5.00E-59
41	comp167886_g2_i2	NAC	3.80E-52
42	comp167986_g1_i3	NAC	1.20E-11
43	comp183686_g1_i1	Nin-like	1.80E-29
44	comp183686_g1_i4	Nin-like	1.50E-29
45	comp178291_g1_i4	Nin-like	2.60E-12
46	comp192945_g1_i1	WRKY	4.20E-42
47	comp175270_g1_i5	WRKY	2.00E-14
48	comp171446_g2_i2	WRKY	1.20E-29
49	comp172170_g1_i2	WRKY	5.30E-54
50	comp174213_g1_i3	WRKY	1.50E-22
51	comp173408_g1_i2	bHLH	1.30E-88
52	comp170315_g1_i1	bHLH	1.40E-09
53	comp180555_g1_i1	bZIP	2.50E-55
54	comp164210_g1_i1	C2H2	1.20E-07
55	comp170878_g1_i1	MYB	2.50E-15
56	comp172113_g1_i6	G2-like	8.20E-13
57	comp181802_g1_i1	LBD	1.90E-44
58	comp181802_g1_i10	LBD	1.70E-44
59	comp181033_g2_i7	M-type	6.70E-13
60	comp171889_g2_i3	MYB	2.70E-34
61	comp170882_g1_i2	NAC	5.90E-13
62	comp167447_g1_i1	NAC	1.20E-69
63	comp177212_g1_i10	Nin-like	4.90E-37
64	comp168942_g2_i1	Nin-like	2.10E-10
65	comp169138_g1_i2	Nin-like	9.50E-31
66	comp172727_g1_i1	MYB	2.00E-138
67	comp184653_g1_i3	EIL	4.30E-21

68	comp191690_g1_i4	FAR1	9.90E-46
69	comp191690_g1_i7	FAR1	7.30E-46
70	comp165387_g1_i1	SBP	6.90E-18
71	comp176176_g2_i5	STAT	3.10E-20
72	comp180555_g1_i5	bZIP	4.50E-33
73	comp190700_g1_i5	GeBP	3.30E-09
74	comp173409_g1_i2	bHLH	2.90E-29

(b) List of other genes name (376 transcripts)

No.	Seqname of Euglena	Annotation	Evalue
1	comp194749_g3_i1	CBPQ_BOVIN	6.80E-78
2	comp149648_g1_i2	phosphatase 2A regulatory subunit PR55	1.90E-143
3	comp193585_g1_i5	transformer-2 homolog alpha-like	1.50E-08
4	comp168732_g1_i4	DESI2_DANRE	2.50E-25
5	comp143128_g1_i2	hypothetical protein TRSC58_03537	7.40E-07
6	comp179681_g1_i7	cytochrome b5	1.30E-17
7	comp195818_g1_i1	BARA_ECO57	3.70E-44
8	comp191589_g2_i5	CSPA_STIAD	1.70E-09
9	comp185070_g1_i6	NAD-dependent deacetylase sirtuin-2	6.90E-89
10	comp172550_g1_i5	zinc finger FYVE domain-containing 16	1.60E-09
11	comp194465_g1_i2	hypothetical protein	1.90E-95
12	comp168488_g1_i4	NLRC3_HUMAN	6.00E-36
13	comp155556_g1_i4	NLRC3_HUMAN	1.30E-49
14	comp194325_g2_i5	HEAT6_XENLA	1.00E-19
15	comp190515_g2_i3	LONF2_HUMAN	2.00E-11
16	comp171403_g1_i2	kinase domain	1.20E-20
17	comp187210_g1_i4	NEK kinase	2.70E-48
18	comp191549_g1_i1	EF-Hand 1, calcium-binding site	0.00E+00
19	comp157089_g1_i2	ANR54_RAT	1.90E-10
20	comp184416_g3_i4	EXGA_NEOFI	6.30E-27
21	comp175031_g1_i1	NSR1_YEAST	1.80E-11
22	comp188287_g2_i2	plastidic atp adp transporter	0.00E+00
23	comp175141_g2_i2	SYS1_DICDI	1.40E-15
24	comp191293_g2_i5	adenylyl cyclase	1.70E-44

25	comp193962_g1_i3	receptor-type adenylate cyclase,	6.50E-19
26	comp195833_g1_i3	ALA3_ARATH	3.10E-110
27	comp184547_g2_i2	F-box-like WD repeat-containing TBL1X	8.70E-151
28	comp186139_g2_i1	methyltransferase family	3.20E-23
29	comp169667_g1_i1	S20A2_XENLA	3.50E-30
30	comp195033_g1_i2	WDFY3_HUMAN	6.10E-105
31	comp189636_g2_i1	IP3KA_HUMAN	2.90E-37
32	comp189386_g3_i1	PSTS_RHILO	5.70E-13
33	comp185938_g1_i4	WDR90_HUMAN	4.90E-40
34	comp165770_g1_i3	Ctr2 family transporter: copper ion CTR-type copper transporter	1.80E-26
35	comp159680_g1_i2	containing DUF1963	9.80E-28
36	comp190086_g1_i2	1-aminocyclopropane-1-carboxylate synthase	2.90E-91
37	comp194030_g3_i7	kinase domain	1.80E-12
38	comp174929_g1_i4	RFA1_HUMAN	3.40E-91
39	comp190329_g1_i16	histidine kinase	1.10E-09
40	comp163029_g1_i2	CIMA_GEOSL	4.20E-149
41	comp172972_g1_i3	hydroxyacylglutathione hydrolase	1.40E-114
42	comp194213_g1_i4	hypothetical protein ABL78_3132	6.90E-10
43	comp176157_g1_i1	Phosphatidylcholine synthase	3.80E-46
44	comp190830_g1_i5	XPG I-region	1.10E-68
45	comp171450_g4_i1	MAT1_EUGGR	3.60E-135
46	comp180019_g4_i5	ADP-ribosyltransferase 1 precursor	7.30E-31
47	comp194522_g3_i8	ferric reductase	6.00E-36
48	comp171509_g2_i1	AAA family ATPase	7.70E-10
49	comp169346_g1_i4	pap fibrillin family	9.00E-16
50	comp184327_g5_i3	GNPAT_BOVIN	9.80E-58
51	comp176272_g1_i2	MUTA_BOVIN	0.00E+00
52	comp162756_g1_i2	Trigger factor	2.40E-23
53	comp180755_g1_i6	secreted partial	5.00E-12
54	comp190343_g1_i3	hypothetical protein EMIHUDRAFT_422811	2.00E-08
55	comp185457_g1_i3	ABCA3_DICDI	1.70E-16
56	comp192399_g1_i17	Hypothetical protein, putative	4.20E-16
57	comp189137_g2_i4	ZNFX1_MOUSE	2.40E-59
58	comp188532_g2_i2	PSTS_METJA	4.20E-13
59	comp194659_g1_i4	peroxidase	8.10E-93

60	comp193424_g1_i4	transporter	2.80E-37
61	comp161732_g1_i1	GORS1_RAT	7.60E-31
62	comp187147_g2_i4	YKP9_SCHPO	7.50E-17
63	comp189758_g2_i7	AARA_DICDI	1.30E-10
64	comp178069_g2_i8	enoyl- hydratase	2.90E-80
65	comp195602_g1_i1	hybrid sensor histidine kinase response regulator	9.50E-11
66	comp190971_g3_i4	CYA1_RHIME	3.80E-09
67	comp171920_g1_i4	TMEM5_HUMAN	2.20E-12
68	comp160863_g1_i1	Y1017_SYNY3	3.00E-12
69	comp189979_g5_i3	Lipase, class 3	1.40E-49
70	comp175962_g1_i2	dual specificity phosphatase 4	9.00E-10
71	comp196310_g3_i2	expression site-associated gene 4 (ESAG4)	0.00E+00
72	comp186428_g1_i3	GRDP1_ARATH	6.00E-30
73	comp188528_g1_i2	AC010924_18 ESTs gb	1.80E-33
74	comp188686_g1_i1	serine threonine kinase	1.30E-68
75	comp174836_g2_i1	Hypothetical protein, putative	1.10E-11
76	comp160024_g1_i1	ferredoxin-NADP+ reductase	2.10E-94
77	comp185403_g2_i2	LDAH_CHICK	3.00E-18
78	comp189878_g1_i5	hypothetical protein GUITHDRAFT_121297	1.60E-28
79	comp166051_g1_i1	DLDH_SCHPO	2.50E-39
80	comp195856_g2_i3	dihydrolipoamide dehydrogenase	5.50E-52
81	comp186473_g2_i9	PSP1_YEAST	2.70E-18
82	comp179511_g1_i2	S-adenosyl-L-methionine-dependent methyltransferase	2.90E-12
83	comp193480_g1_i5	hypothetical protein NAEGRDRAFT_78864	1.80E-10
84	comp186473_g2_i1	cell cycle sequence binding phospho (RBP45),	1.40E-28
85	comp182403_g2_i4	ALA3_ARATH	0.00E+00
86	comp185910_g3_i9	core-2 i-branching beta--n-acetylglucosaminyltransferase family	5.10E-09
87	comp186498_g1_i12	mannosylphosphorylation	1.10E-29
88	comp179947_g2_i1	SMCR8_HUMAN	7.10E-10
89	comp167724_g1_i5	Hypothetical protein, putative	7.20E-12
90	comp174242_g1_i2	Transcription factor	1.70E-56
91	comp181437_g1_i4	fatty acyl- reductase	0.00E+00
92	comp153597_g1_i1	phospholipase	7.00E-33
93	comp182369_g4_i5	feline leukemia virus subgroup C receptor-related 2	4.50E-29
94	comp189490_g1_i19	SET domain-containing	3.90E-29

95	comp184828_g1_i4	SGTA_BOVIN	4.40E-33
96	comp186126_g5_i1	NLRC3_MOUSE	3.40E-13
97	comp192577_g2_i9	probable palmitoyltransferase ZDHHC1 isoform X1	2.00E-39
98	comp170558_g2_i1	conserved hypothetical protein, fragment	4.60E-15
99	comp52240_g1_i2	unnamed product	7.00E-12
100	comp184665_g1_i2	phosphoglycerate mutase	1.00E-26
101	comp184942_g1_i1	ABC transporter G family member 31	3.50E-21
102	comp188033_g1_i3	S-adenosylmethionine carrier 1, chloroplastic mitochondrial	1.40E-23
103	comp188994_g1_i4	predicted protein, partial	8.60E-82
104	comp189878_g1_i3	hypothetical protein GUITHDRAFT_121297	1.60E-28
105	comp190793_g2_i5	glycoside hydrolase	5.60E-93
106	comp179578_g1_i1	SNARE associated Golgi	9.50E-50
107	comp174011_g2_i5	hydroxyacylglutathione hydrolase	2.20E-110
108	comp187927_g1_i1	CAMT1_ARATH	3.50E-34
109	comp173575_g1_i2	TTD14_DROME	2.90E-70
110	comp193875_g2_i5	cyclin-A2predicted protein, partial	4.70E-10
111	comp183521_g1_i2	MNTH_RHILO	6.90E-129
112	comp189030_g1_i5	EGY1_ARATH	2.30E-26
113	comp191687_g2_i2	aldehyde dehydrogenase	3.80E-44
114	comp163805_g2_i1	Type I phosphodiesterase nucleotide pyrophosphatase	1.60E-36
115	comp191473_g1_i4	WD repeat-containing 90	1.60E-65
116	comp182402_g3_i1	ARMC1_MOUSE	4.60E-08
117	comp180639_g1_i2	FBX11_RAT	3.80E-19
118	comp178942_g1_i2	galactose mutarotase	3.90E-130
119	comp158867_g2_i3	F26_ARATH	3.60E-66
120	comp178145_g1_i1	DJ1B_ARATH	4.30E-30
121	comp192514_g1_i3	PDE2A_BOVIN	1.50E-65
122	comp186670_g4_i4	eukaryotic elongation factor-2 kinase	5.40E-48
123	comp187915_g1_i7	MYCE_MICGR	5.60E-12
124	comp179643_g2_i2	Adenylate guanylate cyclase with Chase sensor	1.20E-10
125	comp192568_g1_i2	NLRC3_HUMAN	1.40E-17
126	comp168433_g1_i2	BOB1_ARATH	4.40E-13
127	comp170428_g2_i3	kinase domain	4.40E-09
128	comp183962_g1_i1	RAY1_ARATH	1.10E-09
129	comp183228_g1_i1	zinc metallo ase-like	2.70E-37

130	comp181987_g2_i5	Hypothetical protein, putative	3.50E-11
131	comp194509_g1_i4	HELLS_HUMAN	1.60E-152
132	comp189297_g1_i3	transcription elongation regulator 1 isoform X3	1.10E-27
133	comp178720_g2_i7	THIO_EMENI	1.60E-12
134	comp194338_g1_i4	major facilitator superfamily domain-containing 12	3.30E-62
135	comp195344_g1_i8	e3 ubiquitin- ligase	8.40E-24
136	comp175433_g1_i3	GRSF1_MOUSE	5.60E-33
137	comp188350_g1_i1	SGMC_DICDI	3.80E-29
138	comp190314_g1_i6	BACH_RAT	5.40E-43
139	comp195435_g1_i4	histidine kinase	3.10E-20
140	comp182536_g1_i1	PHOT2_ARATH	2.70E-22
141	comp188049_g1_i3	PHSA_ARATH	8.70E-41
142	comp157624_g1_i2	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	5.90E-105
143	comp178224_g1_i4	kinase domain	1.80E-24
144	comp172139_g1_i2	HIS1_KLULA	5.00E-82
145	comp190368_g2_i2	KAPR_YARLI	1.30E-12
146	comp189939_g1_i8	E3 ubiquitin- ligase SIAH1-like	1.20E-12
147	comp178026_g3_i2	ADP-ribosyltransferase 1 precursor	2.00E-11
148	comp189485_g1_i1	GP107_MOUSE	3.30E-73
149	comp176000_g1_i5	probable S-acyltransferase 4	9.60E-45
150	comp195366_g2_i1	CFA74_HUMAN	6.40E-16
151	comp186019_g1_i1	PRY1_YEAST	4.10E-16
152	comp195971_g1_i3	MEG10_XENTR	6.80E-33
153	comp188962_g1_i5	YAFJ_ECOLI	5.10E-77
154	comp191739_g1_i2	leucine-rich repeat-containing 9	1.50E-115
155	comp194644_g1_i6	OTU-like cysteine protease domain-containing	2.30E-24
156	comp191206_g1_i4	receptor-type adenylate cyclase GRESAG 4,	1.80E-74
157	comp191442_g2_i3	MAM and LDL-receptor class A domain-containing 1	8.30E-13
158	comp195796_g2_i2	TARB1_HUMAN	2.80E-39
159	comp184149_g1_i2	EFCB6_HUMAN	1.90E-08
160	comp193839_g2_i9	tocopherol phytyltransferase	2.10E-58
161	comp186653_g2_i5	serine threonine- kinase 2 19-like	2.70E-36
162	comp191944_g1_i1	CASD1_DANRE	1.60E-07
163	comp186329_g5_i1	RETST_DANRE	3.20E-62
164	comp179182_g1_i1	PREDICTED: alpha-(1,6)-fucosyltransferase	1.50E-07

165	comp167679_g1_i2	FBT2_ARATH	4.60E-08
166	comp180128_g1_i6	TBC1 domain family member 31	3.10E-28
167	comp182359_g1_i5	GSP_ECOLI	1.50E-40
168	comp183078_g2_i6	hypothetical protein H310_03776	1.70E-15
169	comp188423_g1_i4	-glutamine gamma-glutamyltransferase 2-like	9.10E-66
170	comp192944_g1_i1	GRDP1_ARATH	2.20E-45
171	comp183701_g2_i2	lipid phosphate phosphatase 2	5.50E-25
172	comp180998_g3_i1	Cysteine dioxygenase type 1	4.20E-41
173	comp189913_g1_i2	LH18_EUGGR	3.30E-101
174	comp167330_g1_i3	hypothetical protein COCSUDRAFT_68410	1.40E-14
175	comp179243_g1_i2	ECE1_MOUSE	7.00E-94
176	comp190559_g1_i8	CYAA_LEIDO	3.50E-45
177	comp185922_g1_i1	ubiquitin- ligase, PUB49	5.90E-111
178	comp194522_g3_i16	NOXA_DICDI	2.80E-24
179	comp178959_g1_i5	hypothetical protein PBRA_003427	2.40E-99
180	comp191406_g1_i3	SRFR1_ARATH	7.90E-09
181	comp189274_g1_i2	nucleolin	1.70E-13
182	comp183605_g1_i3	NLRC3_HUMAN	6.90E-40
183	comp174228_g1_i1	RA51C_CRIGR	1.40E-23
184	comp187004_g2_i1	YTFP_BACSURecName: Full=Uncharacterized protein YtfP	4.50E-13
185	comp156508_g2_i1	YCXA_EUGGR	0.00E+00
186	comp190515_g1_i2	LON peptidase N-terminal domain and RING finger 1-like	2.40E-12
187	comp190550_g2_i4	MNME_SYNR3	1.20E-147
188	comp146646_g1_i3	Uncharacterized conserved	5.10E-19
189	comp182239_g1_i1	CRY2_ARATH	7.60E-63
190	comp184895_g2_i5	COBT_PSEF5	3.40E-38
191	comp188015_g2_i2	calcineurin-like phosphoesterase,	1.40E-71
192	comp181618_g1_i3	PREDICTED: uncharacterized protein LOC100179133	4.00E-09
193	comp193906_g1_i11	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.10E-07
194	comp186503_g2_i2	formin 5	1.50E-25
195	comp156890_g1_i2	LUL3_ARATH	1.70E-22
196	comp189180_g1_i3	TP4AP_HUMAN	1.70E-14
197	comp189992_g1_i1	ACAC_CHICK	0.00E+00
198	comp190407_g1_i11	MNN4_YEAST	2.60E-13
199	comp192683_g1_i2	dnaJ homolog 1, mitochondrial-like	8.90E-106

200	comp187069_g1_i5	carotenoid 9,10(9 ,10)-cleavage dioxygenase	4.50E-73
201	comp185151_g1_i1	THLA_PEPD6	3.20E-83
202	comp183840_g1_i1	sodium hydrogen exchanger 6-like	1.70E-93
203	comp130210_g1_i1	CHLI_EUGGR	0.00E+00
204	comp176522_g2_i3	PSAB_EUGGR	1.60E-81
205	comp192108_g2_i2	UDP-glucose:glyco glucosyltransferase	0.00E+00
206	comp180579_g1_i7	DDL_RALME	3.10E-09
207	comp187480_g3_i2	lipase family	1.10E-65
208	comp190011_g1_i5	phospholipase D alpha 1	3.30E-163
209	comp177444_g1_i3	UGAL1_ARATH	1.10E-09
210	comp192446_g1_i1	alpha-L-rhamnosidase	2.40E-63
211	comp193424_g1_i16	Y1691_LEGPH	4.00E-33
212	comp141780_g1_i2	PX24D_DICDI	1.50E-09
213	comp189464_g2_i5	FPG_ARATH	9.40E-54
214	comp193210_g1_i4	Y2594_STRCO	3.60E-29
215	comp173140_g1_i2	tyrosine phosphatase	1.40E-14
216	comp196238_g1_i7	PAS domain-containing sensor histidine kinase	5.10E-20
217	comp188742_g1_i6	TIGR02757 family	1.10E-46
218	comp190508_g3_i1	PBP_BRUA2	1.80E-15
219	comp193210_g2_i2	DnaJ domain	2.30E-102
220	comp180336_g1_i5	YB4E_SCHPO	1.50E-20
221	comp149588_g1_i1	TXN4A_HUMAN	7.30E-80
222	comp181964_g1_i1	WAP53_BOVIN	1.50E-66
223	comp165672_g1_i3	AT18A_ARATH	4.10E-89
224	comp196071_g1_i3	IFKB_DICDI	7.20E-17
225	comp186732_g1_i2	mitochondrial carrier	1.00E-42
226	comp166728_g1_i1	stress-induced-phospho 1	1.30E-23
227	comp156743_g1_i2	dynein light chain	1.70E-80
228	comp175811_g1_i1	TVP38_COPC7	5.10E-09
229	comp194293_g1_i6	IP6K3_MOUSE	3.20E-30
230	comp154037_g1_i4	OGFR_MOUSE	3.90E-36
231	comp170841_g1_i8	TRAM LAG1 CLN8 homology domain	1.90E-17
232	comp186550_g1_i7	exonuclease 3 -5 domain-containing 1-like	4.20E-39
233	comp195831_g1_i5	beta-galactosidase	0.00E+00
234	comp192143_g1_i1	DEAD DEAH box helicase	2.60E-16

235	comp181371_g1_i5	hypothetical protein	6.10E-08
236	comp189939_g1_i6	E3 ubiquitin- ligase Siah1	1.80E-09
237	comp172160_g1_i1	MFT2 - Corn MFT	1.80E-13
238	comp188033_g2_i9	SAMC2_ARATH	5.40E-22
239	comp185413_g2_i10	DBP2_PICST	1.20E-43
240	comp188714_g1_i2	FG-GAP repeat	7.40E-18
241	comp192508_g2_i3	serine threonine- phosphatase 6 regulatory ankyrin repeat subunit A-like isoform X1	2.40E-09
242	comp181111_g1_i4	MYBPP_HUMAN	1.00E-11
243	comp193615_g1_i6	AP4B1_HUMAN	2.20E-20
244	comp186862_g1_i1	glycosyl transferase family 2	1.90E-33
245	comp173370_g1_i6	hypothetical protein ACA1_171730	1.80E-23
246	comp192723_g1_i2	KAD2_BOVIN	6.30E-41
247	comp159676_g1_i1	ubiquitin ligase	4.90E-16
248	comp191790_g1_i1	WWP2_HUMAN	1.20E-36
249	comp194353_g2_i5	AChain A, Crystal Structure Of Engineered Northeast Structural Genomics Consortium Target	1.70E-17
250	comp188332_g3_i2	EOGT_DROME	2.50E-08
251	comp179619_g1_i5	NPC2_YARLI	2.20E-12
252	comp194028_g1_i8	ANM7_CRILO	5.60E-33
253	comp63187_g1_i2	predicted protein, partial	7.90E-52
254	comp175930_g2_i4	NLRC3_HUMAN	1.90E-25
255	comp190572_g3_i1	Glycosyltransferase sugar-binding region containing DXD motif	3.40E-16
256	comp155100_g1_i1	TL19_ARATH	9.30E-21
257	comp182081_g1_i10	type II inositol 1,4,5-trisphosphate 5-phosphatase	2.10E-53
258	comp177871_g1_i1	GTPase activating	6.50E-43
259	comp156422_g1_i2	RBL_EUGGR	0.00E+00
260	comp192307_g1_i8	predicted protein	5.60E-17
261	comp192307_g1_i5	predicted protein	9.10E-19
262	comp190716_g1_i11	alpha-type kinase	3.50E-46
263	comp195000_g1_i3	repressed by EFG1 1-like isoform X4	6.70E-21
264	comp188566_g1_i9	MYCF_MICGR	1.00E-33
265	comp165770_g1_i4	Ctr2 family transporter: copper ion CTR-type copper transporter	2.10E-26
266	comp179760_g1_i7	YOS3_SCHPO	2.50E-60
267	comp183117_g1_i14	ZDH11_HUMAN	3.90E-20

268	comp163442_g1_i1	U483_DROME	2.90E-21
269	comp176566_g1_i11	DSP8_ARATH	1.40E-17
270	comp184416_g3_i2	glucan 1,3-beta-glucosidase	2.10E-28
271	comp188879_g1_i4	cAMP specific phosphodiesterase	6.10E-172
272	comp184551_g2_i3	cyclic nucleotide-binding domain	2.50E-07
273	comp190716_g1_i13	alpha-type kinase	2.70E-46
274	comp189196_g1_i1	methyl transferase	1.60E-28
275	comp146534_g3_i2	ATPI_EUGGR	2.30E-149
276	comp185842_g1_i3	NLRC3_HUMAN	5.00E-19
277	comp188247_g1_i2	MA1A1_RABIT	1.30E-101
278	comp179364_g1_i1	ABC transporter substrate-binding	1.10E-41
279	comp188448_g1_i3	alpha-(1,6)-fucosyltransferase	3.40E-20
280	comp185495_g1_i1	PEPD_LACSK	1.40E-24
281	comp187484_g3_i4	hypothetical protein PTSG_01283	5.50E-128
282	comp190975_g2_i9	NLRC3_HUMAN	1.40E-58
283	comp174015_g1_i11	rhodanese	4.50E-15
284	comp166376_g1_i3	DESI2_DANRE	3.30E-25
285	comp179054_g1_i2	S38AB_MOUSE	3.20E-51
286	comp177002_g1_i1	NHP6_DEBHA	6.20E-07
287	comp179769_g1_i2	stress-induced-phospho 1	5.30E-13
288	comp181515_g1_i12	hypothetical protein	8.30E-08
289	comp180978_g1_i5	ATP-grasp domain containing 1	4.40E-132
290	comp180755_g1_i4	secreted predicted protein, partial	4.50E-12
291	comp186240_g1_i2	Tubulin binding	1.40E-23
292	comp190688_g1_i8	Aspartic peptidase	5.80E-19
293	comp171450_g1_i12	PSBD_EUGGR	1.20E-138
294	comp159360_g1_i3	AB1F_ARATH	0.00E+00
295	comp192934_g1_i3	hypothetical protein NAEGRDRAFT_78864	2.20E-10
296	comp171649_g2_i1	PREDICTED: uncharacterized protein LOC106152497	1.80E-31
297	comp184642_g2_i12	membrane-bound adenylyl cyclase	2.20E-40
298	comp167591_g1_i2	beta-type IP39	8.50E-15
299	comp191912_g1_i1	AARA_DICDI	1.10E-10
300	comp164091_g1_i2	lateral signaling target 2 homolog	5.60E-07
301	comp193659_g1_i5	NLRC3_HUMAN	5.40E-30
302	comp179258_g1_i2	holo-[acyl-carrier-] synthase	0.00E+00

303	comp187184_g1_i3	SWET1_XENTR	3.40E-12
304	comp183035_g2_i5	hypothetical protein	2.00E-10
305	comp164821_g2_i3	beta-type IP39	7.60E-09
306	comp181796_g1_i2	GTP pyrophosphokinase	1.30E-74
307	comp186718_g1_i1	exocyst complex component 5	3.10E-08
308	comp182271_g1_i4	ABCD4_MOUSE	2.50E-66
309	comp191850_g1_i4	MENA_SYNY3	5.90E-49
310	comp151977_g1_i1	chromatin assembly factor 1 subunit FAS2	2.00E-68
311	comp194180_g1_i1	NHL repeat containing	1.70E-31
312	comp194215_g1_i1	hypothetical protein AURANDRAFT_67549	5.10E-13
313	comp181427_g1_i2	hypothetical protein NAEGRDRAFT_59723	3.10E-08
314	comp188892_g1_i11	c transferase	3.00E-17
315	comp191609_g1_i1	BRISC and BRCA1-A complex member 1-like	4.70E-14
316	comp192979_g1_i5	Heat shock	0.00E+00
317	comp193640_g1_i3	PAS domain-containing sensor histidine kinase	6.90E-22
318	comp191038_g1_i1	UVRD_RICTY	1.10E-37
319	comp174505_g2_i1	Poly(ADP-ribose) polymerase catalytic domain containing	4.30E-89
320	comp179159_g1_i4	DED1_PHANO	0.00E+00
321	comp188859_g1_i11	hypothetical protein H310_00261	3.30E-09
322	comp183739_g1_i1	PBP_BRUA2	3.50E-14
323	comp190305_g1_i6	hypothetical protein ZOSMA_452G00060	8.70E-10
324	comp190179_g3_i2	KAPR_YARLI	5.60E-17
325	comp188981_g4_i4	ITS3_SCHPO	4.70E-54
326	comp195624_g1_i2	ef hand family	6.70E-24
327	comp194349_g1_i4	kinase domain	2.40E-18
328	comp195127_g1_i7	hypothetical protein	2.30E-08
329	comp184731_g3_i1	NLRC3_MOUSE	3.10E-35
330	comp187437_g1_i3	exostosin	5.00E-20
331	comp190809_g1_i5	fibronectin type-III domain-containing 3A-like	4.60E-50
332	comp192103_g1_i4	AMPP1_PARBD	5.80E-176
333	comp188238_g1_i1	PLA16_ARATH	2.40E-14
334	comp194362_g2_i2	transmembrane	1.90E-13
335	comp183355_g1_i3	PREDICTED: uncharacterized protein LOC103046511 isoform X1	7.90E-80
336	comp193201_g1_i2	CYA1_RHIME	1.50E-31
337	comp159124_g1_i3	plastid terminal oxidase	1.60E-92

338	comp177963_g2_i9	TRMJ_PECAS	1.90E-26
339	comp193657_g1_i7	SPXS1_DICDI	4.60E-45
340	comp181083_g1_i1	predicted protein	4.20E-15
341	comp161932_g2_i2	hypothetical protein SELMODRAFT_102324	1.70E-15
342	comp160864_g1_i4	SAMC1_ARATH	1.20E-10
343	comp184080_g1_i4	hypothetical protein SAMD00019534_007780predicted protein, partial	1.20E-14
344	comp178763_g1_i1	TTD14_DROME	2.20E-34
345	comp187283_g6_i7	YDJX_ECOLI	5.30E-07
346	comp191347_g1_i3	magnesium transporter	4.80E-33
347	comp195534_g2_i1	PIF1_EMENI	4.60E-84
348	comp186851_g1_i1	tRNA-Thr(GGU) m(6)t(6)A37 methyltransferase	1.10E-11
349	comp177153_g1_i3	Mg2+ transporter-e family	7.40E-41
350	comp184908_g1_i7	PP2A3_PARTE	5.30E-42
351	comp191923_g1_i3	STK11_CHICK	3.50E-53
352	comp191578_g1_i5	TREP_THEBR	0.00E+00
353	comp179941_g1_i3	radical SAM domain containing	1.40E-73
354	comp173379_g1_i3	Phytanoyl- dioxygenase	4.40E-84
355	comp179964_g1_i3	TVP38_COPC7	3.40E-09
356	comp188925_g1_i2	PI42C_DANRE	4.30E-13
357	comp182908_g1_i4	unnamed product	4.90E-45
358	comp171241_g1_i5	integrator complex subunit 11	0.00E+00
359	comp176037_g1_i2	inositol 1,4,5-trisphosphate receptor type 3	4.60E-15
360	comp192808_g1_i3	predicted protein	9.50E-13
361	comp194220_g2_i5	ef hand family	6.40E-32
362	comp186502_g2_i3	phosphatidylinositol-4-phosphate 5-kinase	4.20E-55
363	comp191263_g3_i11	conserved unknown protein	5.10E-16
364	comp189471_g3_i7	KLHDB_DROSI	1.90E-18
365	comp194861_g1_i2	hypothetical protein Ctob_007064	4.90E-39
366	comp186260_g3_i1	exostosin	1.90E-11
367	comp195750_g3_i3	threonine synthase	4.00E-32
368	comp183677_g1_i1	FAEB_PIREQ	2.30E-10
369	comp194555_g3_i2	Sodium-dependent proline transporter	2.10E-79
370	comp194271_g1_i3	sn1-specific diacylglycerol lipase beta	9.60E-41
371	comp191569_g2_i5	SGMC_DICDI	8.20E-29
372	comp186188_g1_i15	predicted protein	1.90E-18

373	comp191923_g1_i10	serine threonine- kinase STK11 isoform X1	5.40E-64
374	comp175936_g1_i8	IF4E_CANGA	1.10E-14
375	comp171365_g1_i5	hypothetical protein SacmaDRAFT_2621	1.30E-08
376	comp180639_g2_i3	FBX11_RAT	3.10E-19

Appendix table 22 genes name of area 22

(a) List of TF family of part A (12 transcripts of 9 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp178116_g1_i1	FAR1	1.20E-08
2	comp184227_g1_i1	bHLH	8.90E-07
3	comp185474_g1_i1	C3H	2.30E-09
4	comp190280_g1_i1	C2H2	1.20E-12
5	comp187307_g1_i12	C3H	7.40E-18
6	comp185184_g2_i1	bHLH	4.90E-26
7	comp186095_g1_i1	MYB_related	3.60E-14
8	comp170769_g1_i10	GATA	2.20E-07
9	comp188750_g2_i1	GeBP	2.30E-08
10	comp187898_g1_i3	NF-YA	7.60E-24
11	comp186008_g2_i5	WRKY	1.80E-73
12	comp187307_g1_i8	C3H	2.80E-18

(b) List of other genes name of part A (43 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp181541_g1_i1	TWF_DICDI	6.10E-23
2	comp188599_g3_i2	TOP3B_MOUSE	1.90E-36
3	comp188423_g1_i8	TGM5_HUMAN	4.20E-40
4	comp193426_g1_i4	GPDM_MOUSE	6.30E-140
5	comp192456_g1_i1	CC189_MOUSE	2.00E-17
6	comp196262_g1_i6	plant organellar DNA polymerase	1.80E-47
7	comp187601_g1_i1	CSD_METTH	3.50E-41
8	comp179154_g1_i2	DNA-directed RNA polymerase III subunit RPC5	4.30E-47
9	comp190785_g3_i3	TEB_ARATH	5.20E-160
10	comp192314_g1_i1	CC108_MOUSE	7.60E-131
11	comp192157_g1_i4	OTU domain-containing 7A	1.10E-23
12	comp179308_g1_i1	CDKAL_DROME	2.70E-70
13	comp183216_g1_i3	CRNS1_HUMAN	4.30E-101
14	comp175936_g1_i7	eukarytic translation initiation factor	1.50E-12
15	comp153856_g1_i3	hypothetical protein CHLNCDRAFT_145050	5.30E-58

16	comp178426_g1_i13	KAPS_SCHPO	1.60E-84
17	comp189758_g2_i11	AARA_DICDI	1.20E-10
18	comp170571_g1_i3	mitochondrial import inner membrane translocase subunit Tim17-B	1.30E-30
19	comp188238_g1_i2	PLA16_ARATH	1.90E-14
20	comp180356_g1_i2	SINA_DROME	1.90E-27
21	comp179117_g1_i1	F26_ARATH	9.40E-79
22	comp192841_g2_i11	hypothetical protein Ctob_010698	9.00E-08
23	comp192703_g2_i1	ankyrin repeat	4.10E-28
24	comp186014_g1_i6	6GAL_HYPRU	3.30E-30
25	comp190852_g3_i6	hypothetical protein AMJ87_05245	3.30E-08
26	comp174430_g2_i2	dual specificity phosphatase 22-like	2.20E-08
27	comp174958_g1_i6	cystatin-A	9.80E-20
28	comp186389_g1_i3	Peptidase S16, lon N-terminal	1.10E-35
29	comp188478_g6_i2	adenylate cyclase	4.80E-07
30	comp172770_g8_i4	Endonuclease exonuclease phosphatase	1.20E-27
31	comp194950_g1_i1	ANM7_XENLA	6.10E-18
32	comp105272_g1_i2	EFTU_EUGGR	0.00E+00
33	comp177997_g2_i1	SDP1L_ARATH	1.20E-62
34	comp186448_g1_i1	phage infection	2.70E-31
35	comp187279_g3_i7	uridine kinase	2.40E-66
36	comp188878_g3_i6	EXPR_XANCP	2.00E-72
37	comp190380_g1_i5	NAD-dependent deacetylase sirtuin-6	7.90E-26
38	comp189367_g1_i2	KHA1_YEAST	9.10E-22
39	comp181078_g2_i2	beta-type IP39	3.50E-22
40	comp192951_g1_i2	AChain A, Crystal Structure Of Engineered Northeast Structural Genomics Consortium Target	1.80E-13
41	comp190368_g2_i4	KAPR_YARLI	1.40E-12
42	comp194941_g1_i1	HERC1_HUMAN	1.30E-65
43	comp178248_g1_i3	hypersensitive-induced response 3-like	5.00E-12

(c) List of other genes name of part B (27 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp189480_g1_i2	Rare lipo A	4.30E-29
2	comp165565_g1_i1	TTC8_MOUSE	0.00E+00

3	comp170754_g2_i6	SMYD3_MOUSE	2.50E-07
4	comp181935_g1_i2	CABL2_MOUSE	3.20E-28
5	comp196317_g1_i3	ITPR1_HUMAN	1.60E-55
6	comp195632_g1_i3	Tbc2 translation chloroplastic	4.40E-39
7	comp174015_g1_i13	rhodanese	6.30E-15
8	comp193306_g1_i1	TRAP1_MOUSE	0.00E+00
9	comp190703_g1_i6	hypothetical protein SAMD00019534_055790	2.20E-16
10	comp194462_g1_i8	Transcription factor	1.40E-55
11	comp193370_g2_i3	NEK1_MOUSE	9.90E-67
12	comp190197_g3_i4	HMG2_DROME	2.10E-14
13	comp187000_g1_i7	GEFF_DICDI	2.30E-10
14	comp171176_g1_i5	XRN4_ARATH	2.00E-73
15	comp175223_g2_i2	DCA13_XENLA	3.50E-113
16	comp176677_g2_i2	SYW_NOSS1	3.70E-136
17	comp171450_g1_i2	PSBD_EUGGR	5.40E-180
18	comp194882_g1_i1	SHEP_DROSI	1.00E-10
19	comp171105_g1_i1	ZNT9_PONAB	2.30E-34
20	comp194302_g1_i6	NLRC3_HUMAN	3.90E-54
21	comp190625_g1_i6	voltage-gated ion channel superfamily	3.60E-86
22	comp195342_g1_i2	zinc finger transcription factor sma,	1.10E-27
23	comp188525_g1_i2	phosphatidylinositol 4-phosphate 5-kinase 1	1.10E-13
24	comp184826_g1_i1	predicted protein	2.50E-18
25	comp171241_g1_i3	INT11_DICDI	0.00E+00
26	comp166757_g1_i1	SMS2_MOUSE	2.90E-09
27	comp193906_g1_i9	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07

(d) List of other genes name of part C (31 transcripts)

No.	Seqname of Euglena	Annotation	Evalue
1	comp193627_g1_i8	EF-hand calcium-binding domain-containing 6	3.60E-12
2	comp253604_g1_i1	Hypothetical protein, putative	6.50E-08
3	comp187464_g1_i2	ATP-binding cassette transporter	3.00E-21
4	comp182281_g1_i2	DHKL_DICDI	2.30E-12
5	comp108746_g1_i2	Golgi SNAP receptor complex member 2	2.40E-09
6	comp183895_g1_i1	rieske [2Fe-2S] domain	6.30E-10

7	comp180616_g1_i9	THI4_SCHPO	9.90E-74
8	comp187716_g1_i8	capsular biosynthesis	5.00E-12
9	comp196105_g2_i3	beta subunit of photoactivated adenylyl cyclase	1.60E-16
10	comp179296_g1_i1	hypothetical protein PTSG_10764	2.60E-10
11	comp193353_g2_i4	NOA1_ARATH	1.10E-42
12	comp183323_g1_i1	COPB2_YEAST	2.60E-18
13	comp192673_g2_i2	CYA1_RHIME	3.20E-14
14	comp194541_g3_i5	WD-40 repeat	6.20E-18
15	comp191464_g2_i2	RECQ_ECOLI	1.20E-21
16	comp171660_g1_i2	NOVP_STRNV	1.60E-42
17	comp184190_g1_i3	UBE11_WHEAT	2.00E-100
18	comp189905_g3_i1	OTUD3_MOUSE	8.70E-08
19	comp192844_g1_i2	DEXH1_ARATH	5.00E-87
20	comp190240_g1_i2	DEAD DEAH box helicase	2.50E-16
21	comp176838_g1_i2	hypothetical protein PBRA_003427	1.70E-99
22	comp185965_g1_i2	CNG1_CHICK	5.90E-48
23	comp194344_g1_i16	histidine kinase	3.00E-09
24	comp185343_g2_i7	SPNS1_ARATH	3.00E-37
25	comp187120_g1_i1	peptidyl-prolyl cis-trans isomerase ( ),	9.00E-38
26	comp191508_g1_i1	NLRC3_HUMAN	2.10E-07
27	comp191861_g1_i1	TUB1_CAEBR	2.30E-13
28	comp188008_g1_i3	RING finger	2.00E-36
29	comp183645_g1_i6	RTCA_BOVIN	1.90E-64
30	comp193484_g2_i3	regulator of chromosome condensation RCC1	9.20E-19
31	comp178799_g1_i6	NADB_PSEAE	3.60E-74

Appendix table 23 genes name of area 23

(a) List of TF family of part A (102 transcripts of 25 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp188378_g1_i2	G2-like	9.00E-13
2	comp188378_g5_i4	G2-like	1.70E-11
3	comp187665_g1_i3	C3H	5.60E-12
4	comp195501_g1_i1	C2H2	2.00E-20
5	comp190761_g1_i3	NAC	6.80E-22
6	comp187971_g1_i5	C3H	4.80E-09
7	comp173326_g2_i2	MYB_related	3.10E-25
8	comp195617_g1_i2	C2H2	3.80E-08
9	comp188093_g1_i5	C3H	1.90E-07
10	comp170769_g1_i6	GATA	3.20E-07
11	comp195453_g1_i6	LBD	1.90E-44
12	comp191155_g1_i1	NAC	3.10E-81
13	comp186985_g2_i1	B3	7.70E-57
14	comp186008_g2_i9	WRKY	6.00E-74
15	comp187210_g1_i1	YABBY	8.50E-24
16	comp180760_g1_i1	HB-other	9.80E-07
17	comp188137_g1_i4	MYB_related	4.80E-152
18	comp185588_g1_i1	bHLH	1.90E-23
19	comp189067_g1_i3	G2-like	1.90E-08
20	comp191155_g1_i2	NAC	2.70E-80
21	comp186106_g1_i2	WRKY	6.80E-23
22	comp175254_g1_i3	MYB_related	3.70E-09
23	comp173950_g2_i1	CPP	4.90E-10
24	comp183220_g1_i3	E2F_DP	4.00E-11
25	comp188645_g1_i3	B3	3.50E-09
26	comp174474_g1_i3	GATA	1.20E-22
27	comp188771_g1_i7	MYB_related	1.90E-152
28	comp191207_g1_i3	NAC	6.40E-70
29	comp185723_g4_i2	bHLH	6.60E-12
30	comp184982_g1_i2	HB-other	7.60E-09
31	comp187859_g1_i4	YABBY	1.50E-16
32	comp193899_g3_i6	GATA	9.50E-23

33	comp191207_g1_i4	NAC	3.60E-69
34	comp186219_g1_i3	WRKY	2.00E-12
35	comp188680_g1_i2	B3	4.90E-26
36	comp189459_g2_i2	HB-other	5.60E-55
37	comp185900_g1_i1	bHLH	8.50E-09
38	comp194244_g2_i3	FAR1	1.30E-46
39	comp188771_g1_i8	MYB_related	7.70E-153
40	comp188009_g1_i1	YABBY	3.20E-32
41	comp175483_g1_i2	MYB_related	4.50E-24
42	comp191207_g1_i9	NAC	5.50E-70
43	comp186228_g1_i11	WRKY	9.50E-40
44	comp162724_g1_i3	GATA	1.50E-16
45	comp185184_g1_i4	TCP	1.50E-25
46	comp190348_g2_i3	HB-other	3.10E-08
47	comp185900_g1_i3	bHLH	6.50E-09
48	comp185900_g1_i7	bHLH	8.20E-09
49	comp193842_g1_i1	Nin-like	8.30E-62
50	comp183451_g1_i1	YABBY	3.40E-31
51	comp185942_g1_i1	bHLH	3.70E-17
52	comp194278_g3_i6	WRKY	8.50E-10
53	comp192565_g1_i7	BBR-BPC	6.50E-22
54	comp173138_g1_i2	C3H	9.70E-22
55	comp181098_g1_i2	FAR1	1.20E-37
56	comp188918_g1_i4	MYB_related	5.20E-08
57	comp186765_g1_i4	WRKY	2.40E-22
58	comp184651_g3_i2	YABBY	7.90E-30
59	comp191755_g1_i8	NAC	1.20E-69
60	comp192725_g1_i3	EIL	5.00E-10
61	comp193857_g1_i2	MYB_related	2.20E-27
62	comp188922_g1_i4	B3	3.40E-33
63	comp178162_g1_i2	CPP	8.30E-10
64	comp173380_g1_i1	C3H	3.30E-30
65	comp193203_g1_i1	bHLH	4.50E-34
66	comp173707_g1_i4	C3H	3.50E-30
67	comp189786_g3_i4	MYB_related	5.80E-08

68	comp192182_g1_i1	NF-YA	2.40E-30
69	comp186765_g1_i5	WRKY	2.50E-22
70	comp184821_g1_i2	C2H2	4.20E-11
71	comp175483_g1_i3	MYB_related	9.90E-25
72	comp192182_g1_i2	NF-YA	1.60E-30
73	comp186905_g1_i2	WRKY	1.90E-22
74	comp173707_g1_i5	C3H	3.50E-30
75	comp192182_g1_i5	NF-YA	2.40E-30
76	comp193203_g1_i4	bHLH	4.30E-34
77	comp185085_g1_i4	C2H2	5.20E-47
78	comp173975_g1_i2	C3H	1.20E-18
79	comp186587_g1_i5	MYB	2.50E-12
80	comp186905_g1_i5	WRKY	2.00E-22
81	comp192711_g1_i3	BBR-BPC	1.80E-27
82	comp184664_g1_i2	ERF	4.00E-13
83	comp181098_g1_i3	FAR1	1.30E-37
84	comp194296_g1_i1	FAR1	1.20E-18
85	comp188160_g1_i2	GRAS	1.10E-16
86	comp192037_g1_i9	LBD	1.90E-17
87	comp173920_g1_i3	NF-YB	8.70E-20
88	comp193842_g1_i3	Nin-like	8.30E-62
89	comp185942_g1_i5	bHLH	3.10E-17
90	comp186905_g1_i7	WRKY	2.40E-22
91	comp194278_g3_i2	WRKY	8.50E-10
92	comp191865_g4_i3	bZIP	8.70E-07
93	comp189009_g2_i2	B3	3.00E-26
94	comp185904_g2_i3	C2H2	3.80E-25
95	comp192725_g1_i5	EIL	2.80E-10
96	comp175483_g2_i2	MYB_related	1.20E-24
97	comp195313_g1_i2	MYB_related	6.60E-30
98	comp195309_g1_i8	Nin-like	1.70E-33
99	comp183451_g1_i6	YABBY	2.90E-31
100	comp185168_g1_i1	YABBY	6.50E-21
101	comp188523_g1_i2	MYB	1.40E-12
102	comp174158_g1_i1	C3H	1.20E-29

(b) List of other genes name of part A (288 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp194007_g1_i8	ELAV 2	2.20E-13
2	comp195410_g1_i1	DYH8_HUMAN	0.00E+00
3	comp189526_g1_i3	MFS-type transporter SLC18B1-like	3.50E-10
4	comp190757_g1_i9	CY42_TRYBB	1.10E-20
5	comp179584_g1_i3	Y777_CHLL2	4.80E-09
6	comp166620_g1_i1	predicted protein	4.70E-19
7	comp187655_g1_i2	MRS2_SCHPO	4.20E-13
8	comp188522_g2_i1	NUD24_ARATH	1.00E-43
9	comp188085_g1_i8	kinase,	2.70E-49
10	comp183030_g1_i1	uncharacterized oxidoreductase At4g09670-like	6.40E-55
11	comp188324_g2_i2	predicted protein, partial	5.50E-129
12	comp174015_g1_i19	rhodanese	3.80E-15
13	comp195793_g1_i6	Soluble guanylate cyclase gcy-31	1.20E-09
14	comp184332_g1_i8	PEPD_LACSK	1.30E-24
15	comp184358_g1_i8	CNRN_DICDI	1.90E-48
16	comp183367_g1_i4	TNI3K_MOUSE	2.60E-10
17	comp195050_g1_i4	DNLI4_DICDI	1.10E-124
18	comp194882_g1_i5	SHEP_DROSI	9.30E-11
19	comp188621_g1_i17	copper transporter family	4.60E-13
20	comp180129_g1_i1	TAP Cterminal subfamily	3.20E-18
21	comp178069_g2_i3	enoyl- hydratase	1.00E-79
22	comp193746_g2_i1	CP062_MOUSE	2.40E-110
23	comp183962_g1_i3	UDP-galactose:fucoside alpha-3-galactosyltransferase	1.70E-10
24	comp179109_g1_i5	Y575_SYNY3	6.10E-23
25	comp195512_g1_i6	predicted protein	2.40E-09
26	comp194302_g1_i9	NLRC3_HUMAN	3.70E-54
27	comp192670_g1_i2	Kinesin KIF13A	6.10E-29
28	comp180639_g2_i1	FBX11_RAT	5.60E-19
29	comp148809_g1_i4	DNA endonuclease RBBP8	1.40E-12
30	comp189350_g1_i1	Zn-dependent hydrolase	8.10E-36
31	comp177146_g1_i2	RCSC_SALTY	2.30E-16

32	comp185495_g1_i9	PEPD_LACSK	1.40E-24
33	comp193056_g1_i2	serine threonine- kinase STN7, chloroplastic	1.60E-94
34	comp175439_g1_i4	predicted protein	4.40E-12
35	comp178461_g1_i7	hypothetical protein COCSUDRAFT_63432	2.00E-15
36	comp189758_g2_i3	AARA_DICDI	1.30E-10
37	comp163175_g1_i5	TAR3_ARATH	2.50E-70
38	comp190620_g1_i1	MSH2_CHLAE	4.60E-173
39	comp190925_g1_i3	UVB31_ARATH	8.20E-08
40	comp180809_g1_i2	hypothetical protein AALB_1947	0.00E+00
41	comp172128_g1_i2	PEAMT_SPIOL	1.70E-10
42	comp166391_g1_i2	NLRC3_HUMAN	8.60E-44
43	comp190165_g1_i1	Plastid lipid-associated fibrillin conserved domain	7.80E-119
44	comp175839_g1_i3	FBX11_MOUSE	8.10E-22
45	comp190762_g1_i6	TTD14_DROME	1.40E-74
46	comp193442_g1_i1	drug metabolite transporter superfamily	1.60E-85
47	comp185683_g1_i1	MA1A1_RABIT	3.40E-101
48	comp192253_g1_i6	CJ011_MOUSE	1.90E-14
49	comp177242_g1_i1	lycopene beta cyclase	3.10E-81
50	comp177405_g1_i1	fatty acid desaturase	1.50E-133
51	comp186594_g1_i2	predicted protein	4.40E-14
52	comp178379_g1_i2	ANK3_HUMAN	3.30E-40
53	comp163223_g1_i1	zinc-binding ,	1.20E-68
54	comp186187_g1_i9	phosphatidylethanolaminen-methyltransferase-lik e	5.80E-134
55	comp184727_g1_i1	GDIR_ARATH	5.00E-08
56	comp180844_g1_i1	conserved hypothetical protein	3.10E-07
57	comp190559_g1_i2	CYAA_LEIDO	4.50E-45
58	comp149989_g1_i1	PLDA1_TOBAC	4.70E-168
59	comp181527_g1_i1	L-asparaginase I	5.70E-92
60	comp192519_g1_i1	brucei -specific	2.10E-12
61	comp162177_g1_i3	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	9.60E-54
62	comp178666_g1_i5	TSR3_YEAST	1.20E-21
63	comp180440_g1_i2	1-deoxy-D-xylulose-5-phosphate reductoisomerase	0.00E+00
64	comp188747_g1_i8	hypothetical protein Ctob_000266	7.40E-09
65	comp193875_g2_i10	CCNA2_MESAU	1.10E-10
66	comp174587_g3_i1	hypothetical protein COCSUDRAFT_43198	2.30E-08

67	comp186810_g1_i14	phosphatidylinositol class b	1.10E-52
68	comp174994_g1_i2	PUF68_DROME	3.20E-14
69	comp195905_g1_i2	hypothetical protein CHLREDRAFT_151998	3.60E-69
70	comp169258_g1_i3	Dpy-30 motif	4.20E-28
71	comp196238_g1_i11	PAS domain-containing sensor histidine kinase	5.30E-20
72	comp187283_g6_i2	predicted protein, partial	5.00E-26
73	comp194745_g2_i2	methionine synthase	0.00E+00
74	comp186790_g2_i5	ADP-ribosylation Crystallin J1	3.40E-70
75	comp192241_g1_i6	Y8013_DICDI	1.70E-10
76	comp176522_g2_i4	PSAB_EUGGR	0.00E+00
77	comp195833_g1_i6	ALA3_ARATH	2.80E-110
78	comp182887_g2_i3	hypothetical protein ABB37_09572	2.10E-29
79	comp186900_g1_i1	NNTM_MOUSE	5.60E-09
80	comp175449_g1_i4	NSR1_YEAST	2.00E-11
81	comp190070_g2_i3	major facilitator superfamily (MFS)	1.70E-47
82	comp173370_g1_i5	hypothetical protein ACA1_171730	2.10E-23
83	comp194293_g1_i13	inositol hexakisphosphate kinase 3	3.80E-31
84	comp183429_g1_i4	hypothetical protein crov263	7.70E-17
85	comp188650_g1_i2	RB3GP_RAT	4.80E-41
86	comp191053_g2_i4	TRPV6_HUMAN	3.50E-13
87	comp172243_g1_i1	Vesicle-associated membrane -associated Bpredicted protein, partial	2.00E-08
88	comp171649_g2_i3	PREDICTED: uncharacterized protein LOC106152497	1.10E-31
89	comp175746_g1_i3	coiled-coil domain-containing 6	5.70E-20
90	comp194749_g2_i5	CBPQ_BOVIN	5.10E-81
91	comp140574_g1_i2	bifunctional -like	1.70E-35
92	comp163062_g1_i5	2ABD_CHICK	1.00E-118
93	comp193627_g2_i3	hypothetical protein SDRG_04080	2.70E-12
94	comp170575_g1_i1	DHHC zinc finger multi-pass transmembrane	8.00E-14
95	comp193936_g1_i1	phosphatidylinositol 3-kinase Tor2	0.00E+00
96	comp186560_g2_i1	XYLB_MOUSE	3.80E-121
97	comp188869_g1_i9	probable palmitoyltransferase ZDHHC1 isoform X1	2.00E-39
98	comp173175_g1_i1	expressed unknown protein	3.00E-57
99	comp172015_g1_i3	WD repeat-containing 43-like	1.50E-08
100	comp180208_g1_i3	signal peptide peptidase-like, aspartyl protease family A22B,	1.50E-37
101	comp194271_g1_i6	DGLB_HUMAN	6.80E-15

102	comp179739_g1_i2	hypothetical protein Ctob_000830	1.90E-11
103	comp193509_g1_i12	MFS transporter	1.30E-11
104	comp185810_g1_i1	Y882_HAEINRecName: Full=Uncharacterized protein HI_0882	2.00E-70
105	comp189526_g1_i5	S18B1_MOUSE	1.50E-07
106	comp191431_g1_i2	dymeclin	4.70E-60
107	comp188609_g4_i5	MYCF_MICGR	7.50E-30
108	comp182286_g3_i3	CRFC_ECOLX	1.20E-37
109	comp171363_g1_i1	YITU_BACSU	9.40E-18
110	comp183277_g1_i1	NLRC5_MOUSE	3.70E-12
111	comp194515_g1_i1	UPL6_ARATH	2.30E-98
112	comp178298_g1_i2	GELA_DICDI	1.60E-08
113	comp174680_g1_i3	LIM domain containing	1.60E-14
114	comp184599_g1_i10	MGGB_PETMO	3.40E-19
115	comp177066_g1_i2	sodium-independent anion transporter	2.80E-62
116	comp191788_g2_i8	glycosyltransferase aer61	2.30E-07
117	comp179878_g2_i5	PI5L1_MOUSE	1.30E-09
118	comp170339_g1_i3	rap ran GTPase-activating	2.80E-16
119	comp168113_g1_i1	cAMP-dependent kinase type I regulatory subunit isoform X2	7.90E-77
120	comp173828_g1_i3	SIAH1_CAEBR	4.40E-08
121	comp183412_g1_i4	SGTA_BOVIN	3.50E-33
122	comp192942_g1_i1	PBP_BRUA2	9.10E-31
123	comp176861_g1_i1	PRP31_MOUSE	8.90E-63
124	comp176808_g1_i2	XYNB_BACSU	2.30E-114
125	comp151206_g1_i3	PSBB_EUGGR	1.40E-135
126	comp177152_g1_i1	predicted protein, partial	6.70E-74
127	comp177717_g2_i2	PSTS_HAEIN	2.10E-12
128	comp185620_g1_i3	RTEL1_DANRE	1.70E-161
129	comp184989_g1_i1	Y4139_ARATH	5.90E-39
130	comp173489_g1_i1	dnaJ homolog subfamily C member 22	9.30E-10
131	comp179447_g2_i3	flagellar associated	1.70E-102
132	comp149630_g2_i2	YHCV_BACSU	4.70E-15
133	comp172240_g2_i4	DPOE2_CHICK	5.90E-81
134	comp193049_g1_i4	hypothetical protein TRSC58_05274	5.80E-11
135	comp192614_g1_i6	CLPC2_ORYSJ	0.00E+00
136	comp192442_g1_i2	adenylate cyclase type 10	4.70E-39

137	comp193210_g1_i6	Y2594_STRCO	1.30E-24
138	comp154359_g1_i1	NLRC3_HUMAN	2.50E-45
139	comp195368_g1_i6	TRPV1_RABIT	4.60E-31
140	comp195081_g2_i2	SSH_XENLA	1.20E-15
141	comp167227_g3_i2	PEAR1_HUMAN	4.40E-11
142	comp194905_g1_i2	hypothetical protein AMSG_06883	1.20E-26
143	comp182656_g1_i3	hypothetical protein	9.90E-09
144	comp177058_g1_i4	Y233_STRP6RecName: Full=Uncharacterized protein M6_Spy0233	7.60E-08
145	comp184349_g5_i1	hypothetical protein GUITHDRAFT_105708	5.80E-10
146	comp186186_g1_i2	COB21_ORYSJ	1.00E-17
147	comp188150_g1_i3	PREDICTED: uncharacterized protein LOC100828653 isoform X1	1.50E-07
148	comp193186_g1_i4	MORN5_DANRE	1.90E-19
149	comp172102_g1_i1	type IV secretion Dot	2.00E-20
150	comp190703_g1_i2	hypothetical protein SAMD00019534_055790	2.10E-16
151	comp175038_g1_i5	Tetratricopeptide-like helical	3.00E-46
152	comp196238_g1_i10	PAS domain-containing sensor histidine kinase	4.00E-20
153	comp183844_g1_i4	E3 ubiquitin- ligase AMFR	8.20E-24
154	comp195681_g2_i4	DDX10_HUMAN	2.30E-148
155	comp190491_g1_i8	polyadenylate-binding 8	1.10E-10
156	comp181600_g1_i6	CYA1_RHIME	4.90E-33
157	comp167932_g3_i2	PP1R8_BOVIN	1.50E-29
158	comp189518_g2_i2	IQ calmodulin-binding ,	2.80E-13
159	comp181660_g1_i4	dolichol-p-mannose mannosyltransferase	1.70E-60
160	comp194775_g2_i2	serine threonine- kinase STN7, chloroplastic	3.20E-104
161	comp161874_g1_i2	GST1_SCHPO	2.10E-08
162	comp182448_g2_i1	KPYR_HUMAN	2.70E-118
163	comp194553_g1_i8	AL3B1_HUMAN	2.40E-122
164	comp184559_g1_i1	DEGP1_ARATH	1.20E-115
165	comp176615_g1_i3	hypothetical protein AMSG_03988	5.10E-08
166	comp181356_g2_i3	Histidine phosphatase superfamily, clade-1	4.90E-27
167	comp189023_g1_i3	PDE11_DROME	5.40E-74
168	comp186186_g1_i1	COB21_ORYSJ	9.70E-18
169	comp164955_g1_i3	serine threonine kinase	1.20E-21
170	comp158144_g1_i2	MCES1_ORYSJ	1.50E-39
171	comp176563_g1_i5	ubiquinone biosynthesis methyltransferase	1.10E-14

172	comp195047_g1_i3	membrane-associated protein, putative	4.00E-07
173	comp189908_g2_i4	S35D2_BOVIN	9.40E-63
174	comp192589_g1_i11	PI5K5_ARATH	2.50E-12
175	comp196238_g1_i6	PAS domain-containing sensor histidine kinase	4.80E-20
176	comp161181_g1_i2	CYSKM_ARATH	1.40E-78
177	comp191659_g1_i1	TOM1_NEUCR	7.50E-15
178	comp191225_g1_i1	SPKUL_DICDI	1.50E-42
179	comp193766_g1_i3	SHEP_DROSI	1.00E-10
180	comp170391_g1_i5	MSRA_RHOCS	4.60E-84
181	comp185213_g1_i2	dnaJ homolog subfamily B member 8	3.10E-10
182	comp185922_g2_i3	PPIL2_EMENI	4.30E-78
183	comp170185_g1_i4	hypothetical protein PPERSA_09341	2.50E-13
184	comp154783_g1_i3	OGFR_MOUSE	5.10E-36
185	comp98307_g1_i1	RR2_EUGGR	7.20E-126
186	comp195748_g2_i1	CBPC1_HUMAN	1.00E-99
187	comp183767_g2_i1	YDFJ_BACSU	1.30E-22
188	comp186502_g2_i2	phosphatidylinositol-4-phosphate 5-kinase	2.20E-55
189	comp171258_g1_i2	MYND finger domain	8.90E-60
190	comp194918_g2_i5	CATL_SARPE	3.00E-51
191	comp159286_g1_i1	predicted protein	1.10E-08
192	comp185046_g1_i6	sulfate transporter , chloroplastic-like	2.30E-44
193	comp195833_g1_i16	phospholipid-transporting ATPase 3	5.30E-170
194	comp163237_g1_i5	conserved unknown protein	2.80E-14
195	comp169253_g1_i2	cholestenol delta-isomerase	6.70E-27
196	comp195457_g1_i7	hypothetical protein GUITHDRAFT_142802	6.50E-18
197	comp178282_g1_i1	RUSD1_DANRE	7.70E-39
198	comp195369_g1_i2	phosphate ABC transporter substrate-binding	9.20E-52
199	comp138470_g1_i2	WDR92_MOUSE	2.50E-135
200	comp190380_g2_i2	SIR1_ORYSI	1.40E-21
201	comp143128_g1_i5	hypothetical protein TRSC58_03537	7.00E-07
202	comp192042_g1_i1	COG5_MOUSE	5.50E-56
203	comp177902_g1_i2	CJ011_MOUSE	4.40E-15
204	comp195471_g1_i14	ABC transporter substrate-binding	1.80E-44
205	comp191567_g2_i5	TBC1 domain family member 31-like	2.70E-28
206	comp172972_g1_i2	hydroxyacylglutathione hydrolase	1.30E-114

207	comp172349_g1_i2	DML_EUBBA	9.30E-66
208	comp169346_g1_i2	pap fibrillin family	8.70E-16
209	comp171886_g1_i1	expressed unknown protein	3.00E-57
210	comp188268_g1_i7	Y528_SYNY3	5.30E-09
211	comp191645_g2_i6	SVOP_MOUSE	3.40E-08
212	comp186867_g1_i3	(2Fe-2S)-binding	5.40E-12
213	comp187503_g1_i1	erythroid differentiation-related factor 1	1.20E-31
214	comp195270_g5_i1	signal transduction histidine kinase	2.40E-16
215	comp171295_g1_i1	NLRC3_HUMAN	1.10E-15
216	comp178766_g1_i1	RH46_ARATH	3.10E-132
217	comp185779_g1_i1	NLRC3_HUMAN	1.10E-27
218	comp152092_g1_i1	Y108_SYNY3	4.00E-10
219	comp194110_g1_i5	hypothetical protein	5.10E-20
220	comp156941_g1_i8	DEAD DEAH box helicase	2.90E-30
221	comp166376_g1_i2	DESI2_DANRE	3.10E-25
222	comp184877_g1_i4	hypothetical protein GLOINDRAFT_18528	2.20E-08
223	comp176727_g1_i12	IM30_ORYSJ	2.20E-54
224	comp186313_g2_i2	uracil-DNA glycosylase	2.70E-36
225	comp183017_g1_i1	YRHG_BACSU	5.10E-18
226	comp184321_g2_i5	ALKB_CAUCR	7.30E-10
227	comp185933_g2_i8	NLRC3_HUMAN	6.10E-27
228	comp187411_g2_i8	LOW PSII ACCUMULATION chloroplastic	8.10E-41
229	comp163268_g1_i3	CYSKP_SOLTU	2.20E-86
230	comp193086_g1_i1	EF-Hand calcium-binding site	0.00E+00
231	comp188577_g2_i1	exonuclease 1 isoform X1	3.80E-97
232	comp185939_g2_i1	B12-binding	0.00E+00
233	comp192399_g1_i12	Hypothetical protein, putative	4.40E-16
234	comp193043_g1_i1	GPDM_MOUSE	6.40E-140
235	comp181715_g1_i2	LVHK2_ERYLH	2.60E-08
236	comp168870_g1_i3	receptor-type adenylate cyclase a	8.30E-14
237	comp180165_g1_i1	Digalactosyldiacylglycerol synthase 1, chloroplastic	1.00E-75
238	comp186417_g1_i2	NLRC3_HUMAN	9.90E-24
239	comp177739_g1_i3	adenylyl cyclase	1.60E-08
240	comp187345_g1_i3	BROX_RAT	3.40E-10
241	comp170798_g1_i1	UN104_CAEEL	1.40E-10

242	comp192342_g2_i3	ABC transporter substrate-binding	3.30E-18
243	comp188287_g1_i4	plastidic atp adp transporter	0.00E+00
244	comp189716_g1_i6	DDAC_ENTAG	1.10E-07
245	comp171049_g1_i3	PTR36_ARATH	3.40E-32
246	comp175966_g1_i2	hypothetical protein SDRG_02977	4.70E-41
247	comp195928_g2_i6	VP13A_DICDI	1.60E-57
248	comp158911_g1_i2	KSHB_MYCS2	3.90E-19
249	comp188290_g1_i2	hypothetical protein GUITHDRAFT_81225	1.00E-31
250	comp178752_g1_i12	alpha subunit of photoactivated adenylyl cyclase	4.10E-23
251	comp188376_g1_i5	PREDICTED: uncharacterized protein LOC105847846	3.90E-31
252	comp190368_g1_i1	KAPR_YARLI	1.20E-12
253	comp174242_g1_i5	Transcription factor	2.70E-56
254	comp191215_g1_i3	poly(ADP-ribose) glycohydrolase-like	1.60E-50
255	comp192990_g1_i2	FYV1_DICDI	3.40E-66
256	comp192307_g1_i1	predicted protein	4.70E-17
257	comp164653_g1_i5	Hypothetical protein, putative	7.50E-12
258	comp190203_g1_i3	MTP9_ARATH	2.50E-37
259	comp193858_g1_i2	Y1181_ARATH	2.90E-121
260	comp168035_g1_i2	CY42_TRYBB	1.20E-13
261	comp171450_g1_i4	PSBD_EUGGR	1.70E-73
262	comp170275_g1_i1	Poly (A) RNA polymerase cid14	6.00E-51
263	comp175114_g1_i1	AIG2 family	1.90E-18
264	comp187148_g3_i2	ATP-binding cassette transporter,	6.20E-57
265	comp192542_g3_i5	tRNA-splicing ligase homolog	0.00E+00
266	comp187484_g3_i1	LUC7L_DICDI	2.50E-24
267	comp155632_g1_i2	NEK1_HUMAN	4.30E-50
268	comp193845_g1_i5	AZOR_DECAR	1.70E-85
269	comp163237_g1_i1	conserved unknown protein	3.00E-14
270	comp186473_g1_i5	PSP1_YEAST	2.60E-18
271	comp171458_g1_i1	zinc finger DHHC domain-containing	1.20E-19
272	comp191570_g2_i2	hypothetical protein PBRA_008009	3.90E-11
273	comp184275_g1_i4	hypothetical protein C922_02208	6.50E-07
274	comp184347_g1_i5	serine protease	8.30E-80
275	comp134876_g1_i4	hypothetical protein TRSC58_03537	7.40E-07
276	comp177684_g1_i4	CQ053_HUMANRecName: Full=Uncharacterized protein C17orf53	1.30E-09

277	comp161429_g1_i1	probable 1-acyl-sn-glycerol-3-phosphate acyltransferase 4	2.60E-53
278	comp156303_g1_i2	malate synthase-isocitrate lyase	1.60E-45
279	comp187790_g1_i5	PYR1_SQUAC	0.00E+00
280	comp191609_g1_i3	BRISC and BRCA1-A complex member 1-like	4.70E-14
281	comp185001_g2_i4	OTU family cysteine protease	2.20E-33
282	comp177739_g1_i1	adenylyl cyclase	1.70E-08
283	comp187834_g1_i7	EXD2_HUMAN	9.90E-36
284	comp187308_g1_i1	CID13_SCHPO	5.80E-13
285	comp178883_g1_i8	phosphoglycerate mutase family domain containing	7.90E-16
286	comp191205_g2_i9	hypothetical protein SPRG_15286	4.70E-08
287	comp176522_g2_i1	PSAB_EUGGR	0.00E+00
288	comp182143_g1_i1	RPM1_CAEEL	4.30E-37

(c) List of TF family of part B (19 transcripts of 12 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp182154_g2_i2	Whirly	1.10E-09
2	comp190140_g2_i6	LBD	8.60E-28
3	comp166780_g1_i2	HB-other	1.10E-29
4	comp181025_g1_i2	WRKY	1.60E-93
5	comp193939_g1_i2	YABBY	8.50E-30
6	comp183118_g2_i4	B3	4.30E-21
7	comp184981_g1_i5	MYB	2.20E-21
8	comp185491_g1_i8	NAC	1.10E-22
9	comp187072_g3_i2	NAC	1.30E-11
10	comp168168_g1_i3	HB-other	2.30E-09
11	comp183950_g1_i1	B3	3.10E-39
12	comp191738_g1_i2	GATA	4.40E-13
13	comp186437_g3_i1	NAC	3.80E-17
14	comp194229_g1_i3	YABBY	8.10E-44
15	comp181588_g1_i4	C2H2	4.50E-10
16	comp185491_g2_i3	MYB	1.60E-138
17	comp165764_g1_i1	Trihelix	2.00E-57
18	comp181025_g1_i5	WRKY	2.90E-93
19	comp191709_g1_i7	bZIP	7.20E-09

(d) List of other genes name of part B (53 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp171509_g2_i4	AAA family ATPase	5.60E-10
2	comp160291_g1_i2	SGT1A_ARATH	2.90E-40
3	comp182397_g2_i1	MET16_HUMAN	1.90E-46
4	comp182751_g1_i1	GAS8_CHLRE	3.50E-124
5	comp193296_g1_i1	heterogeneous nuclear ribonucleo H	1.10E-11
6	comp176502_g1_i2	predicted protein, partial	2.70E-54
7	comp195731_g1_i11	PAS domain-containing sensor histidine kinase	3.40E-20
8	comp164861_g1_i3	translation initiation factor eIF2 alpha subunit	5.10E-91
9	comp189361_g1_i5	cyclin-A2predicted protein, partial	6.60E-10
10	comp193659_g1_i2	NLRC3_HUMAN	2.50E-30
11	comp193927_g1_i2	wd40 repeat	3.90E-20
12	comp156508_g6_i1	YCX7_EUGGR	1.30E-20
13	comp187368_g1_i4	DBP2_ASHGO	9.40E-89
14	comp182705_g2_i2	GPKOW_XENLA	2.70E-13
15	comp190042_g1_i2	Phytanoyl- dioxygenase	1.90E-29
16	comp194383_g1_i1	AARA_DICDI	2.30E-08
17	comp189908_g1_i5	S35D2_BOVIN	5.40E-64
18	comp160061_g1_i1	polyketide cyclase	4.90E-26
19	comp180184_g1_i3	TIL_ARATH	3.40E-10
20	comp187325_g2_i4	glycosyl transferase	1.10E-17
21	comp194646_g2_i6	TRIM2_CALJA	9.10E-09
22	comp194006_g1_i9	cAMP specific phosphodiesterase	2.30E-35
23	comp181313_g1_i3	multiple RNA-binding domain-containing 1	5.30E-104
24	comp193527_g1_i1	UDP-sugar pyrophosphorylase 1	1.60E-118
25	comp171843_g1_i2	DNA glycosylase	1.10E-17
26	comp186361_g1_i1	KAPR_CAEEL	3.30E-11
27	comp169203_g1_i1	vacuolar sorting 9 (vps9) domain containing	6.10E-12
28	comp183772_g1_i6	hypothetical protein AMSG_06512	1.10E-10
29	comp191569_g2_i3	SGMC_DICDI	5.40E-29
30	comp185355_g2_i1	RUS3_ARATH	7.30E-25
31	comp157088_g1_i2	Y182_METEA	1.50E-17

32	comp188199_g3_i1	receptor-type adenylate cyclase GRESAG 4,	3.00E-47
33	comp188986_g1_i9	YH010_HUMAN	8.10E-35
34	comp195767_g2_i10	OTU7B_HUMAN	3.40E-23
35	comp187263_g1_i1	CYA1_MYCBO	2.90E-17
36	comp165761_g1_i2	mitochondrial paraflagellar rod (PFC16)	2.30E-35
37	comp191344_g1_i3	Lmbr1-like motif	6.30E-48
38	comp191053_g2_i8	TRPV6_HUMAN	3.30E-13
39	comp183232_g5_i2	ATCA3_ARATH	8.10E-25
40	comp181515_g1_i11	hypothetical protein	1.30E-07
41	comp182446_g1_i6	metallophosphoesterase domain-containing 1	4.90E-61
42	comp202551_g1_i1	RK20_EUGGR	5.60E-65
43	comp179644_g1_i6	intraflagellar transport IFT88	0.00E+00
44	comp157565_g1_i1	GST1_SCHPO	2.10E-08
45	comp180703_g1_i7	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3,	0.00E+00
46	comp185804_g1_i2	pbs lyase heat-like repeat	2.30E-10
47	comp174989_g1_i1	ZDH13_ARATH	2.90E-43
48	comp194470_g1_i1	hypothetical protein RirG_215870	1.20E-39
49	comp183218_g2_i2	glycosyl transferase family GT77 cell wall polysaccharide biosynthesis	5.40E-27
50	comp204489_g1_i1	MP20_DROME	1.50E-104
51	comp170790_g1_i2	uncharacterized endoplasmic reticulum membrane -like	5.10E-33
52	comp180165_g1_i3	DGDG1_SOYBN	8.40E-70
53	comp194000_g1_i1	kinesin motor family	5.40E-08

(e) List of TF family of part C (10 transcripts of 9 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp178919_g1_i2	CPP	1.20E-22
2	comp159227_g1_i1	bHLH	9.60E-32
3	comp145127_g1_i2	C3H	1.20E-08
4	comp187760_g1_i7	EIL	6.80E-13
5	comp162433_g1_i2	FAR1	2.70E-58
6	comp178433_g1_i5	Nin-like	1.20E-11
7	comp192560_g1_i3	WRKY	9.20E-31
8	comp156504_g1_i3	YABBY	9.20E-31
9	comp161701_g1_i6	GeBP	1.60E-25

10	comp159486_g1_i1	bHLH	1.50E-08

(f) List of other genes name of part C (54 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp166062_g1_i5	TM222_HUMAN	5.90E-30
2	comp189117_g1_i4	CBPC1_XENLA	1.70E-92
3	comp190920_g2_i4	hypothetical protein OCBIM_22033985mgpredicted protein, partial	2.20E-100
4	comp193531_g3_i2	hypothetical protein EUGRSUZ_E03521	3.90E-08
5	comp183741_g1_i5	calmodulin mutant syncam9	7.90E-09
6	comp170754_g2_i1	isoform a	7.20E-21
7	comp191532_g1_i8	NEDD4-like E3 ubiquitin- ligase WWP2	1.70E-45
8	comp190591_g1_i2	PARP domain-containing	8.60E-11
9	comp195319_g2_i1	membrane-bound adenylyl cyclase	4.30E-19
10	comp180836_g4_i7	hypothetical protein CHLNCDRAFT_145050	1.10E-55
11	comp172770_g8_i3	Endonuclease exonuclease phosphatase	9.60E-30
12	comp147074_g1_i1	hypothetical protein VOLCADRAFT_100878	7.40E-60
13	comp171920_g1_i2	transmembrane 5 isoform X2	2.60E-16
14	comp183103_g1_i1	Plastid lipid-associated fibrillin conserved domain	1.70E-32
15	comp189444_g2_i9	kinase domain	6.50E-12
16	comp179410_g1_i3	CI171_DANRERecName: Full=Uncharacterized protein C9orf171 homolog	1.30E-08
17	comp181401_g1_i1	regulator of chromosome condensation RCC1	1.10E-23
18	comp184575_g2_i3	membrane-associated protein, putative	4.10E-08
19	comp186766_g1_i4	Y2800_NOSS1	7.30E-14
20	comp190340_g2_i2	ELAV2_RAT	3.70E-21
21	comp195606_g3_i5	RIBD_CHLPN	1.40E-56
22	comp179782_g1_i6	VACHB_DANRE	5.30E-21
23	comp195049_g1_i4	CID11_SCHPO	1.70E-14
24	comp174393_g1_i1	hypothetical protein THAOC_31588	1.90E-60
25	comp195537_g2_i2	regulatory associated of mTOR	3.70E-156
26	comp193849_g2_i12	hypothetical protein	2.10E-20
27	comp187666_g2_i3	kinase domain	3.20E-17
28	comp186653_g2_i11	KPK2_ARATH	1.60E-34
29	comp193509_g1_i1	MFS transporter	4.50E-11
30	comp192546_g1_i4	membrane-associated ,	1.20E-11

31	comp184700_g1_i4	PBUO_BACSU	2.10E-51
32	comp183107_g1_i1	DPOA2_MOUSE	3.90E-27
33	comp193577_g1_i6	Phosphate transporter PHO1	1.80E-60
34	comp174749_g1_i2	PI42C_DANRE	5.60E-10
35	comp181483_g1_i1	TPA: RING zinc finger domain superfamily	2.30E-12
36	comp175387_g2_i5	hypothetical protein BA6E_12465	3.60E-13
37	comp167840_g1_i2	sphingomyelin synthase-related 1	3.10E-18
38	comp194103_g2_i5	CYNT_MYCS2	8.60E-26
39	comp194784_g1_i4	WD repeat-containing mio-A-like	6.00E-30
40	comp187529_g2_i1	ALIS3_ARATH	4.10E-56
41	comp165052_g1_i2	WAXS5_ARATH	3.60E-22
42	comp194318_g1_i4	CYAA_YEAST	5.50E-30
43	comp191406_g1_i7	SRFR1_ARATH	4.90E-09
44	comp191132_g1_i5	pantothenate kinase 1	4.40E-52
45	comp191923_g1_i7	STK11_CHICK	4.70E-53
46	comp179734_g2_i1	PIF1_XENLA	4.10E-108
47	comp196348_g2_i1	dynein heavy chain 7, axonemal-like	0.00E+00
48	comp189896_g1_i3	copper transporter family	2.20E-13
49	comp196152_g1_i5	Y8013_DICDI	6.80E-11
50	comp191053_g1_i1	transient receptor potential cation channel subfamily V member 6-like	1.10E-19
51	comp188600_g1_i5	TSC2_RAT	2.60E-11
52	comp182711_g1_i5	TRK2_YEAST	5.90E-38
53	comp174891_g2_i2	PABP_CANAL	8.20E-23
54	comp187028_g2_i7	phospho phosphatase	1.30E-12

(g) List of other genes name of part D (41 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp189196_g2_i8	methyl transferase	4.70E-29
2	comp175038_g1_i13	Tetratricopeptide-like helical	2.00E-46
3	comp140961_g1_i1	ATP synthase F0 subunit 6 (mitochondrion)	1.30E-18
4	comp190200_g1_i2	KCD17_HUMAN	3.20E-12
5	comp189150_g2_i5	IMCL1_ORYSJ	3.00E-58
6	comp161981_g1_i1	endo-1,3-beta-glucanase	1.20E-07
7	comp169913_g1_i6	PTBP3_ARATH	5.60E-08

8	comp193408_g2_i1	Y5520_ARATH	5.30E-147
9	comp181261_g4_i1	KAPR_YARLI	6.00E-14
10	comp193056_g1_i4	serine threonine- kinase STN7, chloroplastic	2.50E-94
11	comp178370_g1_i2	NLRC3_HUMAN	4.30E-33
12	comp185650_g1_i3	carnitine O-palmitoyltransferase 2, mitochondrial	4.30E-154
13	comp195238_g1_i7	OHK3_ORYSI	1.40E-28
14	comp191297_g3_i5	ABC transporter permease	7.20E-24
15	comp175256_g1_i1	aspartic protease	1.60E-13
16	comp165642_g1_i2	hypothetical protein GUITHDRAFT_149743	3.70E-20
17	comp193010_g1_i2	hypothetical protein H310_01051	4.50E-29
18	comp188468_g1_i7	TRPV4_MOUSE	2.60E-16
19	comp194028_g1_i4	ANM7_DANRE	1.80E-18
20	comp191127_g1_i2	CRYD_XENLA	1.80E-40
21	comp188879_g1_i10	PDE11_DROME	8.30E-74
22	comp185447_g1_i2	glycosyltransferase aer61	7.00E-49
23	comp178763_g2_i4	TTD14_DROME	2.10E-34
24	comp190042_g1_i3	Phytanoyl- dioxygenase	3.90E-22
25	comp170990_g1_i4	Y305_SYNY3	1.30E-29
26	comp173380_g1_i4	pumilio homolog 5	6.90E-68
27	comp190901_g1_i1	hypothetical protein MOQ_000545	5.70E-08
28	comp183042_g3_i1	hypothetical protein COCSUDRAFT_68410	1.20E-14
29	comp194695_g1_i4	TRI37_MOUSE	4.10E-89
30	comp187787_g1_i2	PAP6_ARATH	2.40E-19
31	comp193338_g1_i1	IDE_HUMAN	3.00E-72
32	comp181113_g1_i2	hypothetical protein US73_C0005G0005	1.20E-24
33	comp192620_g2_i5	tetratricopeptide repeat 28-like isoform X1	1.40E-34
34	comp184385_g1_i9	IF4E3_ORYSJ	3.00E-10
35	comp193855_g1_i3	kinase domain	2.40E-12
36	comp189992_g1_i3	ACAC_CHICK	0.00E+00
37	comp189538_g1_i5	PAS domain-containing sensor histidine kinase	2.00E-26
38	comp194950_g1_i4	ANM7_XENLA	2.20E-18
39	comp184299_g1_i3	ASAT1_ARATH	5.70E-07
40	comp189661_g2_i1	DYL3_CHLRE	1.90E-08
41	comp128383_g1_i1	NLRC3_HUMAN	7.90E-40

Appendix table 24 genes name of area 24

(a) List of TF family (264 transcripts of 35 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp195400_g1_i2	ARF	1.30E-25
2	comp186940_g1_i1	B3	6.80E-27
3	comp192565_g1_i3	BBR-BPC	8.10E-22
4	comp182076_g1_i6	bHLH	3.10E-23
5	comp182133_g1_i3	C3H	9.00E-08
6	comp181866_g2_i3	C3H	3.40E-34
7	comp186783_g1_i5	C3H	4.60E-13
8	comp181432_g1_i4	C3H	4.00E-17
9	comp182121_g1_i1	C3H	9.60E-17
10	comp182121_g1_i2	C3H	8.30E-17
11	comp182121_g1_i4	C3H	8.90E-17
12	comp183682_g1_i6	C3H	5.90E-18
13	comp182505_g1_i3	E2F_DP	4.10E-11
14	comp182083_g2_i7	ERF	1.10E-07
15	comp183261_g1_i1	ERF	9.90E-54
16	comp177893_g2_i2	FAR1	4.90E-41
17	comp187913_g2_i4	FAR1	2.00E-47
18	comp188798_g3_i3	FAR1	1.50E-08
19	comp188798_g3_i10	FAR1	1.20E-08
20	comp186291_g1_i7	G2-like	6.30E-14
21	comp186291_g2_i8	G2-like	5.60E-14
22	comp186291_g2_i9	G2-like	6.00E-14
23	comp187174_g1_i2	FAR1	1.20E-38
24	comp189253_g1_i4	NAC	5.70E-15
25	comp189253_g1_i5	NAC	5.90E-15
26	comp172834_g1_i4	HB-other	2.50E-09
27	comp186004_g1_i2	C3H	2.40E-09
28	comp194858_g2_i1	LBD	5.10E-13
29	comp192037_g1_i3	LBD	1.70E-17
30	comp191983_g1_i6	LBD	1.70E-17
31	comp194715_g3_i1	MIKC	4.60E-07
32	comp186587_g1_i1	MYB	2.50E-12

33	comp186587_g1_i2	MYB	2.60E-12
34	comp186587_g1_i4	MYB	2.60E-12
35	comp184756_g3_i4	MYB_related	2.00E-13
36	comp186775_g1_i1	NAC	5.20E-15
37	comp186775_g1_i3	NAC	7.90E-15
38	comp188829_g1_i5	NAC	3.20E-53
39	comp188829_g1_i14	NAC	2.40E-53
40	comp188829_g2_i3	NAC	4.40E-53
41	comp190129_g1_i2	NAC	3.10E-11
42	comp195449_g2_i4	NF-YA	5.50E-09
43	comp187898_g1_i2	NF-YA	6.80E-24
44	comp190350_g2_i4	Nin-like	1.20E-11
45	comp190536_g2_i10	Nin-like	1.40E-12
46	comp192062_g2_i3	Nin-like	2.30E-47
47	comp192221_g1_i2	Nin-like	4.10E-36
48	comp162763_g1_i2	CPP	2.00E-15
49	comp179590_g1_i3	CPP	5.30E-08
50	comp181360_g1_i1	YABBY	1.10E-62
51	comp181096_g1_i4	WRKY	4.70E-64
52	comp181096_g1_i7	WRKY	1.70E-64
53	comp183543_g1_i3	MYB_related	2.50E-45
54	comp184137_g1_i8	STAT	3.50E-07
55	comp178501_g1_i2	Trihelix	5.60E-12
56	comp194139_g1_i6	WRKY	1.20E-07
57	comp184207_g3_i1	WRKY	4.10E-09
58	comp184207_g3_i3	WRKY	4.00E-09
59	comp184845_g1_i7	WRKY	3.60E-28
60	comp195019_g2_i5	WRKY	2.90E-09
61	comp195019_g2_i8	WRKY	2.90E-09
62	comp184598_g1_i3	WRKY	1.60E-73
63	comp184598_g1_i7	WRKY	1.50E-73
64	comp182154_g2_i5	Whirly	1.30E-09
65	comp180869_g4_i2	YABBY	5.50E-19
66	comp180869_g4_i3	YABBY	5.70E-19
67	comp195382_g1_i1	bHLH	9.20E-43

68	comp183070_g1_i2	bHLH	3.90E-08
69	comp184505_g3_i3	bHLH	8.30E-08
70	comp184505_g3_i4	bHLH	8.60E-08
71	comp183992_g2_i7	bHLH	1.00E-40
72	comp182884_g1_i4	bHLH	2.30E-45
73	comp182647_g1_i1	bHLH	3.00E-07
74	comp184855_g1_i1	bHLH	1.90E-23
75	comp195081_g1_i3	bHLH	2.10E-14
76	comp184818_g1_i3	bHLH	9.30E-22
77	comp195400_g1_i1	ARF	1.80E-25
78	comp194796_g1_i4	ARR-B	1.60E-08
79	comp184451_g1_i1	B3	1.60E-41
80	comp184773_g1_i3	B3	4.20E-18
81	comp185372_g1_i2	B3	3.70E-08
82	comp192565_g1_i2	BBR-BPC	8.10E-22
83	comp182907_g1_i5	C2H2	6.00E-16
84	comp182133_g1_i2	C3H	8.90E-08
85	comp181285_g1_i7	C3H	4.70E-09
86	comp186806_g1_i1	C3H	1.30E-24
87	comp186249_g1_i4	C3H	5.50E-10
88	comp180925_g1_i2	C3H	2.00E-52
89	comp181653_g1_i2	C3H	5.70E-25
90	comp181180_g1_i1	C3H	1.70E-16
91	comp182464_g1_i1	C3H	3.70E-44
92	comp188601_g1_i1	CAMTA	1.30E-09
93	comp182505_g1_i2	E2F_DP	4.00E-11
94	comp187492_g1_i3	EIL	5.90E-13
95	comp186518_g1_i2	FAR1	2.20E-28
96	comp187913_g1_i3	FAR1	2.00E-47
97	comp187822_g1_i12	FAR1	5.50E-20
98	comp186291_g1_i4	G2-like	6.20E-14
99	comp186291_g2_i12	G2-like	6.00E-14
100	comp170576_g2_i1	GATA	3.20E-07
101	comp194735_g1_i4	GRAS	2.20E-38
102	comp194735_g1_i5	GRAS	1.00E-38

103	comp189976_g1_i5	GRAS	4.70E-44
104	comp190849_g1_i2	GRAS	2.90E-07
105	comp183838_g1_i4	GRAS	4.50E-13
106	comp187961_g3_i6	GeBP	4.20E-11
107	comp174103_g1_i1	HB-other	8.70E-09
108	comp171273_g1_i1	HB-other	1.90E-31
109	comp194858_g2_i3	LBD	6.90E-11
110	comp191983_g1_i2	LBD	1.70E-17
111	comp191983_g1_i8	LBD	1.80E-17
112	comp186337_g1_i4	MYB	1.00E-49
113	comp186587_g1_i3	MYB	2.70E-12
114	comp193068_g2_i3	MYB_related	6.80E-08
115	comp180388_g1_i3	MYB_related	3.50E-17
116	comp193139_g1_i3	MYB_related	1.20E-28
117	comp171133_g1_i1	MYB_related	9.60E-17
118	comp193062_g1_i16	MYB_related	1.20E-11
119	comp182513_g1_i2	MYB_related	5.60E-09
120	comp168249_g1_i5	MYB_related	6.30E-41
121	comp193026_g3_i5	MYB_related	1.20E-49
122	comp193026_g3_i7	MYB_related	6.40E-45
123	comp192343_g1_i1	MYB_related	5.30E-60
124	comp180984_g1_i3	MYB_related	1.10E-23
125	comp183543_g1_i2	MYB_related	4.60E-45
126	comp187811_g1_i1	NAC	1.50E-21
127	comp187811_g1_i6	NAC	1.60E-21
128	comp186956_g1_i2	NAC	3.20E-16
129	comp190266_g1_i4	NAC	3.90E-07
130	comp190266_g1_i5	NAC	3.70E-07
131	comp187475_g1_i1	NAC	3.90E-119
132	comp187475_g1_i3	NAC	2.00E-119
133	comp188987_g1_i1	NAC	2.80E-70
134	comp188987_g1_i2	NAC	3.00E-70
135	comp188987_g1_i3	NAC	2.90E-70
136	comp188987_g1_i5	NAC	3.60E-70
137	comp188987_g1_i6	NAC	5.00E-70

138	comp189733_g2_i2	NF-YA	1.40E-39
139	comp192838_g1_i6	Nin-like	3.30E-09
140	comp178501_g2_i4	Trihelix	5.60E-12
141	comp177668_g1_i1	Trihelix	3.10E-10
142	comp184207_g3_i4	WRKY	4.00E-09
143	comp184845_g1_i5	WRKY	3.90E-29
144	comp181096_g1_i5	WRKY	5.60E-64
145	comp195019_g2_i4	WRKY	2.90E-09
146	comp184598_g1_i4	WRKY	3.20E-30
147	comp185643_g1_i1	WRKY	5.70E-36
148	comp185188_g1_i1	WRKY	4.40E-36
149	comp185188_g1_i4	WRKY	4.70E-36
150	comp182240_g1_i2	WRKY	9.40E-76
151	comp194503_g2_i3	WRKY	5.20E-12
152	comp194139_g1_i7	WRKY	1.10E-07
153	comp181025_g1_i6	WRKY	2.90E-93
154	comp181025_g1_i7	WRKY	2.70E-93
155	comp183410_g1_i3	WRKY	1.40E-20
156	comp181589_g1_i7	YABBY	6.50E-28
157	comp181589_g1_i9	YABBY	1.00E-27
158	comp192990_g2_i5	bHLH	1.40E-60
159	comp192727_g1_i6	bHLH	4.20E-94
160	comp195640_g1_i5	bHLH	3.70E-24
161	comp195242_g1_i4	bHLH	6.80E-32
162	comp183004_g2_i2	bHLH	3.20E-12
163	comp185184_g2_i3	bHLH	5.00E-26
164	comp182567_g1_i1	bHLH	3.50E-12
165	comp182567_g1_i3	bHLH	3.70E-12
166	comp192693_g1_i3	bHLH	4.80E-28
167	comp196244_g1_i1	ARF	8.40E-25
168	comp184451_g1_i2	B3	2.70E-41
169	comp180967_g1_i3	BES1	1.00E-20
170	comp182123_g1_i6	bHLH	3.40E-23
171	comp182076_g1_i2	bHLH	2.70E-23
172	comp184100_g2_i3	C2H2	6.60E-08

173	comp182133_g1_i1	C3H	9.10E-08
174	comp185320_g1_i7	C3H	2.30E-28
175	comp185320_g1_i11	C3H	5.00E-29
176	comp181432_g1_i3	C3H	4.80E-17
177	comp187307_g1_i6	C3H	6.90E-18
178	comp187307_g1_i17	C3H	6.90E-18
179	comp182505_g1_i4	E2F_DP	4.30E-11
180	comp182083_g2_i1	ERF	1.30E-07
181	comp182083_g2_i9	ERF	3.50E-08
182	comp190570_g1_i2	NAC	6.10E-21
183	comp190570_g2_i7	NAC	5.90E-21
184	comp187307_g1_i2	C3H	3.10E-18
185	comp178027_g1_i2	FAR1	1.20E-20
186	comp178027_g1_i6	FAR1	1.10E-20
187	comp177893_g2_i1	FAR1	4.60E-41
188	comp188798_g3_i8	FAR1	1.30E-08
189	comp186291_g1_i8	G2-like	5.70E-14
190	comp186291_g2_i3	G2-like	6.10E-14
191	comp190849_g1_i1	GRAS	2.50E-07
192	comp179897_g1_i1	GRAS	9.00E-13
193	comp194799_g1_i1	GRAS	5.50E-40
194	comp189976_g1_i2	GRAS	2.00E-20
195	comp187592_g3_i4	GRAS	1.80E-18
196	comp172401_g1_i5	HB-other	2.90E-24
197	comp171202_g1_i1	HB-other	3.20E-24
198	comp171202_g1_i3	HB-other	4.20E-24
199	comp175469_g1_i1	HB-other	8.90E-09
200	comp169916_g1_i1	HB-other	9.90E-23
201	comp185474_g1_i4	C3H	2.30E-09
202	comp186004_g1_i1	C3H	2.50E-09
203	comp181285_g1_i4	C3H	1.70E-09
204	comp181285_g1_i6	C3H	3.80E-09
205	comp179368_g1_i2	FAR1	3.80E-42
206	comp180402_g1_i1	FAR1	3.30E-13
207	comp192037_g1_i5	LBD	1.80E-17

208	comp192037_g1_i7	LBD	1.90E-17
209	comp191983_g1_i4	LBD	1.70E-17
210	comp192107_g1_i6	LBD	3.70E-24
211	comp194691_g1_i1	M-type	2.90E-10
212	comp180388_g1_i2	MYB_related	4.50E-13
213	comp193139_g1_i1	MYB_related	1.20E-28
214	comp192858_g1_i4	MYB_related	5.70E-28
215	comp191094_g1_i1	C2H2	1.10E-07
216	comp182907_g1_i2	C2H2	6.30E-16
217	comp189567_g1_i2	NAC	4.20E-10
218	comp192182_g1_i6	NF-YA	8.00E-32
219	comp192182_g1_i7	NF-YA	2.50E-30
220	comp191708_g1_i3	Nin-like	4.40E-12
221	comp191970_g2_i2	Nin-like	1.60E-07
222	comp184417_g2_i1	C2H2	2.70E-21
223	comp162130_g1_i1	CPP	2.60E-15
224	comp181360_g1_i10	YABBY	1.60E-62
225	comp187492_g1_i9	EIL	6.00E-13
226	comp187822_g2_i2	FAR1	1.50E-19
227	comp184578_g1_i1	WRKY	3.10E-67
228	comp181118_g3_i1	YABBY	5.50E-50
229	comp195470_g1_i12	YABBY	7.20E-29
230	comp168249_g1_i6	MYB_related	7.80E-41
231	comp191775_g1_i6	MYB_related	1.00E-49
232	comp184774_g1_i1	MYB_related	1.40E-35
233	comp184774_g1_i7	MYB_related	8.80E-36
234	comp183543_g1_i1	MYB_related	2.40E-45
235	comp185507_g1_i6	GRAS	5.20E-13
236	comp184317_g1_i1	STAT	1.30E-19
237	comp189534_g1_i1	TALE	1.30E-37
238	comp188085_g2_i3	TALE	2.20E-08
239	comp146337_g1_i2	TCP	7.30E-114
240	comp184679_g1_i1	Trihelix	1.30E-07
241	comp185767_g3_i1	GeBP	4.00E-07
242	comp187961_g1_i3	GeBP	1.40E-10

243	comp187162_g1_i1	GeBP	1.90E-35
244	comp184087_g1_i1	VOZ	3.20E-24
245	comp181533_g1_i1	WRKY	1.90E-43
246	comp185188_g1_i2	WRKY	4.40E-36
247	comp185188_g1_i3	WRKY	9.20E-36
248	comp194747_g2_i2	WRKY	1.50E-57
249	comp186008_g2_i11	WRKY	6.50E-74
250	comp184598_g1_i5	WRKY	6.70E-74
251	comp184598_g1_i10	WRKY	9.40E-74
252	comp182637_g1_i3	WRKY	9.20E-30
253	comp183410_g1_i4	WRKY	1.40E-20
254	comp195586_g1_i2	WRKY	4.80E-26
255	comp195586_g1_i4	WRKY	1.90E-24
256	comp183070_g1_i1	bHLH	4.00E-08
257	comp196216_g1_i2	bHLH	6.10E-24
258	comp192844_g1_i1	bHLH	1.70E-50
259	comp195178_g1_i6	bHLH	4.50E-24
260	comp183087_g1_i2	bHLH	1.30E-08
261	comp183004_g2_i1	bHLH	3.00E-12
262	comp183530_g3_i2	bHLH	1.50E-18
263	comp182884_g1_i5	bHLH	2.20E-45
264	comp182669_g1_i1	bHLH	6.30E-19

(b) List of other genes name (738 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp191652_g3_i1	PDE1_CAEEL	5.10E-62
2	comp181865_g1_i4	hypothetical protein	6.70E-10
3	comp192179_g1_i5	membrane-bound adenylyl cyclase	3.80E-41
4	comp196159_g1_i1	ABC transporter substrate-binding	1.00E-20
5	comp186647_g1_i2	had-superfamily subfamily iia	6.00E-32
6	comp176840_g1_i2	KAD_DESMR	4.90E-109
7	comp149348_g1_i1	hypothetical protein GUITHDRAFT_166518	4.60E-24
8	comp194995_g1_i3	violaxanthin de- chloroplastic	7.10E-92
9	comp171840_g3_i5	TTL_YEAST	1.10E-11

10	comp144924_g1_i1	hypothetical protein NAEGRDRAFT_58984	3.00E-09
11	comp182126_g1_i5	RA51B_MOUSE	8.80E-53
12	comp186675_g1_i2	transmembrane ,	5.50E-08
13	comp190506_g2_i3	lipolytic G-D-S-L family	4.60E-11
14	comp189979_g5_i2	PLA15_ARATH	3.70E-10
15	comp188035_g1_i2	presequence protease 1, chloroplastic mitochondrial-like	0.00E+00
16	comp192399_g1_i18	Hypothetical protein, putative	2.80E-16
17	comp182293_g3_i4	kelch domain-containing 3	2.00E-22
18	comp172240_g2_i11	DPOE2_CHICK	3.80E-81
19	comp190138_g1_i1	EXO1_ARATH	4.60E-94
20	comp169791_g1_i1	membrane-bound adenylyl cyclase	2.80E-123
21	comp179505_g3_i2	uncharacterized oxidoreductase At4g09670-like	1.70E-59
22	comp188190_g1_i1	TBC domaincontaining	1.00E-20
23	comp195767_g2_i4	OTU7B_HUMAN	3.40E-23
24	comp189341_g4_i7	receptor-type adenylate	3.10E-10
25	comp186779_g2_i2	Y328_SYNY3	2.10E-26
26	comp185679_g1_i1	nicotinate-nucleotide adenylyltransferase	6.80E-52
27	comp190339_g1_i3	pilus assembly	5.40E-81
28	comp189536_g1_i1	hypothetical protein CAOG_002773	1.60E-07
29	comp192846_g1_i3	hypothetical protein A1Q2_02392	1.20E-07
30	comp176504_g3_i4	PREDICTED: uncharacterized protein LOC104613003	1.50E-18
31	comp188270_g1_i3	ENG2_SCHPO	8.50E-27
32	comp191889_g1_i8	YGL4_YEAST	1.40E-37
33	comp168732_g1_i1	DESI2_DANRE	2.80E-25
34	comp188869_g1_i2	ZDH11_MOUSE	3.60E-29
35	comp177914_g1_i1	LUL1_ARATH	3.30E-09
36	comp191132_g1_i1	PANK2_ORYSJ	3.60E-48
37	comp195012_g1_i2	p-type atpase (p-atpase) superfamily	1.00E-59
38	comp173930_g1_i2	DEK1_ORYSJ	2.10E-39
39	comp162419_g1_i3	cyclin-K	2.80E-16
40	comp176721_g1_i2	TTD14_DROME	8.70E-70
41	comp175529_g1_i3	Hel1p	2.20E-54
42	comp195282_g3_i2	Unc104-like kinesin	7.60E-24
43	comp153609_g1_i1	palmitoyltransferase ZDHHC15 isoform X2	1.80E-28
44	comp189930_g1_i5	adenylate cyclase	1.90E-20

45	comp176834_g1_i4	WDR20_HUMAN	1.90E-42
46	comp189563_g1_i3	CAS_ARATH	4.10E-15
47	comp166620_g1_i2	predicted protein	4.20E-19
48	comp190018_g1_i1	tubby-related 3 isoform X1	3.50E-14
49	comp178741_g1_i2	PA1B3_BOVIN	1.70E-21
50	comp166235_g1_i2	RTPR_EUGGR	1.10E-18
51	comp194170_g2_i3	sialate O-acetylesterase	3.40E-19
52	comp190348_g1_i2	AL4A1_DICDI	0.00E+00
53	comp184665_g2_i4	vesicle-associated membrane	5.00E-45
54	comp192856_g3_i2	IF4E3_DANRE	3.20E-12
55	comp194462_g1_i28	Transcription factor	9.00E-58
56	comp193197_g1_i2	guanylate cyclase	1.00E-65
57	comp192233_g1_i8	class II glutamine amidotransferase	1.00E-95
58	comp195428_g1_i6	MTLD_CROS8	4.40E-18
59	comp189528_g1_i2	Mir1p	3.90E-28
60	comp174028_g1_i3	ankyrin	8.70E-16
61	comp194648_g1_i10	mitochondrial fission process 1	1.80E-11
62	comp185795_g1_i7	expressed unknown protein	7.80E-29
63	comp195263_g1_i2	hybrid sensor histidine kinase response regulator	3.40E-18
64	comp183701_g2_i4	phosphatidate phosphatase ppapdc1b	5.90E-19
65	comp180836_g3_i5	hypothetical protein CHLNCDRAFT_145050	1.20E-55
66	comp20849_g1_i2	SPRY domain-containing SOCS box	1.50E-65
67	comp191819_g1_i2	PREDICTED: uncharacterized protein LOC103330366	2.10E-08
68	comp190368_g1_i8	cyclic nucleotide-binding	1.70E-19
69	comp188490_g1_i4	polyadenylate-binding	1.10E-18
70	comp188461_g1_i10	pre-mRNA-processing 40A	2.10E-08
71	comp192688_g3_i2	GSTL1_ARATH	1.20E-14
72	comp150399_g1_i4	YCX9_EUGGR	0.00E+00
73	comp171176_g1_i1	XRN4_ARATH	1.40E-73
74	comp184648_g4_i3	MAF1_SCHPO	3.50E-21
75	comp192403_g1_i2	PBL9F_EUGGR	2.20E-13
76	comp179921_g1_i2	LAAT1_CAEEL	6.00E-22
77	comp195334_g1_i1	phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2predicted protein	8.10E-67
78	comp182889_g2_i2	unknown protein	1.30E-17
79	comp178977_g1_i2	RRAGA_HUMAN	6.00E-150

80	comp187210_g2_i2	NEK kinase	8.20E-64
81	comp195900_g1_i3	PCYAA_EUGLO	4.80E-17
82	comp127343_g1_i2	CTR1_SOLIN	3.80E-51
83	comp182483_g2_i3	membrane	2.00E-49
84	comp184349_g5_i3	hypothetical protein GUITHDRAFT_105708	3.30E-10
85	comp190433_g1_i1	calcium-binding CML	1.20E-77
86	comp179505_g3_i1	Y4967_ARATH	9.30E-42
87	comp180703_g1_i6	DED1_PHANO	0.00E+00
88	comp195431_g4_i4	Y1796_ARATH	9.20E-83
89	comp190305_g1_i3	extended synaptotagmin-1	6.70E-10
90	comp181483_g3_i6	TPA: RING zinc finger domain superfamily	3.80E-12
91	comp161194_g1_i2	SSRP1_XENLA	8.10E-45
92	comp188819_g1_i3	spermatogenesis-associated 4	1.70E-27
93	comp162660_g1_i3	DBP5_LODEL	2.20E-48
94	comp193677_g1_i3	YIS7_YEAST	1.90E-108
95	comp178713_g2_i2	T9SS C-terminal target domain-containing	5.50E-17
96	comp183363_g1_i1	hypothetical protein PBRA_000501	1.60E-15
97	comp191609_g1_i4	BRISC and BRCA1-A complex member 1-like	3.00E-14
98	comp188468_g1_i3	TRPV5_MOUSE	1.40E-14
99	comp161120_g1_i1	e3 ubiquitin- ligase	4.80E-16
100	comp192786_g2_i2	PAS domain-containing sensor histidine kinase	2.70E-13
101	comp187601_g1_i5	CSD_METTH	1.00E-40
102	comp163086_g1_i1	predicted protein	2.30E-08
103	comp165267_g1_i1	cyclopropane-fatty-acyl-phospholipid synthase	2.20E-80
104	comp190629_g2_i3	ankyrin repeat	6.90E-29
105	comp191183_g1_i3	FAD-dependent oxidoreductase	2.10E-47
106	comp188892_g1_i1	SRFR1_ARATH	9.10E-09
107	comp193751_g1_i4	TMM43_RAT	2.50E-22
108	comp193480_g1_i2	hypothetical protein NAEGRDRAFT_78864	1.80E-10
109	comp195579_g1_i7	hypothetical protein PBRA_002490	1.50E-26
110	comp195073_g2_i7	hypothetical protein AGDE_02734	2.00E-10
111	comp192852_g1_i2	RBBP6_HUMAN	3.10E-21
112	comp179810_g1_i1	ZC21C_MOUSE	2.40E-11
113	comp188445_g2_i1	glycosyltransferases (COG1216)	3.10E-22
114	comp182809_g2_i1	RCA_VIGRR	3.10E-85

115	comp154886_g1_i3	#NAME?	1.40E-125
116	comp172899_g2_i4	FBT2_ARATH	7.10E-08
117	comp188246_g1_i1	udp-n-acetylglucosamine--peptide n-acetylglucosaminyltransferase	1.70E-29
118	comp194882_g1_i3	SHEP_DROSI	9.90E-11
119	comp182231_g2_i2	PLDA1_ORYSJ	7.20E-179
120	comp189630_g1_i3	P2C10_ORYSJ	8.60E-41
121	comp168113_g1_i3	cAMP-dependent kinase type I regulatory subunit isoform X2	6.00E-77
122	comp177917_g1_i3	M2OM_DICDI	2.30E-50
123	comp181380_g1_i2	DCUP_HUMAN	5.50E-110
124	comp178833_g1_i1	Tetratricopeptide repeat domain	8.40E-19
125	comp177380_g1_i1	Mitochondrial transcription termination factor family isoform 2	8.20E-11
126	comp193197_g1_i5	guanylate cyclase	6.20E-64
127	comp179644_g1_i2	IFT88_MOUSE	0.00E+00
128	comp191596_g1_i2	NLTP_RABIT	6.20E-19
129	comp182924_g1_i2	arylsulfate sulfotransferase	1.20E-08
130	comp138431_g1_i1	PSBB_EUGGR	0.00E+00
131	comp173087_g1_i2	OFUT2_DROME	1.80E-09
132	comp195006_g1_i3	Y8013_DICDI	3.10E-10
133	comp194662_g1_i5	ITIH3_RABIT	6.00E-07
134	comp191205_g2_i10	hypothetical protein SPRG_15286	1.10E-07
135	comp167764_g1_i2	Uvr REP helicase	8.80E-30
136	comp195702_g1_i1	hypothetical protein	9.60E-22
137	comp184016_g1_i7	hypothetical protein CHLNCDRAFT_53531	2.30E-10
138	comp194117_g1_i4	hypothetical protein CHLNCDRAFT_53531	2.30E-11
139	comp170391_g4_i3	MSRA_RHOCS	6.50E-84
140	comp170803_g2_i2	PUT1_ORYSJ	1.10E-132
141	comp173239_g1_i8	beta-type IP39	1.60E-14
142	comp190162_g1_i8	predicted protein, partial	2.40E-38
143	comp172231_g1_i2	AXEA1_PRER2	1.10E-18
144	comp184152_g3_i3	ankyrin repeat domain-containing 50	5.30E-15
145	comp182065_g1_i3	DUF21 domain-containing At5g52790-like	4.40E-102
146	comp185495_g1_i4	PEPD_LACSK	1.40E-24
147	comp181532_g1_i2	LANC2_MOUSE	1.10E-55
148	comp169502_g1_i3	transporter permease	1.00E-28
149	comp175494_g1_i5	RFA1_HUMAN	4.30E-91

150	comp164733_g1_i1	PTPS_DROME	3.80E-29
151	comp190355_g1_i5	PAS domain-containing sensor histidine kinase	4.10E-13
152	comp191016_g1_i2	hypothetical protein EMIHUDRAFT_113418predicted protein, partial	1.20E-08
153	comp195305_g1_i1	FKB62_ARATH	6.20E-19
154	comp184859_g1_i6	glycosyltransferase sugar-binding region containing DXD motif	4.30E-29
155	comp156363_g1_i2	hypothetical protein GUITHDRAFT_166346	5.80E-73
156	comp175835_g1_i1	palmitoleoyl- carboxylesterase NOTUM	6.20E-32
157	comp172315_g1_i6	kinase domain	9.00E-14
158	comp184137_g1_i3	GGPPS_MUCCL	2.50E-35
159	comp195261_g1_i9	ADCY1_HUMAN	2.60E-12
160	comp185999_g4_i1	probable phosphatase 2C 8 isoform X1	2.60E-18
161	comp174477_g1_i1	predicted protein	2.80E-44
162	comp193585_g2_i3	TRA2A_MOUSE	2.00E-07
163	comp196140_g1_i2	NLRC3_HUMAN	2.20E-22
164	comp172899_g2_i12	FBT2_ARATH	7.20E-08
165	comp192446_g1_i4	alpha-L-rhamnosidase	2.20E-63
166	comp190428_g4_i6	PLCD4_XENLA	2.60E-74
167	comp192048_g1_i5	myosin heavy chain kinase	9.00E-41
168	comp175372_g1_i4	SR140_HUMAN	5.50E-19
169	comp180775_g1_i2	PFR1_TRYBB	6.60E-14
170	comp182543_g1_i3	DSF1_YEAST	4.20E-107
171	comp190716_g1_i8	AK1_DICDI	2.90E-40
172	comp186529_g1_i1	CHO2_ASPFU	4.00E-101
173	comp168860_g1_i5	dnaJ homolog subfamily B member 12	4.40E-34
174	comp187503_g1_i4	erythroid differentiation-related factor 1	1.90E-33
175	comp183887_g1_i9	hypothetical protein EMIHUDRAFT_197272	3.90E-16
176	comp186072_g2_i3	RA51C_CRIGR	3.60E-23
177	comp194417_g1_i14	poly(ADP-ribose) polymerase	3.70E-10
178	comp164626_g1_i2	GDIR_ARATH	2.10E-08
179	comp181497_g1_i7	ubiquinone biosynthesis methyltransferase	2.70E-14
180	comp177612_g1_i4	SRPR_DROME	3.10E-114
181	comp193902_g1_i3	TTCA_GEOSL	1.60E-32
182	comp196030_g1_i5	EXPR_XANCP	2.50E-45
183	comp193564_g1_i7	predicted protein, partial	2.60E-55
184	comp179995_g1_i3	predicted protein, partial	0.00E+00

185	comp171544_g2_i1	hypothetical protein SPRG_19192	2.10E-10
186	comp173615_g1_i2	GNPAT_BOVIN	5.80E-47
187	comp196033_g3_i1	MYND finger domain	1.50E-59
188	comp171755_g1_i2	lysine decarboxylase domain-containing	1.10E-100
189	comp182271_g1_i7	ABC transporter ATP-binding	4.00E-71
190	comp187915_g1_i5	MYCE_MICGR	3.90E-12
191	comp179316_g1_i2	TTL_BOVIN	2.20E-45
192	comp195557_g1_i4	hypothetical protein H257_09746	3.50E-08
193	comp154563_g1_i2	hypothetical protein GUITHDRAFT_109974	6.30E-46
194	comp194220_g2_i6	ef hand family	6.80E-32
195	comp143711_g1_i2	ANT1_YEAST	1.30E-21
196	comp185742_g1_i1	phosphoribulokinase uridine kinase family	2.40E-72
197	comp177876_g1_i3	PRP31_DANRE	1.50E-63
198	comp186300_g2_i6	sex-lethal homolog isoform X4	1.20E-17
199	comp168245_g1_i3	SFC1_ARATH	1.10E-53
200	comp174680_g1_i1	PAXI_CHICK	4.20E-11
201	comp163584_g1_i3	zinc finger predicted protein, partial	6.50E-36
202	comp193424_g1_i11	transporter	7.80E-38
203	comp188307_g1_i5	SDRD_STAAW	2.30E-09
204	comp189896_g1_i11	copper transporter family	4.60E-15
205	comp192614_g1_i3	CLPAA_SOLLC	0.00E+00
206	comp179762_g1_i4	eukaryotic translation initiation factor 4E-like isoform X4	4.00E-18
207	comp188566_g1_i1	MYCF_MICGR	1.80E-33
208	comp180254_g1_i3	COQ4_XENLA	5.30E-64
209	comp166800_g1_i6	Vesicle transport SFT2B	1.20E-08
210	comp184539_g1_i6	RNA binding	4.90E-90
211	comp194898_g1_i1	hypothetical protein ACA1_171730	1.30E-10
212	comp194917_g2_i4	F184A_HUMAN	1.40E-11
213	comp171192_g1_i1	RNA-binding family isoform 1	2.50E-14
214	comp184340_g1_i2	IF2C_EUGGR	2.60E-39
215	comp186014_g1_i7	6GAL_HYPRU	2.70E-30
216	comp185222_g1_i3	glycosyl transferase family 2	1.10E-16
217	comp178116_g1_i2	SAM-dependent methyltransferase	2.90E-76
218	comp194957_g1_i5	ATP-dependent DNA helicase Q4	4.30E-78
219	comp185151_g1_i2	THLA_PEPD6	8.80E-102

220	comp194004_g1_i9	adenylate cyclase,	2.40E-31
221	comp191897_g1_i1	YGL4_YEAST	1.00E-37
222	comp195163_g1_i9	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07
223	comp180775_g1_i3	PFR1_TRYBB	4.50E-14
224	comp189792_g1_i2	TRM61_HUMAN	1.80E-49
225	comp164216_g1_i1	CAH1_CHIHA	1.20E-29
226	comp182491_g2_i7	QORL1_BOVIN	9.80E-14
227	comp180109_g2_i3	ankyrin repeat-containing ,	5.60E-24
228	comp180662_g1_i1	hypothetical protein SPRG_14866	2.40E-07
229	comp174844_g1_i2	Plastidic atp adp transporter	7.20E-135
230	comp194362_g2_i3	transmembrane	1.80E-13
231	comp174989_g1_i7	ZDH13_ARATH	6.90E-43
232	comp191805_g1_i1	RSPH1_MOUSE	6.00E-14
233	comp195567_g2_i6	VMPA2_LOXIN	8.50E-24
234	comp175561_g1_i5	hypothetical protein	2.10E-27
235	comp150399_g1_i5	YCX9_EUGGR	0.00E+00
236	comp191823_g1_i4	tectonin beta-propeller repeat-containing 1-like	4.10E-10
237	comp166967_g1_i2	GP157_MOUSE	2.50E-08
238	comp178171_g1_i3	Neurotrypsin	3.10E-07
239	comp164733_g1_i6	PTPS_DROME	7.00E-31
240	comp177910_g1_i13	hypothetical protein PHALS_08252	1.50E-08
241	comp187283_g6_i12	YDJX_ECOLI	5.10E-07
242	comp182279_g1_i1	BABA1_RAT	1.60E-11
243	comp175936_g1_i6	IF4E_CANGA	1.10E-14
244	comp192589_g1_i7	phosphatidylinositol-4-phosphate 5-kinase	4.60E-19
245	comp192456_g3_i5	CORA_THEMA	5.30E-28
246	comp180162_g1_i2	hypothetical protein Esi_0020_0151	1.30E-40
247	comp175049_g1_i4	family magnesium ion transporter	2.50E-25
248	comp164734_g1_i1	hypothetical protein TRIADDRAFT_57460	9.40E-11
249	comp184721_g3_i1	MGAT3_RAT	7.00E-21
250	comp187053_g1_i1	6-phosphofructo-2-kinase fructose-2,6-bisphosphatase 4-like isoform X1	1.20E-58
251	comp187065_g1_i8	eukaryotic elongation factor 2 kinase-like	5.50E-21
252	comp170841_g1_i5	TRAM LAG1 CLN8 homology domain	2.10E-17
253	comp165294_g1_i2	EKI_ARATH	5.40E-50
254	comp181126_g1_i1	YRHG_BACSU	5.00E-18

255	comp181865_g1_i3	hypothetical protein	6.90E-10
256	comp173020_g1_i1	WDR20_HUMAN	1.70E-42
257	comp189948_g1_i4	outer membrane-stress sensor serine endopeptidase	3.80E-08
258	comp183382_g1_i3	phosphoglycerate mutase family domain containing	7.40E-16
259	comp190793_g2_i3	glycoside hydrolase	1.90E-86
260	comp188721_g1_i3	DIOXL_ARATH	5.10E-44
261	comp171129_g1_i2	mitochondrial carrier	1.20E-45
262	comp194529_g1_i5	ankyrin repeat	8.70E-33
263	comp190378_g2_i2	TANC1_RAT	7.50E-14
264	comp184685_g1_i7	YIS7_YEAST	4.20E-78
265	comp195793_g1_i1	Atrial natriuretic peptide receptor 1predicted protein, partial	2.90E-11
266	comp172763_g1_i1	Hypothetical protein, putative	3.70E-23
267	comp174564_g1_i4	CCDC6_HUMAN	8.50E-18
268	comp180333_g1_i4	MF6LB_XENLA	9.20E-10
269	comp191905_g1_i3	CLPC1_ARATH	0.00E+00
270	comp165657_g1_i4	cysteine synthase	4.80E-66
271	comp184386_g1_i1	Hypothetical protein, putative	8.30E-81
272	comp181894_g1_i1	ubiquitin conjugation factor E4 B	2.70E-116
273	comp186861_g1_i5	PTBP2_HUMAN	8.30E-68
274	comp194098_g2_i1	RL40_BRARP	7.50E-27
275	comp195290_g2_i4	photosystem I assembly Ycf3	6.30E-10
276	comp185654_g4_i6	ubiquitin carboxyl-terminal hydrolase 49	1.70E-43
277	comp190512_g1_i3	Y8236_DICDI	2.60E-99
278	comp180501_g1_i1	hypothetical protein PBRA_000501	5.80E-19
279	comp188053_g2_i4	asteroid homolog 1-like	3.80E-13
280	comp190216_g2_i4	predicted protein	4.10E-18
281	comp159846_g1_i2	ATE1_ARATH	6.50E-45
282	comp175063_g1_i1	ARF_AJECA	4.50E-12
283	comp176506_g1_i3	VIS_VIBS1	2.70E-09
284	comp186829_g2_i1	jumonji domain-containing	7.50E-24
285	comp173828_g1_i7	SIAH1_CAEBR	4.60E-08
286	comp178426_g1_i14	KAPS_SCHPO	5.90E-85
287	comp180552_g1_i3	THIM_CHLAD	3.30E-53
288	comp184614_g1_i5	CLPAA_SOLLC	0.00E+00
289	comp193533_g1_i1	S phase cyclin A-associated in the endoplasmic reticulum-like	1.20E-22

290	comp186447_g3_i2	HSOP3_ARATH	9.70E-11
291	comp185604_g1_i1	PUMP4_ARATH	1.10E-56
292	comp190418_g2_i5	CYA1_RHIME	1.70E-18
293	comp181333_g2_i4	IM30_ORYSJ	1.40E-53
294	comp195139_g1_i1	TOR_DICDI	0.00E+00
295	comp184304_g1_i2	adenylate guanylate cyclase	5.40E-47
296	comp189589_g1_i3	TAS_ECOLI	2.10E-59
297	comp178369_g1_i4	YRHG_BACSU	2.90E-12
298	comp153362_g1_i3	VIS_VIBS1	2.00E-07
299	comp188956_g1_i2	vacuolar sorting vps16	6.00E-170
300	comp193453_g5_i2	CYA1_MYCBO	7.00E-29
301	comp174141_g1_i4	hypothetical protein GUITHDRAFT_164479	7.00E-31
302	comp180295_g1_i5	MNTH_RHILO	6.90E-129
303	comp188092_g1_i1	predicted protein	3.10E-73
304	comp190381_g7_i1	BETC_RHIME	3.60E-17
305	comp178929_g1_i2	doublecortin protein	4.00E-11
306	comp189677_g1_i4	PCYAA_EUGLO	1.20E-163
307	comp172938_g1_i2	Y233_STRP6RecName: Full=Uncharacterized protein M6_Spy0233	1.60E-09
308	comp149586_g1_i4	AB1F_ARATH	0.00E+00
309	comp183396_g1_i2	phospholipid:diacylglycerol acyltransferase	5.30E-136
310	comp188300_g1_i1	GPX1_SYNY3	1.10E-49
311	comp176403_g1_i2	KELC_DROME	4.60E-11
312	comp183431_g4_i1	Hypothetical protein, putative	2.20E-18
313	comp180948_g1_i2	Wd-repeat ,	3.00E-51
314	comp188870_g2_i3	SNF1-related kinase catalytic subunit alpha KIN10-like isoform X2	1.80E-136
315	comp186829_g1_i4	jumonji domain-containing	4.90E-41
316	comp175533_g1_i7	S35F5_PONAB	3.70E-42
317	comp195811_g1_i5	Eukaryotic translation initiation factor 4E	8.60E-11
318	comp190512_g1_i5	tetratricopeptide repeat domain containing	5.10E-127
319	comp173088_g1_i3	solute carrier family 35 member F1-like	1.50E-44
320	comp195831_g1_i2	BGAL_PSYIN	7.40E-170
321	comp191160_g1_i3	ANK3_HUMAN	4.30E-15
322	comp115089_g1_i2	kynurenine--oxoglutarate transaminase 1 isoform X1	3.30E-123
323	comp180251_g1_i7	myosin heavy chain kinase	8.00E-42
324	comp179536_g1_i1	hypothetical protein PBRA_008002	8.80E-90

325	comp178602_g1_i3	TIGR02757 family	5.80E-49
326	comp175825_g1_i2	Triacylglycerol lipase precursor	2.80E-07
327	comp174403_g1_i2	ATXA_LEIDO	5.10E-39
328	comp194495_g1_i2	MHCKB_DICDI	6.00E-25
329	comp179644_g1_i5	IFT88_MOUSE	0.00E+00
330	comp187449_g1_i2	translation initiation factor IF-2	0.00E+00
331	comp193298_g1_i2	actin cross-linking	2.00E-12
332	comp171840_g3_i3	TTL_SCHPO	9.50E-09
333	comp179077_g1_i1	PCAT1_HUMAN	1.70E-40
334	comp177551_g1_i2	ASHR1_ARATH	1.40E-26
335	comp185679_g1_i5	nicotinate-nucleotide adenylyltransferase	5.70E-52
336	comp163001_g1_i2	Tellurite resistance methyltransferase, , core	1.70E-18
337	comp191202_g1_i5	ECHP_DANRE	7.60E-148
338	comp187915_g1_i1	MYCE_MICGR	5.70E-12
339	comp170803_g1_i1	PUT1_ORYSJ	9.90E-133
340	comp193236_g1_i5	HECD1_MOUSE	5.70E-08
341	comp168174_g1_i1	PPA5_PIG	7.20E-46
342	comp175643_g1_i3	2,3-diketo-5-methylthiopentyl-1-phosphate enolase	7.80E-98
343	comp175981_g1_i1	TMPS9_RAT	5.30E-16
344	comp178262_g1_i14	HPXO_KLEP7	2.80E-11
345	comp171649_g2_i2	PREDICTED: uncharacterized protein LOC106152497	1.40E-31
346	comp177981_g1_i3	5 -3 exoribonuclease 3-like	2.30E-83
347	comp188621_g1_i13	copper transporter family	2.20E-13
348	comp184170_g1_i2	mannosyl-oligosaccharide 1,2-alpha-mannosidase	1.60E-128
349	comp170616_g1_i1	TTL_SCHPO	4.10E-41
350	comp184816_g1_i4	Y8013_DICDI	4.80E-11
351	comp180787_g1_i1	SIAE_HUMAN	2.90E-18
352	comp179447_g2_i4	CFA58_CHLRE	1.90E-66
353	comp174929_g1_i5	replication factor-A 1 (rpa1)	6.20E-97
354	comp191477_g8_i4	E3 ubiquitin- ligase sina	2.10E-10
355	comp190945_g1_i1	CAH7_MOUSE	3.90E-25
356	comp195141_g1_i4	histidine kinase	2.70E-20
357	comp184665_g3_i8	phosphoglycerate mutase	6.60E-27
358	comp194354_g3_i2	PSME4_MOUSE	3.80E-55
359	comp196286_g3_i8	TCPR1_CHICK	5.00E-07

360	comp159163_g1_i1	AIR9_ARATH	3.80E-14
361	comp194189_g1_i5	5 -nucleotidase	1.20E-110
362	comp187470_g2_i4	Bdf1p	5.40E-15
363	comp179060_g1_i1	MGDP1_MOUSE	1.50E-14
364	comp175533_g1_i6	S35F5_PONAB	2.50E-42
365	comp177425_g1_i4	NLRC3_HUMAN	7.10E-15
366	comp192838_g1_i4	hypothetical protein L914_01153	1.60E-50
367	comp179149_g2_i2	unnamed protein product	5.00E-09
368	comp158064_g1_i2	inner membrane PPF-1, chloroplastic	6.10E-42
369	comp194415_g1_i2	BIG2_MOUSE	1.30E-08
370	comp196016_g1_i14	MELK_DANRE	5.90E-10
371	comp169706_g1_i1	tryptophan aminotransferase-related 4-like	3.70E-73
372	comp179066_g1_i4	CYSP3_SOLLC	4.00E-103
373	comp187834_g1_i8	hypothetical protein VOLCADRAFT_104579	2.00E-08
374	comp191338_g1_i4	NLRC3_HUMAN	5.60E-59
375	comp194710_g1_i3	MTFP1_CAEEL	7.80E-11
376	comp187755_g1_i4	hypothetical protein CHLREDRAFT_206193	8.90E-07
377	comp189855_g3_i4	gonidia-specific KA_k47	1.50E-19
378	comp181464_g1_i2	Poly [ADP-ribose] polymerase 1	1.60E-10
379	comp171612_g1_i1	AVT5_SCHPO	5.00E-55
380	comp187059_g1_i8	DMC1_MOUSE	4.40E-12
381	comp163260_g1_i1	ATG12_ASPCL	3.50E-19
382	comp174343_g1_i1	replication factor C, subunit 1	1.60E-110
383	comp187056_g1_i2	NMRL1_CHICK	4.40E-63
384	comp191861_g1_i8	tubby-related 3	5.80E-14
385	comp177331_g1_i1	Y3433_MYCBORecName: Full=Uncharacterized protein Mb3433	5.70E-09
386	comp179210_g1_i1	Tumor susceptibility gene 101	7.90E-29
387	comp196175_g1_i2	HAMP domain	1.90E-12
388	comp168108_g1_i1	steroid isomerase	7.60E-35
389	comp171450_g1_i11	PSBD_EUGGR	6.30E-98
390	comp181541_g1_i2	tyrosine kinase 9 actin monomer-binding ,	9.00E-27
391	comp190501_g1_i5	chloroplast fructose-1,6-bisphosphatase	1.80E-174
392	comp186187_g1_i11	CHO2_ASPFU	2.70E-101
393	comp32960_g1_i1	RL44B_YEAST	2.00E-32
394	comp172899_g2_i11	folate pteridine transporter	1.00E-15

395	comp175146_g1_i1	Patatin like	2.00E-38
396	comp174805_g1_i1	hypothetical protein	3.90E-07
397	comp193875_g2_i12	CCNA2_MESAU	4.90E-11
398	comp185518_g1_i5	desumoylating isopeptidase 2	3.10E-33
399	comp184327_g5_i2	GNPAT_BOVIN	2.50E-60
400	comp169258_g1_i2	Dpy-30 motif	1.70E-44
401	comp189573_g1_i6	regulator of chromosome condensation RCC1	1.80E-34
402	comp188566_g1_i7	MYCF_MICGR	1.30E-33
403	comp179504_g2_i7	tankyrase isoform X2	2.90E-12
404	comp186861_g1_i3	hnRNPL PTB hephaestus splicing factor subfamily	1.30E-74
405	comp173139_g1_i2	JANA_DROPS	8.90E-15
406	comp184289_g1_i2	DIT1_SPIOL	8.80E-143
407	comp177296_g1_i2	oligoribonuclease, mitochondrial	1.70E-72
408	comp188286_g1_i7	PK2_DICDI	1.30E-90
409	comp191032_g1_i1	predicted protein	1.00E-07
410	comp116569_g1_i1	ZRAB2_CHICK	1.70E-09
411	comp187480_g3_i1	LIP_RHIMI	5.90E-32
412	comp189217_g3_i5	CY43_TRYBB	6.30E-20
413	comp192830_g1_i2	MNN4_YEAST	3.60E-09
414	comp175486_g1_i7	PYRG_DICDI	2.20E-161
415	comp190830_g1_i7	MKT1_YEAST	2.40E-11
416	comp170693_g5_i2	PP1R8_BOVIN	1.10E-21
417	comp195141_g1_i1	histidine kinase	2.70E-20
418	comp184026_g1_i4	DHKL_DICDI	2.20E-12
419	comp193496_g2_i1	CHI11_METAN	7.70E-09
420	comp195349_g1_i4	sugar transporter SWEET1 isoform X1	3.00E-16
421	comp166728_g1_i4	HSOP3_ARATH	7.00E-24
422	comp190019_g2_i7	CYAA_TRYEQ	1.60E-29
423	comp191410_g1_i4	hypothetical protein A1Q2_02392	2.20E-07
424	comp182012_g1_i5	type III domain and Immunoglobulin-like fold domain-containing	5.20E-08
425	comp190949_g1_i3	AARA_DICDI	4.90E-07
426	comp193057_g1_i4	serine threonine- phosphatase 6 regulatory ankyrin repeat subunit b	7.60E-18
427	comp161443_g1_i1	dihydropyrimidinase	0.00E+00
428	comp193049_g1_i3	hypothetical protein TRSC58_05274	5.70E-11
429	comp195476_g1_i8	PTEN_CANFA	3.20E-57

430	comp192793_g3_i2	EDRF1_PONAB	9.60E-36
431	comp190712_g1_i5	hypothetical protein COCSUDRAFT_31865	1.40E-19
432	comp192354_g1_i2	hypothetical protein	4.00E-11
433	comp183887_g3_i2	hypothetical protein EMIHUDRAFT_197272	4.40E-16
434	comp186315_g2_i1	2-heptaprenyl-1,4-naphthoquinone methyltransferase	3.00E-10
435	comp190031_g1_i12	PREDICTED: uncharacterized protein LOC106806144	5.40E-20
436	comp192446_g1_i5	alpha-L-rhamnosidase	2.40E-63
437	comp196137_g1_i2	UVR8_ARATH	7.70E-15
438	comp184230_g2_i4	hypothetical protein	1.10E-08
439	comp149921_g1_i2	LYRM4_TAEGU	7.50E-18
440	comp178617_g1_i6	glycosyl transferase	3.80E-21
441	comp192096_g1_i3	glycosyl transferase	2.40E-23
442	comp180626_g1_i24	Os03g0719500predicted protein, partial	8.50E-159
443	comp182271_g1_i2	ABCD4_MOUSE	3.30E-66
444	comp177612_g1_i2	SRPR_DROME	4.20E-115
445	comp186502_g2_i5	phosphatidylinositol-4-phosphate 5-kinase	4.10E-55
446	comp184818_g1_i5	MAP kinase	6.10E-44
447	comp175996_g1_i1	ATPbinding Cassette (ABC) superfamily	1.50E-08
448	comp182507_g1_i10	DDL_RALME	4.50E-09
449	comp185561_g1_i1	Y1181_ARATH	2.00E-94
450	comp170477_g1_i8	CIMA_GEOSL	3.50E-149
451	comp189110_g1_i1	NLRC3_HUMAN	2.00E-17
452	comp192618_g1_i1	hypothetical protein MNEG_5574	1.40E-09
453	comp193641_g1_i1	ABC transporter D family	0.00E+00
454	comp186463_g1_i7	ser thr kinase, ABC1 family	0.00E+00
455	comp174852_g3_i7	PAPS4_ARATH	2.30E-60
456	comp188879_g1_i3	PDE11_DROME	5.90E-74
457	comp191424_g2_i2	glutaredoxin-1 isoform X2	2.90E-102
458	comp193657_g1_i5	SPXS1_DICDI	6.80E-45
459	comp195191_g1_i3	magnesium chelatase	0.00E+00
460	comp187809_g1_i2	cobyric acid synthase	6.10E-120
461	comp175155_g1_i1	flagellar associated	1.80E-22
462	comp182900_g1_i4	GTP cyclohydrolase I	3.10E-68
463	comp189948_g1_i6	outer membrane-stress sensor serine endopeptidase	4.60E-08
464	comp194426_g3_i5	KAPR_DICDI	5.50E-12

465	comp192045_g1_i1	adenylate cyclase	6.20E-59
466	comp168530_g1_i1	guanylate cyclase	5.30E-07
467	comp183157_g1_i2	lysophospholipid acyltransferase LPEAT1-like isoform X1	5.50E-23
468	comp191532_g1_i5	NEDD4-like E3 ubiquitin- ligase WWP2predicted protein, partial	1.10E-45
469	comp180616_g3_i3	THI41_PHYPA	2.60E-122
470	comp195426_g3_i1	beta-phosphoglucomutase	0.00E+00
471	comp181135_g1_i1	hypothetical protein SDRG_05032	3.80E-80
472	comp192124_g1_i3	PLDA1_RICCO	4.90E-35
473	comp194187_g1_i2	hypothetical protein AMSG_03068	2.00E-21
474	comp180626_g1_i15	DYRK2_DICDI	2.70E-63
475	comp182098_g1_i1	PCAT1_HUMAN	1.60E-40
476	comp192618_g1_i2	hypothetical protein MNEG_5574	1.40E-09
477	comp192811_g1_i5	Y1760_MYCTU	3.60E-11
478	comp178426_g1_i8	KAPS_SCHPO	1.30E-81
479	comp191200_g1_i2	LGRC_BREPA	2.30E-57
480	comp195006_g1_i8	hybrid sensor histidine kinase response regulator	1.00E-20
481	comp182293_g3_i3	kelch domain-containing 3	2.10E-22
482	comp189549_g1_i5	RNZ_STRSV	2.00E-67
483	comp176781_g2_i2	EFHB_MOUSE	4.30E-09
484	comp171068_g1_i1	SAE1_HUMAN	1.30E-45
485	comp183078_g2_i2	hypothetical protein H310_03776	1.60E-15
486	comp195888_g1_i5	GALE_METJA	1.70E-16
487	comp180668_g1_i4	RS20_CYAA5	7.10E-13
488	comp184790_g1_i1	DAAF5_MOUSE	3.10E-35
489	comp178977_g1_i3	RRAGA_HUMAN	2.40E-150
490	comp185375_g1_i9	Eukaryotic translation initiation factor 4E-3	1.10E-19
491	comp178799_g1_i3	L-aspartate oxidase	8.50E-81
492	comp164413_g1_i2	FLU_ARATH	1.20E-12
493	comp180261_g1_i3	hypothetical protein EMIHUDRAFT_435586	9.30E-58
494	comp187279_g3_i2	uridine kinase	2.60E-69
495	comp183520_g1_i1	PIBF1_HUMAN	5.50E-17
496	comp178763_g2_i6	ATPase AAA	4.50E-42
497	comp171377_g1_i1	F221A_XENLA	1.80E-11
498	comp180374_g1_i3	Y5279_ARATH	6.60E-97
499	comp171491_g1_i2	na+ solute symporter	6.50E-75

500	comp165500_g1_i2	-tyrosine phosphatase 3	2.00E-15
501	comp186161_g1_i1	PLCD4_HUMAN	1.20E-54
502	comp195163_g1_i4	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.70E-07
503	comp189959_g1_i1	transmembrane and coiled-coil domain-containing 3 isoform X1	2.10E-81
504	comp173992_g1_i2	hypothetical protein GUITHDRAFT_157081	1.60E-38
505	comp191827_g2_i6	UVR3_ARATH	9.40E-29
506	comp172515_g1_i1	IF20A_XENLA	3.50E-14
507	comp188859_g1_i17	hypothetical protein H310_00261	1.90E-10
508	comp176522_g1_i1	PSAA_EUGGR	0.00E+00
509	comp178713_g2_i1	T9SS C-terminal target domain-containing	6.00E-17
510	comp191706_g1_i4	POXA_DICDI	7.30E-59
511	comp189519_g1_i6	YFZ3_SCHPO	1.80E-09
512	comp192140_g1_i4	predicted protein	2.90E-30
513	comp184183_g3_i1	ITPR1_MOUSE	2.40E-42
514	comp187520_g3_i2	membrane-bound adenylyl cyclase	1.90E-65
515	comp193939_g1_i6	KRAC_DICDI	4.60E-97
516	comp182551_g1_i3	SAC9_ARATH	2.60E-35
517	comp192028_g1_i1	unnamed protein product	6.50E-28
518	comp194495_g1_i5	MHCKB_DICDI	8.00E-25
519	comp189137_g2_i14	NFX1-type zinc finger-containing 1-like	9.00E-73
520	comp185622_g1_i1	BFR2_SCHPO	3.80E-14
521	comp183228_g1_i2	zinc metallo ase-like	4.70E-37
522	comp176563_g1_i10	ubiquinone biosynthesis methyltransferase	1.00E-14
523	comp193667_g2_i3	CRYD_GLOVI	2.70E-97
524	comp187219_g2_i2	gonidia-specific KA_k47	8.80E-16
525	comp47594_g1_i1	aerotolerance regulator	4.20E-12
526	comp191918_g1_i3	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase-like	4.10E-44
527	comp189948_g1_i7	outer membrane-stress sensor serine endopeptidase	3.20E-08
528	comp195349_g1_i1	SWET1_MOUSE	1.10E-09
529	comp185167_g1_i2	MKS1_HUMAN	7.50E-07
530	comp183153_g1_i1	CFA53_MOUSE	3.40E-16
531	comp189588_g2_i1	beta-glucan elicitor receptor	2.50E-176
532	comp180837_g1_i3	metallophosphoesterase	1.60E-46
533	comp192391_g2_i4	hypothetical protein AURANDRAFT_67549	4.00E-13

534	comp184037_g1_i1	hypothetical protein EMIHUDRAFT_113418predicted protein, partial	1.10E-08
535	comp185483_g1_i2	biotin carboxylase	1.80E-88
536	comp182366_g1_i4	bile Acid:Na+ symporter family	2.10E-39
537	comp185654_g4_i2	UBP49_MOUSE	1.30E-40
538	comp176563_g1_i8	ubiquinone biosynthesis methyltransferase	1.70E-14
539	comp186975_g1_i2	ATG4A_BOVIN	7.70E-30
540	comp175038_g1_i7	Tetratricopeptide-like helical	4.40E-46
541	comp190150_g2_i3	endoribonuclease Dicer homolog 1	5.10E-53
542	comp148589_g1_i1	DENR_MAGO7	4.50E-25
543	comp191943_g4_i2	Spermidine synthase	4.90E-08
544	comp172492_g1_i1	PRY1_YEAST	2.20E-21
545	comp194169_g1_i3	KAPR_DICDI	2.60E-13
546	comp176840_g3_i1	adenylate kinase	3.30E-113
547	comp187139_g2_i1	exonuclease 3 -5 domain-containing 1-like	1.20E-33
548	comp183455_g1_i2	trypanothione reductase	0.00E+00
549	comp152332_g1_i2	AZOR_DECAR	2.10E-67
550	comp172640_g1_i1	THDH_SCHPO	5.30E-136
551	comp179995_g1_i2	predicted protein, partial	0.00E+00
552	comp196043_g1_i1	TIGR02452 family	4.60E-45
553	comp183613_g3_i9	translation initiation factor 4F, cap-binding subunit	4.90E-61
554	comp181935_g1_i3	CABL2_MOUSE	1.70E-28
555	comp188242_g1_i4	PFR1_TRYBB	2.10E-89
556	comp178752_g1_i9	alpha subunit of photoactivated adenylyl cyclase	7.40E-23
557	comp192964_g1_i4	Tbc2 translation factor, chloroplastic	2.20E-30
558	comp178140_g1_i4	CY41_TRYBB	8.90E-27
559	comp180261_g1_i4	hypothetical protein EMIHUDRAFT_435586	1.60E-57
560	comp186384_g1_i2	SEC_ARATH	2.90E-21
561	comp174587_g3_i2	hypothetical protein COCSUDRAFT_43198	1.20E-09
562	comp168860_g1_i6	DJC18_BOVIN	1.90E-30
563	comp169528_g1_i1	CTL2_DICDI	1.40E-33
564	comp189959_g1_i10	transmembrane and coiled-coil domain-containing 3 isoform X1	9.20E-81
565	comp175878_g1_i2	hypothetical protein GUITHDRAFT_108701	4.60E-32
566	comp174141_g1_i1	hypothetical protein GUITHDRAFT_164479	7.90E-31
567	comp194964_g1_i6	adenylate guanylate cyclase	2.00E-08
568	comp178385_g1_i11	ubiquinone biosynthesis	4.20E-81

569	comp193581_g2_i7	histidine kinase	2.40E-09
570	comp161874_g1_i3	GST1_SCHPO	2.10E-08
571	comp194523_g1_i5	predicted protein, partial	1.10E-08
572	comp186228_g1_i3	NPK1-related kinase 2	8.60E-79
573	comp161903_g1_i1	calcium homeostasis endoplasmic reticulum	2.50E-11
574	comp188190_g1_i2	GRTP1_XENTR	6.10E-18
575	comp184497_g1_i2	PLT5_ARATH	3.00E-28
576	comp160506_g1_i1	Ribosomal S21	1.20E-15
577	comp192187_g1_i2	hypothetical protein PBRA_000501	1.30E-37
578	comp193439_g2_i5	SLBP2_XENLA	1.00E-08
579	comp167661_g1_i3	CC151_MOUSE	3.70E-27
580	comp193677_g1_i11	cellulase (glycosyl hydrolase family 5) domain	0.00E+00
581	comp193396_g2_i5	HERC4_RAT	5.60E-15
582	comp189019_g3_i3	major facilitator superfamily	3.60E-22
583	comp193215_g2_i1	Armadillo betacatenin-like repeat domain containing	6.70E-11
584	comp192403_g1_i8	beta subunit of photoactivated adenylyl cyclase	6.20E-13
585	comp180381_g1_i1	PAS domain-containing sensor histidine kinase	2.70E-13
586	comp193150_g1_i4	RSPH9_HUMAN	1.20E-25
587	comp185800_g2_i4	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, chloroplastic	0.00E+00
588	comp182885_g2_i1	methyltransferase	4.60E-23
589	comp175996_g1_i4	ATPbinding Cassette (ABC) superfamily	1.40E-08
590	comp159812_g1_i2	MCES1_ORYSJ	2.20E-28
591	comp176525_g1_i3	CAN15_MOUSE	3.30E-41
592	comp174958_g1_i8	cystatin-A	1.00E-19
593	comp187208_g1_i2	DUS1_ARATH	1.90E-27
594	comp179493_g3_i4	KAPR_YARLI	6.20E-14
595	comp181589_g1_i14	3-phosphoinositide dependent kinase-1	2.40E-85
596	comp176502_g1_i4	predicted protein, partial	1.20E-48
597	comp193627_g2_i8	hypothetical protein SDRG_04080	3.50E-12
598	comp192486_g1_i7	alpha subunit	1.70E-19
599	comp166286_g1_i2	CC189_MOUSE	1.70E-17
600	comp180850_g1_i9	Glycosyltransferase AER61, uncharacterized	2.00E-16
601	comp170446_g1_i4	RH46_ARATH	2.60E-132
602	comp177794_g2_i4	hypothetical protein ABL78_5049	4.70E-08
603	comp192155_g2_i1	alpha-1,2-galactosyltransferase gmh3-like	2.70E-19

604	comp179243_g1_i4	ECE1_MOUSE	9.20E-103
605	comp186560_g5_i10	xylulose kinase	3.20E-129
606	comp177716_g1_i1	formate transporter	1.90E-113
607	comp190858_g1_i2	phosphatase 1 regulatory subunit SDS22	8.60E-12
608	comp179481_g1_i3	CDC23_ARATH	6.20E-122
609	comp186209_g1_i3	hypothetical protein L916_17225	3.00E-09
610	comp182887_g1_i3	hypothetical protein ABB37_09572	1.60E-29
611	comp193180_g1_i3	MHCKA_DICDI	5.40E-36
612	comp160801_g1_i1	RTPR_EUGGR	1.20E-18
613	comp172899_g2_i3	FBT2_ARATH	9.50E-08
614	comp195521_g2_i1	adenylate guanylate cyclase catalytic domain	4.40E-08
615	comp184939_g1_i2	nibrin	1.40E-26
616	comp195268_g1_i1	hypothetical protein	1.70E-23
617	comp184436_g1_i7	amino acid permease	1.20E-52
618	comp190925_g1_i8	UVB31_ARATH	9.50E-08
619	comp186014_g1_i5	6GAL_HYPRU	2.40E-30
620	comp176932_g2_i2	kinase domain	1.30E-30
621	comp188172_g1_i1	GTDC1_DROME	3.50E-50
622	comp191349_g3_i1	sel1 domain repeat-containing	2.90E-35
623	comp164267_g1_i1	AVP_HORVU	4.90E-41
624	comp170415_g1_i5	haloalkane dehalogenase	1.60E-15
625	comp190028_g2_i4	NLTP_RABIT	8.50E-16
626	comp192676_g1_i3	Carbohydrate-binding and sugar hydrolysis	2.50E-08
627	comp195318_g1_i8	predicted protein	3.90E-15
628	comp183454_g1_i2	THLA_PEPD6	7.20E-81
629	comp174994_g1_i6	PUF68_DROME	3.40E-14
630	comp183030_g1_i3	Y4967_ARATH	8.30E-37
631	comp194992_g1_i4	ANR50_HUMAN	1.10E-16
632	comp166947_g1_i1	TEBP_MOUSE	1.60E-07
633	comp190279_g1_i2	ABC transporter C family member 2-like	0.00E+00
634	comp170391_g4_i1	MSRA_RHOCS	3.80E-84
635	comp175544_g2_i2	PR5_ARATH	4.80E-20
636	comp195944_g1_i3	polycystin-2-like isoform X2	9.50E-31
637	comp183078_g1_i5	hypothetical protein H310_03776	1.70E-15
638	comp185440_g1_i3	VTI11_ARATH	1.10E-17

639	comp191516_g1_i2	LRRC9_HUMAN	5.80E-55
640	comp187555_g3_i4	myosin heavy chain kinase	1.80E-40
641	comp190752_g2_i3	metallophosphoesterase	6.50E-157
642	comp191055_g1_i12	TPR domain-containing	1.00E-11
643	comp184026_g1_i3	DHKL_DICDI	2.30E-12
644	comp196096_g1_i3	TCRG1_HUMAN	1.10E-28
645	comp188532_g2_i3	PSTS_METJA	5.40E-14
646	comp170895_g2_i3	mitochondrial carrier	1.90E-52
647	comp166062_g1_i2	TM222_HUMAN	3.40E-30
648	comp175355_g1_i1	SAV_SULAC	1.80E-16
649	comp192453_g4_i2	SAMC1_ARATH	1.10E-21
650	comp183647_g3_i4	membrane-associated ,	4.20E-08
651	comp187682_g1_i5	UDP-galactopyranose mutase	1.10E-135
652	comp188423_g1_i3	TGM5_HUMAN	3.80E-40
653	comp174432_g1_i1	beta-type IP39	7.50E-09
654	comp196297_g2_i1	ADP-ribosyltransferase 1 precursor	1.30E-08
655	comp194643_g1_i1	calpain-like cysteine peptidase	1.70E-08
656	comp175194_g1_i2	DIOXL_ARATH	3.40E-42
657	comp187666_g2_i4	kinase domain	8.30E-19
658	comp170275_g1_i6	Poly (A) RNA polymerase cid14	4.80E-51
659	comp178094_g1_i5	glycosyl transferase	1.80E-18
660	comp180660_g1_i2	RBE1_CANAL	1.20E-19
661	comp261183_g1_i1	AAEL017413- partial	5.00E-08
662	comp189908_g2_i6	S35D2_BOVIN	1.70E-62
663	comp185293_g1_i6	metallophosphoesterase domain-containing 1	2.30E-61
664	comp189876_g7_i9	predicted protein	4.40E-21
665	comp192703_g2_i2	ankyrin repeat	8.30E-29
666	comp163104_g1_i1	PCS3_LOTJA	1.60E-67
667	comp187064_g1_i2	PLA16_ARATH	2.40E-13
668	comp192589_g1_i8	phosphatidylinositol-4-phosphate 5-kinase	3.00E-19
669	comp194698_g1_i6	PRMS_MAIZE	2.10E-11
670	comp186741_g1_i2	acyltransferase At1g54570, chloroplastic	1.90E-89
671	comp176965_g1_i3	COMB_STRCO	6.10E-22
672	comp195319_g2_i5	membrane-bound adenylyl cyclase	3.90E-19
673	comp182937_g3_i1	RNH2_PARUW	1.20E-50

674	comp161429_g1_i5	LPAT4_ARATH	1.50E-39
675	comp172496_g1_i1	AMERL_HUMAN	1.30E-40
676	comp187984_g1_i4	exostosin family	1.30E-16
677	comp191840_g1_i1	predicted protein, partial	3.90E-121
678	comp193823_g1_i4	BPRX_DICNO	4.30E-72
679	comp191510_g1_i1	enhancer of polycomb group	7.80E-43
680	comp191026_g1_i1	monogalactosyldiacylglycerol synthase	1.50E-84
681	comp181303_g1_i2	EF-Hand calcium-binding site	7.10E-87
682	comp186290_g1_i4	CMGC MAPK kinase	2.10E-132
683	comp187682_g2_i6	GLF_ECOLI	3.10E-47
684	comp185260_g1_i9	kinase A regulatory subunit	1.30E-54
685	comp189939_g1_i2	E3 ubiquitin- ligase SIAH1-like	1.20E-12
686	comp181497_g1_i6	ubiquinone biosynthesis methyltransferase	3.00E-14
687	comp179410_g1_i2	CI171_DANRERecName: Full=Uncharacterized protein C9orf171 homolog	3.50E-09
688	comp191609_g1_i2	BRISC and BRCA1-A complex member 1-like	4.50E-14
689	comp163315_g1_i1	S39AC_MACFA	7.20E-09
690	comp191944_g1_i2	CASD1_DANRE	1.50E-07
691	comp187431_g1_i1	RUFY2_PONAB	7.10E-07
692	comp167661_g1_i1	CC151_MOUSE	3.80E-27
693	comp184080_g1_i5	hypothetical protein SAMD00019534_007780predicted protein, partial	1.20E-14
694	comp180673_g1_i8	YDK2_SCHPO	8.50E-18
695	comp178874_g1_i5	MET22_HUMAN	2.40E-09
696	comp135330_g1_i1	hypothetical protein GUITHDRAFT_161344predicted protein, partial	8.60E-13
697	comp191553_g1_i8	receptor-type adenylate cyclase GRESAG 4,	9.70E-32
698	comp184221_g1_i1	hypothetical protein	8.50E-09
699	comp190947_g2_i5	S12A4_HUMAN	2.60E-119
700	comp184224_g1_i3	alpha beta hydrolase family	3.50E-18
701	comp191447_g2_i2	hypothetical protein VOLCADRAFT_97022	4.00E-12
702	comp179296_g1_i2	hypothetical protein PTSG_10764	2.80E-10
703	comp194969_g1_i2	AT133_HUMAN	3.90E-32
704	comp191613_g3_i2	CY42_TRYBB	1.40E-47
705	comp177794_g2_i6	hypothetical protein ABL78_4992	1.00E-07
706	comp189755_g1_i2	AMPD_ARATH	7.90E-147
707	comp189575_g1_i2	ADPRH_HUMAN	2.40E-44
708	comp189137_g2_i13	ZNFX1_MOUSE	2.00E-59

709	comp161920_g1_i1	glycosyl family 43	3.00E-50
710	comp187105_g1_i3	POLH_ARATH	8.90E-83
711	comp191053_g2_i1	TRPV6_HUMAN	3.50E-13
712	comp181191_g1_i2	PTH_SULIA	1.90E-16
713	comp185495_g1_i10	PEPD_LACSK	1.50E-24
714	comp186753_g1_i1	ANR27_PONAB	1.90E-07
715	comp163743_g1_i2	AIDA_DANRE	1.30E-31
716	comp193688_g2_i2	MKT1_YEAST	2.80E-11
717	comp189695_g1_i5	Y663_METTH	6.10E-66
718	comp141180_g1_i3	CSPA_STIAD	1.00E-18
719	comp179230_g1_i1	SDRD_STAAW	5.20E-40
720	comp171450_g1_i9	PSBC_EUGGR	0.00E+00
721	comp193947_g1_i6	hypothetical protein STCU_06761	6.30E-08
722	comp188631_g1_i1	hypothetical protein DDB_G0269288	2.40E-22
723	comp193484_g2_i1	regulator of chromosome condensation RCC1	9.30E-19
724	comp192784_g1_i2	sporulation	6.70E-09
725	comp179760_g1_i5	YOS3_SCHPO	3.00E-60
726	comp186246_g1_i1	receptor-type adenylate cyclase,	3.10E-46
727	comp175309_g1_i6	Neurotrypsin	3.20E-07
728	comp193674_g1_i10	OTU7B_MOUSE	1.70E-23
729	comp189878_g1_i1	hypothetical protein GUITHDRAFT_121297	1.50E-28
730	comp157799_g1_i1	ATG26_YEAS7	2.00E-13
731	comp163276_g1_i2	GLTP1_ARATH	2.60E-32
732	comp192403_g1_i7	PBL9F_EUGGR	2.60E-13
733	comp162090_g1_i1	PIN1_ARATH	2.80E-37
734	comp159204_g1_i2	AB17A_RAT	1.40E-18
735	comp147834_g1_i1	ZNHI1_HUMAN	1.00E-16
736	comp189084_g1_i1	kinase domain	1.60E-08
737	comp195163_g1_i10	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.00E-07
738	comp181632_g1_i1	PDXK_ARATH	1.80E-88

Appendix table 25 genes name of area 25

(a) List of TF family (63 transcripts of 27 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp189899_g1_i3	C2H2	5.70E-28
2	comp190627_g1_i2	C3H	1.80E-13
3	comp158853_g1_i2	HB-other	7.30E-10
4	comp190065_g1_i3	LBD	2.90E-20
5	comp189670_g1_i2	LBD	5.40E-19
6	comp190835_g1_i2	M-type	7.00E-13
7	comp179113_g1_i4	MIKC	4.10E-42
8	comp183261_g1_i3	MYB	1.30E-52
9	comp183386_g3_i3	NAC	2.40E-07
10	comp184733_g2_i2	NAC	5.70E-30
11	comp184733_g2_i3	NAC	5.10E-30
12	comp190350_g2_i3	Nin-like	1.00E-11
13	comp183064_g2_i3	B3	1.60E-16
14	comp195646_g1_i5	GeBP	2.80E-07
15	comp190804_g3_i2	YABBY	2.70E-30
16	comp187889_g2_i1	bZIP	1.50E-21
17	comp183064_g2_i2	B3	1.60E-16
18	comp190958_g2_i3	C3H	1.00E-12
19	comp191478_g3_i2	C3H	2.50E-07
20	comp195485_g1_i2	C3H	4.20E-08
21	comp191534_g1_i12	C3H	3.60E-08
22	comp175684_g2_i1	E2F_DP	4.50E-19
23	comp179058_g1_i1	E2F_DP	1.30E-25
24	comp179058_g1_i2	E2F_DP	1.20E-25
25	comp180644_g1_i3	ERF	8.00E-118
26	comp178939_g1_i1	GATA	8.40E-08
27	comp157424_g1_i2	HB-other	9.50E-18
28	comp183896_g2_i2	MYB	2.40E-15
29	comp184981_g1_i1	MYB	2.50E-21
30	comp194235_g1_i3	NAC	3.90E-11
31	comp181387_g1_i2	NAC	1.10E-15
32	comp183386_g2_i1	NAC	2.30E-07

33	comp184133_g1_i3	NAC	2.80E-28
34	comp183768_g1_i2	TALE	1.20E-10
35	comp150391_g1_i2	Trihelix	9.00E-11
36	comp183284_g1_i5	VOZ	6.50E-13
37	comp179633_g1_i1	WRKY	2.40E-09
38	comp180004_g1_i2	WRKY	1.50E-15
39	comp180498_g1_i2	bHLH	9.80E-09
40	comp189160_g2_i1	bZIP	1.80E-16
41	comp187889_g2_i2	bZIP	1.50E-21
42	comp178211_g1_i3	C2H2	3.10E-34
43	comp178149_g1_i2	C2H2	1.50E-28
44	comp191337_g2_i1	C3H	1.60E-10
45	comp179240_g1_i2	CAMTA	1.30E-11
46	comp188886_g2_i2	CPP	1.30E-38
47	comp183118_g1_i4	B3	3.30E-21
48	comp183118_g1_i5	B3	2.40E-21
49	comp176471_g1_i1	FAR1	3.00E-114
50	comp194244_g2_i2	FAR1	7.20E-32
51	comp174585_g2_i6	GRAS	4.40E-18
52	comp177978_g1_i3	C2H2	4.80E-12
53	comp171050_g1_i1	C3H	4.00E-11
54	comp183896_g1_i2	NAC	8.20E-10
55	comp191474_g1_i1	NF-YC	2.70E-12
56	comp184137_g1_i5	STAT	3.50E-07
57	comp160135_g1_i2	Trihelix	1.80E-09
58	comp162969_g1_i8	GeBP	1.40E-25
59	comp195646_g1_i4	GeBP	3.00E-07
60	comp180004_g1_i3	WRKY	8.90E-16
61	comp180480_g2_i1	bHLH	2.80E-08
62	comp180480_g2_i3	bHLH	3.10E-08
63	comp186117_g1_i1	bZIP	1.70E-21

(b) List of other genes name (202 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp194338_g1_i6	major facilitator superfamily domain-containing 12	1.70E-62
2	comp191536_g1_i10	SHIA_ECOLI	2.20E-12
3	comp167797_g1_i2	Tetratricopeptide repeat (fragment)	5.90E-12
4	comp183468_g1_i3	KMO_XENTR	3.70E-77
5	comp189338_g1_i11	membrane-bound adenylyl cyclase	1.20E-39
6	comp182969_g1_i1	ABCC9_MOUSE	1.10E-18
7	comp187325_g1_i5	glycosyl transferase	1.40E-17
8	comp189768_g2_i4	RIBC1_HUMAN	4.70E-15
9	comp196279_g2_i6	ADCY1_MOUSE	3.90E-12
10	comp180857_g1_i2	PTR36_ARATH	2.50E-11
11	comp173405_g1_i2	MEG5_MAIZE	3.00E-07
12	comp185727_g1_i2	PPD2_ARATH	3.60E-07
13	comp176974_g1_i1	VDE_LACSA	8.60E-52
14	comp178742_g1_i1	insulin-degrading enzyme	1.50E-25
15	comp189940_g1_i1	FAS1_YEAST	2.30E-19
16	comp185440_g2_i7	VTI11_ARATH	2.70E-17
17	comp180703_g1_i3	DED1_PHANO	0.00E+00
18	comp192489_g1_i4	Sedoheptulose-1,7-bisphosphatase, chloroplastic	0.00E+00
19	comp196132_g1_i2	ef hand family	5.80E-35
20	comp178883_g1_i2	phosphoglycerate mutase family domain containing	7.80E-16
21	comp162097_g1_i1	LPSA_DICNO	1.20E-27
22	comp174011_g2_i14	hydroxyacylglutathione hydrolase	8.60E-114
23	comp192048_g2_i6	AK1_DICDI	8.20E-40
24	comp188567_g1_i2	conserved hypothetical protein	1.20E-39
25	comp176022_g1_i1	DDL_BORPD	2.80E-13
26	comp187179_g2_i2	CNOT4_MOUSE	1.40E-22
27	comp176060_g1_i2	glycosyl transferase	1.80E-17
28	comp177924_g1_i6	AAPT2_ARATH	4.60E-53
29	comp190652_g2_i2	epidermal retinol dehydrogenase 2-like	3.10E-34
30	comp180616_g1_i3	THI41_PHYPA	4.10E-122
31	comp169439_g2_i1	NHX6_ARATH	4.00E-42
32	comp176403_g1_i3	KELC_DROME	4.30E-11
33	comp195579_g1_i5	hypothetical protein PBRA_000501	1.80E-26

34	comp195204_g1_i1	hypothetical protein THAOC_26672	1.50E-27
35	comp195290_g2_i5	photosystem I assembly Ycf3	6.50E-10
36	comp186993_g1_i3	SAMH1_HUMAN	4.70E-89
37	comp175762_g1_i2	E3 ubiquitin- ligase XIAP-like	1.10E-12
38	comp178488_g1_i2	ADH1_GEOSE	4.30E-42
39	comp188981_g4_i9	phosphatidylinositol-4-phosphate-5-kinase-likep ro tein	4.70E-60
40	comp169913_g1_i4	polypyrimidine tract-binding homolog 2-like	1.00E-08
41	comp195414_g1_i6	PABP6_ARATH	1.20E-13
42	comp183743_g2_i1	lipase, putative	8.50E-30
43	comp155380_g1_i2	ADT3_BOVIN	8.20E-89
44	comp175206_g1_i3	PLANT CADMIUM RESISTANCE 2-like	1.30E-09
45	comp179544_g1_i3	CCD39_CHLRE	7.50E-106
46	comp176144_g1_i3	ELOV4_MOUSE	6.10E-37
47	comp179603_g2_i5	AMZ2_MACFA	1.00E-29
48	comp191696_g2_i8	TAER_MYCUA	8.00E-26
49	comp195127_g1_i2	probable beta-1,3-galactosyltransferase 19	2.80E-16
50	comp192811_g1_i4	Y1760_MYCTU	3.30E-11
51	comp173974_g2_i4	enoyl- hydratase	2.70E-60
52	comp157300_g1_i2	CYC6_EUGVI	3.90E-32
53	comp194807_g1_i5	RUSD1_DANRE	2.80E-18
54	comp167089_g1_i3	-tyrosine phosphatase 3	2.00E-15
55	comp175860_g1_i1	mcKusick-Kaufman Bardet-Biedl syndromes chaperonin	1.40E-10
56	comp195294_g1_i3	Ankyrin repeat	3.00E-79
57	comp180558_g1_i1	hypothetical protein PBRA_000836	4.70E-46
58	comp145585_g1_i2	NLRC3_HUMAN	4.10E-34
59	comp170185_g3_i2	hypothetical protein PPERSA_09341	1.90E-13
60	comp172639_g1_i5	SAMC1_ARATH	1.20E-10
61	comp180616_g3_i4	THI4_SCHPO	9.20E-74
62	comp196132_g1_i4	ef hand family	5.70E-35
63	comp163945_g1_i3	PSB5_DICDI	7.30E-94
64	comp184553_g1_i1	hypothetical protein	5.20E-08
65	comp190712_g1_i8	hypothetical protein COCSUDRAFT_31865	1.30E-19
66	comp180948_g1_i3	Wd-repeat ,	1.70E-51
67	comp181758_g1_i3	unnamed protein product	1.40E-14
68	comp178324_g2_i2	EFHC2_DANRE	1.00E-59

69	comp151206_g1_i2	PSBB_EUGGR	0.00E+00
70	comp179399_g1_i5	thiol protease aleurain-like	1.60E-104
71	comp152886_g1_i2	mitochondrial tryptophan rich sensory 1	1.90E-08
72	comp190831_g1_i1	CY43_TRYBB	1.30E-10
73	comp186376_g1_i10	Phosphoglycerate mutase	1.80E-76
74	comp163805_g2_i4	ENPP1_MOUSE	3.00E-09
75	comp190909_g2_i3	cGMP-dependent kinase egl-4	2.80E-16
76	comp182711_g1_i3	Trk1 potassium transporter	2.70E-42
77	comp187308_g1_i2	CID13_SCHPO	1.20E-12
78	comp175049_g1_i2	YF2E_SCHPO	3.70E-19
79	comp180651_g1_i1	inositol-hexakisphosphate kinase	1.80E-23
80	comp184944_g1_i2	Low temperature viability	1.10E-12
81	comp193767_g1_i4	CYAA_LEIDO	5.60E-28
82	comp191319_g2_i3	TOM1_NEUCR	1.10E-14
83	comp179760_g1_i3	ubiquinone biosynthesis	1.90E-80
84	comp189674_g1_i2	Y090_METJA	2.20E-50
85	comp184939_g1_i1	nibrin	1.20E-26
86	comp170954_g2_i1	cardiolipin synthase	1.10E-31
87	comp184640_g1_i4	calpain-like cysteine peptidase	4.50E-17
88	comp183181_g1_i1	predicted protein	9.50E-33
89	comp190070_g2_i8	SPNS1_ARATH	7.60E-37
90	comp168939_g1_i6	hypothetical protein Ctob_006924	8.00E-24
91	comp184555_g1_i2	LONF1_HUMAN	7.90E-41
92	comp171107_g1_i3	ATP-binding Cassette (ABC) superfamily	7.30E-14
93	comp185375_g2_i9	IF4E2_MOUSE	2.10E-18
94	comp190383_g2_i4	kinase domain	4.10E-11
95	comp181835_g1_i3	GDIR_ARATH	6.00E-08
96	comp189908_g1_i1	UDP-glucuronic acid UDP-N-acetylgalactosamine transporter	7.00E-68
97	comp192121_g1_i4	flagellar associated	9.70E-14
98	comp195895_g1_i4	uracil-DNA glycosylase	6.10E-09
99	comp176797_g1_i2	ELAV 3 isoform X8	5.60E-21
100	comp175449_g1_i5	NSR1_YEAST	2.10E-11
101	comp166967_g1_i6	GP157_MOUSE	2.30E-08
102	comp185484_g1_i1	GCR1_ARATH	5.80E-22
103	comp180651_g1_i3	IP6K1_MOUSE	1.10E-19

104	comp184300_g1_i1	DNA repair endonuclease XPF-like	1.80E-73
105	comp184939_g2_i1	nibrin	2.80E-26
106	comp191923_g1_i6	serine threonine- kinase STK11 isoform X1	4.70E-64
107	comp195823_g1_i1	membrane-bound adenylyl	1.70E-08
108	comp177345_g2_i2	Hel1p	7.20E-55
109	comp180639_g1_i3	FBX11_RAT	6.10E-19
110	comp158183_g1_i1	ODC_DICDI	4.00E-65
111	comp190549_g2_i1	PSD2_ORYSJ	1.60E-69
112	comp158577_g1_i6	zinc finger with UFM1-specific peptidase domain	9.60E-53
113	comp192636_g2_i3	xyloglucan galactosyltransferase KATAMARI1-like	2.80E-09
114	comp193789_g1_i4	peptidase S8	6.00E-90
115	comp186653_g2_i12	serine threonine- kinase 2 19-like	3.80E-36
116	comp162703_g3_i1	AP4S_ARATH	1.20E-59
117	comp171450_g1_i14	PSBD_EUGGR	9.60E-46
118	comp180520_g2_i1	SWIM zinc finger domain-containing	1.00E-16
119	comp161194_g1_i3	SSRP1_XENLA	1.10E-44
120	comp176720_g1_i5	predicted protein, partial	1.10E-53
121	comp162368_g1_i1	glycosyl transferase	9.70E-93
122	comp195571_g1_i4	Proteophosphoglycan ppg4	1.70E-40
123	comp188287_g2_i5	TLC2_ARATH	2.00E-150
124	comp189383_g1_i1	PCYOX_ARATH	2.90E-48
125	comp175439_g1_i8	predicted protein	2.50E-18
126	comp186036_g1_i1	Hypothetical protein, putative	4.80E-56
127	comp166051_g2_i4	AVT5_SCHPO	5.40E-55
128	comp195280_g1_i2	zinc phosphodiesterase ELAC 1-like	5.00E-70
129	comp181462_g3_i3	ADHC1_MYCS2	1.20E-78
130	comp168488_g1_i6	predicted protein, partial	7.90E-54
131	comp182199_g1_i3	NLRC3_HUMAN	4.30E-32
132	comp191011_g1_i11	Y8236_DICDI	5.10E-99
133	comp190383_g2_i6	kinase domain	4.20E-11
134	comp172635_g1_i2	RAY1_ARATH	7.10E-10
135	comp191595_g4_i1	GONS4_ARATH	1.40E-50
136	comp194634_g2_i1	DYHC_DICDI	1.40E-123
137	comp174353_g2_i17	protein of unknown function DUF427	9.20E-40
138	comp185032_g4_i1	UNC47_CAEEL	2.60E-14

139	comp143008_g1_i1	BnaC09g29270D	7.20E-42
140	comp165739_g1_i1	PSMD4_ARATH	4.70E-68
141	comp175448_g1_i8	hypothetical protein CHLNCDRAFT_138448	1.00E-09
142	comp194090_g1_i1	ankyrin repeat	1.30E-35
143	comp156508_g1_i1	YCX8_EUGGR	3.70E-54
144	comp189677_g1_i1	alpha subunit of photoactivated adenylyl cyclase	3.70E-162
145	comp188566_g1_i3	MYCF_MICGR	8.40E-34
146	comp188870_g1_i5	KIN11_ARATH	3.90E-129
147	comp193927_g1_i4	FBW1B_HUMAN	1.10E-13
148	comp187741_g1_i6	TPA_exp: flp	2.70E-09
149	comp170754_g1_i2	isoform a	7.20E-21
150	comp172725_g1_i10	probable G- coupled receptor 157	2.80E-16
151	comp188053_g2_i5	asteroid homolog 1-like	4.00E-13
152	comp175891_g2_i6	S38AB_MOUSE	1.20E-45
153	comp178385_g1_i9	ubiquinone biosynthesis	1.10E-80
154	comp193517_g2_i4	ADP-ribosyltransferase 1	1.00E-09
155	comp186480_g3_i2	Polyadenylate-binding 1	4.00E-23
156	comp168258_g1_i1	isoform B-like	7.20E-10
157	comp174774_g2_i8	tetracycline resistance MFS efflux pump	1.50E-24
158	comp193657_g1_i2	SPXS1_DICDI	4.10E-45
159	comp188376_g1_i1	PREDICTED: uncharacterized protein LOC105847846	2.60E-31
160	comp194281_g1_i6	hybrid sensor histidine kinase response regulator	1.10E-20
161	comp182185_g2_i4	conserved unknown protein	2.00E-24
162	comp175529_g1_i11	IBR domain	8.20E-50
163	comp187304_g1_i1	hypothetical protein THAOC_37756predicted protein, partial	3.90E-28
164	comp181946_g1_i2	EGCSE_CYANO	7.80E-100
165	comp156890_g1_i1	LUL3_ARATH	1.60E-22
166	comp165657_g1_i3	cysteine synthase	5.70E-66
167	comp183520_g1_i3	PIBF1_HUMAN	7.00E-17
168	comp171450_g1_i18	PSBD_EUGGR	1.00E-95
169	comp189895_g1_i7	RPE_HUMAN	1.30E-82
170	comp186374_g3_i7	DNA-deoxyinosine glycosylase	1.40E-19
171	comp179373_g1_i1	UPTG_PEA	5.00E-150
172	comp186237_g1_i2	TPR repeat-containing	7.60E-13
173	comp174912_g1_i2	leucine-rich repeat and guanylate kinase domain-containing isoform X1	1.20E-18

174	comp171050_g1_i5	TTLL1_TETTS	2.80E-57
175	comp193469_g3_i3	threonine synthase	6.30E-33
176	comp167863_g1_i7	RL3_NOSP7	1.10E-64
177	comp193550_g1_i1	PIGO_HUMAN	2.00E-69
178	comp188879_g1_i5	PDE11_DROME	6.70E-74
179	comp192159_g5_i8	hypothetical protein	4.40E-20
180	comp185943_g1_i2	probable arabinosyltransferase ARAD1	1.40E-19
181	comp178404_g1_i2	PTH_GLOVI	4.90E-26
182	comp188515_g1_i6	FRO4_ARATH	2.50E-22
183	comp173406_g1_i2	YU87_ARATH	3.30E-12
184	comp184179_g2_i9	RFWD2_MOUSE	2.20E-07
185	comp192577_g2_i3	probable palmitoyltransferase ZDHHC1 isoform X1	1.70E-39
186	comp188721_g1_i2	DIOXL_ARATH	6.80E-44
187	comp194245_g1_i4	conserved unknown protein	1.00E-16
188	comp163063_g1_i2	N-formylglutamate amidohydrolase	5.70E-18
189	comp178414_g1_i5	MAP kinase phosphatase 6	3.70E-24
190	comp193381_g2_i2	IQ calmodulin-binding motif family	4.90E-99
191	comp189761_g1_i5	hypothetical protein SDRG_06217	5.20E-40
192	comp185800_g2_i3	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase, chloroplastic	0.00E+00
193	comp191011_g1_i5	tetratricopeptide repeat domain containing	5.10E-126
194	comp187161_g1_i3	hypothetical protein GUITHDRAFT_105708	3.20E-10
195	comp183642_g1_i1	PP1R8_MOUSE	3.20E-17
196	comp188546_g1_i7	UGT52_DICDI	7.80E-16
197	comp187414_g1_i2	early light induced -like 5, chloroplast precursor	8.10E-14
198	comp182374_g1_i3	predicted protein	3.10E-26
199	comp180381_g1_i4	PAS domain-containing sensor histidine kinase	2.80E-13
200	comp186146_g1_i6	FDFT_ARATH	7.10E-99
201	comp188952_g1_i1	ADP-ribosyltransferase 1 precursor	2.30E-09
202	comp176120_g1_i1	RIHA_SHESR	2.10E-62

Appendix table 26 genes name of area 26

(a) List of TF families (153 transcripts of 34 TF families)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp166325_g1_i3	B3	4.00E-20
2	comp115071_g1_i2	C3H	2.90E-24
3	comp189060_g2_i3	FAR1	9.90E-41
4	comp165204_g1_i1	G2-like	7.70E-16
5	comp161284_g2_i4	GRAS	5.00E-18
6	comp185506_g1_i1	GeBP	2.20E-19
7	comp179746_g3_i1	HB-other	4.80E-08
8	comp187410_g2_i1	ARR-B	7.90E-08
9	comp170533_g1_i1	MYB	4.40E-13
10	comp158180_g1_i3	MYB_related	4.10E-59
11	comp156701_g1_i4	MYB_related	2.80E-59
12	comp162688_g1_i1	NAC	5.50E-13
13	comp162688_g1_i3	NAC	5.60E-13
14	comp166071_g1_i4	NAC	1.50E-29
15	comp164778_g1_i1	NAC	1.50E-29
16	comp189839_g1_i6	NF-YA	1.30E-08
17	comp193661_g2_i2	C3H	5.50E-11
18	comp193661_g2_i6	C3H	5.30E-11
19	comp171662_g1_i1	YABBY	5.20E-35
20	comp182667_g1_i3	FAR1	1.20E-17
21	comp187760_g1_i1	EIL	6.70E-13
22	comp165526_g1_i1	FAR1	3.80E-32
23	comp162969_g1_i4	GeBP	1.50E-25
24	comp175353_g1_i1	WOX	7.60E-33
25	comp165515_g1_i3	WRKY	2.50E-122
26	comp167863_g1_i3	WRKY	1.80E-09
27	comp167863_g1_i5	WRKY	2.00E-09
28	comp167863_g1_i10	WRKY	2.10E-09
29	comp161141_g2_i3	bHLH	1.10E-30
30	comp167735_g1_i1	bHLH	7.80E-14
31	comp163987_g1_i1	bHLH	1.90E-21
32	comp160030_g1_i2	AP2	2.40E-52

33	comp160030_g1_i3	AP2	3.10E-52
34	comp159167_g1_i2	B3	2.40E-07
35	comp157204_g1_i3	B3	3.00E-59
36	comp181747_g1_i1	BBR-BPC	8.50E-22
37	comp161535_g1_i1	C2H2	5.50E-15
38	comp146844_g1_i1	C2H2	1.70E-09
39	comp162222_g1_i2	C2H2	4.60E-08
40	comp149499_g1_i2	C2H2	8.90E-17
41	comp131074_g1_i1	C3H	2.90E-09
42	comp193661_g2_i10	C3H	5.30E-11
43	comp193661_g2_i11	C3H	5.30E-11
44	comp193661_g2_i12	C3H	5.60E-11
45	comp194079_g1_i5	C3H	1.10E-20
46	comp194087_g1_i3	C3H	3.70E-16
47	comp144913_g1_i2	C3H	2.20E-08
48	comp110311_g1_i1	C3H	2.40E-11
49	comp179590_g1_i1	CPP	3.90E-08
50	comp189007_g2_i2	EIL	3.60E-19
51	comp188263_g1_i1	FAR1	8.50E-41
52	comp183900_g1_i3	FAR1	3.80E-33
53	comp162768_g1_i2	G2-like	5.90E-16
54	comp172113_g1_i4	G2-like	8.80E-13
55	comp171725_g2_i1	G2-like	3.20E-13
56	comp173068_g1_i4	GeBP	6.30E-09
57	comp172263_g1_i4	GeBP	3.10E-07
58	comp162969_g1_i1	GeBP	1.60E-25
59	comp180434_g1_i4	GeBP	5.80E-12
60	comp184042_g2_i2	GeBP	4.90E-08
61	comp179746_g3_i2	HB-other	4.70E-08
62	comp178049_g1_i6	HB-other	1.20E-13
63	comp178444_g1_i2	HB-other	2.30E-08
64	comp179607_g2_i1	HRT-like	7.30E-10
65	comp172861_g1_i2	LBD	4.90E-25
66	comp172903_g1_i3	LBD	9.70E-20
67	comp181033_g2_i3	M-type	2.00E-13

68	comp168178_g1_i1	MYB	1.80E-56
69	comp168531_g1_i1	MYB	1.80E-56
70	comp163408_g1_i2	MYB	9.30E-61
71	comp170533_g1_i2	MYB	3.90E-13
72	comp170533_g1_i3	MYB	4.30E-13
73	comp166556_g1_i10	MYB	2.80E-135
74	comp160199_g1_i3	NAC	9.50E-22
75	comp164945_g1_i1	NAC	8.10E-15
76	comp184589_g3_i5	NF-YC	1.50E-35
77	comp184589_g4_i4	NF-YC	1.30E-35
78	comp187823_g1_i1	SRS	4.80E-15
79	comp166380_g1_i1	STAT	2.10E-16
80	comp171011_g2_i4	WRKY	1.50E-52
81	comp168795_g2_i1	WRKY	3.70E-19
82	comp192596_g1_i1	WRKY	1.10E-46
83	comp192596_g1_i2	WRKY	1.20E-46
84	comp154901_g1_i4	WRKY	1.30E-20
85	comp154901_g1_i5	WRKY	1.90E-20
86	comp192497_g1_i2	WRKY	1.00E-09
87	comp156093_g1_i2	YABBY	1.70E-57
88	comp167474_g1_i2	YABBY	2.70E-29
89	comp159034_g1_i3	bHLH	2.40E-24
90	comp166597_g1_i3	bHLH	3.70E-09
91	comp169739_g1_i2	bHLH	2.00E-13
92	comp157822_g1_i2	bHLH	7.80E-34
93	comp168197_g1_i1	bHLH	4.50E-12
94	comp163987_g1_i4	bHLH	1.60E-21
95	comp163987_g1_i11	bHLH	1.90E-21
96	comp169328_g1_i2	bHLH	1.80E-21
97	comp161975_g1_i2	bZIP	1.20E-07
98	comp165325_g2_i1	B3	1.90E-17
99	comp159504_g1_i1	B3	6.80E-08
100	comp149163_g1_i1	B3	3.20E-17
101	comp180967_g1_i1	BES1	2.80E-21
102	comp131611_g1_i1	C3H	4.20E-08

103	comp154387_g1_i3	C3H	2.80E-34
104	comp115071_g1_i3	C3H	1.30E-17
105	comp172655_g1_i2	CAMTA	2.30E-20
106	comp190337_g1_i3	S1Fa-like	9.70E-28
107	comp172680_g1_i4	ERF	8.90E-15
108	comp167218_g1_i2	Dof	9.80E-19
109	comp175907_g1_i1	ERF	4.90E-09
110	comp183024_g1_i2	FAR1	3.20E-09
111	comp154405_g1_i2	FAR1	5.40E-26
112	comp161284_g2_i2	GRAS	4.80E-18
113	comp161284_g2_i3	GRAS	4.70E-18
114	comp183358_g1_i2	Nin-like	4.40E-09
115	comp187410_g2_i2	ARR-B	7.90E-08
116	comp187410_g1_i5	ARR-B	1.60E-08
117	comp187410_g1_i11	ARR-B	1.70E-08
118	comp164678_g1_i3	FAR1	2.50E-146
119	comp170487_g1_i3	LBD	2.20E-29
120	comp173611_g3_i7	LBD	6.80E-31
121	comp181033_g2_i5	M-type	1.60E-13
122	comp163408_g1_i1	MYB	8.40E-60
123	comp168860_g1_i4	MYB	9.30E-12
124	comp149649_g1_i2	MYB	5.10E-76
125	comp158180_g1_i1	MYB_related	3.60E-59
126	comp156701_g1_i2	MYB_related	3.80E-59
127	comp140515_g1_i2	C3H	4.40E-17
128	comp150486_g1_i2	C3H	1.30E-09
129	comp160199_g1_i2	NAC	9.80E-22
130	comp164510_g1_i1	NAC	1.20E-69
131	comp166071_g1_i1	NAC	1.00E-29
132	comp175390_g1_i5	NF-YA	2.00E-36
133	comp188608_g1_i3	NF-YC	1.20E-12
134	comp180612_g1_i1	Nin-like	1.00E-10
135	comp150399_g1_i2	MYB	1.60E-10
136	comp193661_g2_i1	C3H	5.50E-11
137	comp193661_g2_i8	C3H	5.50E-11

138	comp189206_g1_i4	SRS	4.30E-15
139	comp187823_g1_i4	SRS	5.00E-15
140	comp161615_g1_i2	B3	3.20E-11
141	comp180434_g1_i2	GeBP	6.50E-12
142	comp167416_g1_i2	WOX	7.10E-22
143	comp169479_g1_i3	WRKY	1.90E-09
144	comp169479_g1_i7	WRKY	2.60E-10
145	comp168795_g2_i2	WRKY	3.80E-19
146	comp155632_g1_i1	WRKY	6.10E-21
147	comp164501_g1_i1	YABBY	1.70E-28
148	comp164501_g2_i2	YABBY	1.60E-28
149	comp167639_g1_i2	bHLH	2.60E-15
150	comp157295_g1_i2	bHLH	9.20E-17
151	comp162250_g1_i2	bHLH	4.60E-27
152	comp163987_g1_i9	bHLH	1.80E-21
153	comp158804_g1_i1	bZIP	1.10E-12

(b) List of other genes names (681 transcripts)

No.	Seqname of *Euglena*	Annotation	Evalue
1	comp188033_g1_i2	SAMC2_ARATH	5.00E-22
2	comp185228_g1_i1	family transcriptional regulator	4.40E-22
3	comp181618_g1_i2	PREDICTED: uncharacterized protein LOC100179133	1.20E-24
4	comp172315_g1_i1	kinase domain	7.50E-14
5	comp170891_g1_i1	RLF_ARATH	2.30E-18
6	comp178385_g1_i1	ubiquinone biosynthesis	9.70E-81
7	comp177478_g1_i1	Zinc CCCH-type	9.80E-46
8	comp153609_g1_i2	palmitoyltransferase ZDHHC15 isoform X2	3.90E-28
9	comp193804_g1_i5	GLCD_BACSU	1.20E-29
10	comp194021_g1_i3	conserved unknown protein	1.00E-16
11	comp190939_g2_i2	FH13_ARATH	1.10E-34
12	comp164344_g1_i4	formamidopyrimidine-DNA glycosylase	1.30E-92
13	comp193370_g2_i6	NEK1_MOUSE	1.10E-66
14	comp192134_g1_i2	EGY1_ARATH	1.50E-26
15	comp175933_g2_i1	serine threonine kinase	2.60E-09

16	comp185590_g1_i1	TRI56_HUMAN	8.00E-09
17	comp192399_g1_i15	Hypothetical protein, putative	3.30E-16
18	comp191461_g1_i5	KHA1_YEAST	2.30E-24
19	comp166051_g2_i2	AVT5_SCHPO	2.50E-55
20	comp188049_g1_i5	amino acid permease	5.10E-65
21	comp188559_g1_i2	hypothetical protein AURANDRAFT_69315	1.40E-38
22	comp175967_g1_i4	predicted protein, partial	4.00E-40
23	comp181588_g1_i2	DNJC7_DICDI	1.60E-60
24	comp189653_g2_i2	helicase and polymerase-containing TEBICHI isoform X1	3.30E-176
25	comp188461_g1_i1	PR40A_ARATH	4.30E-10
26	comp190674_g1_i1	MPP10_HUMAN	6.70E-25
27	comp194609_g1_i4	phosphate carrier , mitochondrial	6.70E-50
28	comp155365_g1_i4	rho GDP-dissociation inhibitor 1-like	4.10E-09
29	comp194856_g3_i3	KPRS_HELPY	2.10E-37
30	comp193924_g1_i3	dynein heavy chain 5, axonemal	0.00E+00
31	comp191532_g1_i3	PUB1_SCHPO	1.20E-43
32	comp173792_g1_i2	PAH2_ARATH	1.80E-64
33	comp182924_g1_i1	arylsulfate sulfotransferase	1.10E-08
34	comp195476_g1_i2	phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase and dual-specificity phosphatase PTEN isoform X2	1.80E-62
35	comp194362_g2_i4	transmembrane	2.50E-13
36	comp192683_g1_i8	dnaJ homolog 1, mitochondrial-like	7.70E-106
37	comp193895_g1_i6	cyclic nucleotide-binding domain	1.60E-08
38	comp190709_g1_i1	ABC transporter substrate-binding	1.80E-22
39	comp152332_g1_i1	AZOR_AROAE	1.20E-23
40	comp194968_g1_i1	hypothetical protein	2.70E-11
41	comp183256_g1_i3	DPYD_HUMAN	0.00E+00
42	comp192547_g1_i3	EFP_SYNE7	1.30E-55
43	comp153597_g1_i2	phospholipase	4.40E-33
44	comp172408_g1_i1	PHMT_MYCLE	2.70E-30
45	comp179643_g2_i1	Adenylate guanylate cyclase with Chase sensor	1.80E-10
46	comp191782_g2_i2	PLK5_MOUSE	9.50E-10
47	comp185493_g2_i6	hypothetical protein	1.90E-11
48	comp155291_g1_i2	tetratricopeptide repeat 1	1.20E-23
49	comp158289_g1_i3	pterin-4-alpha-carbinolamine dehydratase	3.10E-17

50	comp196310_g3_i8	expression site-associated gene 4 (ESAG4)	0.00E+00
51	comp192355_g1_i3	WECH_DROME	1.50E-07
52	comp193064_g1_i5	hypothetical protein	2.20E-58
53	comp180831_g1_i3	kinase domain	1.10E-11
54	comp188045_g2_i2	predicted protein	6.90E-19
55	comp195049_g1_i5	CID11_SCHPO	1.80E-14
56	comp195863_g1_i6	adenylate cyclase	7.20E-13
57	comp175038_g1_i3	Tetratricopeptide-like helical	2.70E-46
58	comp189576_g1_i2	importin-11 isoform X1	1.70E-72
59	comp193328_g2_i6	E3 ubiquitin- ligase ARIH2-like	1.00E-25
60	comp180709_g1_i3	UVH3_ARATH	4.40E-37
61	comp195665_g1_i7	MAV_MYCA1	3.20E-09
62	comp184148_g1_i1	Choline dehydrogenase, mitochondrial	1.10E-31
63	comp185125_g1_i8	PAS domain-containing sensor histidine kinase	2.70E-14
64	comp192089_g2_i1	GP107_MOUSE	3.80E-73
65	comp193947_g1_i1	hypothetical protein STCU_06761	4.60E-08
66	comp193215_g2_i2	AARA_DICDI	1.90E-07
67	comp188625_g1_i7	predicted protein, partial	4.00E-53
68	comp194189_g1_i1	5NTD_LUTLO	3.30E-32
69	comp158867_g2_i2	F26_ARATH	3.50E-66
70	comp195602_g1_i4	Multi-sensor Hybrid Histidine Kinase	4.30E-07
71	comp180626_g1_i4	Os03g0719500predicted protein, partial	5.10E-159
72	comp161194_g1_i6	SSRP1_XENLA	9.40E-45
73	comp191477_g8_i3	E3 ubiquitin- ligase sina	2.70E-10
74	comp187915_g1_i2	MYCE_MICGR	3.60E-12
75	comp178203_g2_i1	CC067_RATRecName: Full=Uncharacterized protein C3orf67 homolog	4.10E-23
76	comp193581_g2_i3	hybrid sensor histidine kinase response regulator	7.30E-10
77	comp189825_g1_i4	multiple C2 and transmembrane domain-containing 2-like	1.60E-09
78	comp175448_g1_i9	hypothetical protein CHLNCDRAFT_138448	4.20E-10
79	comp195126_g1_i5	hypothetical protein PTSG_12820	4.60E-12
80	comp187345_g1_i1	BROX_RAT	3.40E-10
81	comp190339_g1_i6	pilus assembly	8.80E-86
82	comp178996_g2_i4	ribulose-1,5 bisphosphate carboxylase oxygenase large subunit N-methyltransferase, chloroplastic isoform X2	4.00E-39
83	comp176563_g1_i2	ubiquinone biosynthesis methyltransferase	1.70E-14

84	comp194006_g1_i2	PDE11_RAT	1.90E-24
85	comp178385_g1_i4	ubiquinone biosynthesis	1.30E-80
86	comp158289_g1_i1	pterin-4-alpha-carbinolamine dehydratase	3.00E-17
87	comp177489_g1_i1	kinase domain	1.80E-10
88	comp192478_g1_i3	multi-sensor hybrid histidine kinase	6.00E-35
89	comp186722_g2_i1	SDRD_STAAW	9.30E-07
90	comp180933_g1_i1	PSAA_EUGGR	0.00E+00
91	comp173258_g2_i3	ABC transporter f family member 2	0.00E+00
92	comp187558_g1_i4	predicted protein	1.30E-09
93	comp181821_g1_i1	probable G- coupled receptor 157	8.20E-32
94	comp186944_g1_i3	HCP_GEOMG	1.60E-139
95	comp193906_g1_i3	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07
96	comp175262_g1_i2	replication factor A1	1.90E-85
97	comp195006_g1_i5	hybrid sensor histidine kinase response regulator	9.30E-21
98	comp171049_g1_i4	MFS general substrate transporter	8.40E-44
99	comp181191_g1_i6	peptidyl-tRNA hydrolase	4.50E-15
100	comp152060_g1_i3	NLRC3_HUMAN	3.60E-34
101	comp185001_g2_i3	Y4757_DICDI	1.40E-09
102	comp183092_g1_i2	QUIRKY-like	7.10E-23
103	comp185000_g1_i1	predicted protein, partial	1.10E-16
104	comp179644_g1_i3	intraflagellar transport IFT88	0.00E+00
105	comp195174_g2_i2	calpain-like cysteine peptidase	1.40E-10
106	comp186936_g3_i1	CB21_EUGGR	1.10E-23
107	comp181633_g1_i3	ZDHC2_MOUSE	6.10E-43
108	comp153567_g1_i3	hypothetical protein GUITHDRAFT_166518	1.10E-23
109	comp191002_g2_i2	PUTP_SALTY	5.80E-10
110	comp194462_g1_i10	Transcription factor	8.30E-56
111	comp171755_g1_i3	lysine decarboxylase domain-containing	1.60E-100
112	comp177377_g1_i4	NLRC3_HUMAN	4.30E-20
113	comp186188_g1_i6	predicted protein	2.10E-18
114	comp195480_g1_i4	parp domain-containing	3.40E-11
115	comp189471_g1_i2	KLHDB_DROSI	1.10E-18
116	comp175946_g1_i1	SIR2 family transcriptional regulator	1.60E-94
117	comp190617_g2_i3	SMT1_ARATH	7.90E-97
118	comp190664_g1_i2	ACADS_RAT	5.10E-167

119	comp185779_g1_i2	NLRC3_HUMAN	2.40E-23
120	comp190604_g2_i3	RIMP_THEEB	2.40E-15
121	comp196310_g3_i1	CYAA_TRYEQ	1.00E-46
122	comp189538_g1_i7	PAS domain-containing sensor histidine kinase	2.00E-26
123	comp189812_g1_i11	PREDICTED: uncharacterized protein LOC106806144	2.30E-20
124	comp187981_g2_i2	WD-40 repeat	2.00E-18
125	comp183520_g1_i2	PIBF1_HUMAN	4.30E-12
126	comp194338_g1_i7	MFS12_MOUSE	1.20E-45
127	comp175309_g1_i1	Neurotrypsin	3.10E-07
128	comp170308_g2_i1	DGK1_ARATH	2.70E-14
129	comp195163_g1_i5	PREDICTED: uncharacterized protein LOC103949034 isoform X1	2.00E-07
130	comp195607_g1_i2	polyadenylate-binding 5	1.10E-16
131	comp194353_g2_i2	AChain A, Crystal Structure Of Engineered Northeast Structural Genomics Consortium Target	1.70E-17
132	comp188088_g1_i1	IF4E_DICDI	1.30E-14
133	comp178888_g1_i5	IF4E_RAT	1.40E-45
134	comp168732_g1_i2	desumoylating isopeptidase 2	1.90E-32
135	comp129541_g1_i1	HOS4_YEAST	1.90E-07
136	comp185046_g1_i3	SULT2_CHLRE	1.20E-37
137	comp192327_g1_i2	amino acid permease	3.60E-74
138	comp185091_g1_i1	BRAP_HUMAN	1.30E-48
139	comp187373_g1_i2	TBC domain-containing predicted protein, partial	1.80E-84
140	comp192048_g1_i3	AK1_DICDI	5.30E-40
141	comp183496_g1_i1	ATR_ORYSJ	0.00E+00
142	comp179154_g1_i1	RPC5_MOUSE	3.10E-21
143	comp173380_g1_i7	PUM5_ARATH	2.10E-60
144	comp183078_g2_i3	alpha beta hydrolase fold	5.60E-15
145	comp180704_g1_i2	ARLY_PARD8	1.90E-88
146	comp156380_g1_i2	MFS transporter, FLVCR family, disrupted in renal carcinoma 2	1.20E-58
147	comp189117_g1_i3	peptidase M14, carboxypeptidase A	1.80E-74
148	comp191050_g2_i3	ankyrin repeat	6.50E-18
149	comp180131_g1_i2	DNA-directed RNA polymerase isoform 1	2.60E-148
150	comp175839_g1_i2	FBX11_RAT	7.50E-22
151	comp190793_g2_i4	glycoside hydrolase	1.10E-92
152	comp162679_g1_i2	testis-expressed sequence 11 -like	5.50E-10

153	comp192798_g2_i2	N-lysine methyltransferase SETD6 isoform X2	6.00E-39
154	comp180057_g1_i2	GLOD4_MOUSE	1.10E-62
155	comp164077_g1_i2	CEMA_NEPOL	2.50E-24
156	comp181089_g1_i2	predicted protein	8.00E-11
157	comp189538_g1_i3	PAS domain-containing sensor histidine kinase	1.90E-26
158	comp182900_g1_i1	GTP cyclohydrolase I	1.90E-68
159	comp182507_g1_i17	DDL_RALME	2.60E-09
160	comp191923_g1_i8	serine threonine- kinase STK11 isoform X1	5.00E-64
161	comp194536_g1_i6	DYH6_HUMAN	0.00E+00
162	comp193386_g1_i2	vacuolar sorting-associated 13 family	2.50E-40
163	comp188625_g1_i9	predicted protein, partial	3.70E-53
164	comp192932_g1_i1	E3 ubiquitin- ligase HECTD1 isoform X4	3.80E-17
165	comp154198_g1_i3	Hypothetical protein, putative	2.50E-09
166	comp191132_g1_i2	PANK2_ORYSJ	4.70E-48
167	comp184429_g1_i3	NLRC3 isoform X1	3.70E-21
168	comp173221_g1_i1	GEMI2_MOUSE	2.80E-08
169	comp190378_g2_i3	TANC1_RAT	7.90E-14
170	comp176840_g3_i4	KAD_DESMR	2.00E-110
171	comp174505_g1_i1	Poly(ADP-ribose) polymerase catalytic domain containing	2.20E-89
172	comp180789_g2_i3	gamma-taxilin isoform X1	7.80E-24
173	comp187263_g2_i8	CYA1_MYCBO	2.60E-17
174	comp182271_g1_i3	ABCD4_MOUSE	9.50E-67
175	comp185776_g1_i2	KLHDB_CULQU	5.20E-30
176	comp181796_g1_i8	RSH1C_ARATH	3.20E-56
177	comp188738_g1_i7	MTA70_DROME	2.60E-33
178	comp172228_g1_i5	dual specificity phosphatase 7	1.30E-46
179	comp190923_g3_i3	NLRC3_HUMAN	2.30E-40
180	comp182421_g1_i4	Squamous cell carcinoma antigen recognized by T-cells 3	4.50E-38
181	comp192137_g2_i2	ABC transporter substrate-binding	7.10E-39
182	comp195163_g1_i2	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07
183	comp182993_g1_i2	EGT2_SCHPO	1.00E-25
184	comp185373_g1_i2	DGLA_HUMAN	6.80E-40
185	comp155365_g1_i5	rho GDP-dissociation inhibitor 1-like	3.60E-09
186	comp164725_g1_i3	PREDICTED: uncharacterized protein LOC105646130	3.40E-18
187	comp195579_g1_i1	hypothetical protein PBRA_000501	1.30E-26

188	comp175891_g2_i3	S38AB_MOUSE	2.10E-51
189	comp188625_g1_i3	NLRC3_MOUSE	1.60E-37
190	comp184418_g1_i4	Proteophosphoglycan ppg4	5.30E-35
191	comp173684_g1_i3	RANB9_XENTR	5.20E-15
192	comp172487_g1_i1	hypothetical protein, conserved	7.60E-09
193	comp174430_g2_i6	DUS1B_ARATH	1.90E-07
194	comp182891_g1_i1	PHR_POTTR	4.50E-108
195	comp187600_g2_i4	#NAME?	6.40E-72
196	comp183741_g1_i6	calmodulin mutant syncam9	8.20E-09
197	comp143255_g2_i1	RL34A_YEAST	1.20E-19
198	comp189660_g3_i6	FRO4_ARATH	4.30E-17
199	comp187000_g1_i11	GEFF_DICDI	2.30E-10
200	comp187069_g1_i4	CCD1_ONCHC	3.60E-68
201	comp192307_g1_i4	predicted protein	4.60E-17
202	comp164527_g1_i2	S14L1_HUMAN	4.10E-07
203	comp192990_g2_i3	FYV1_DICDI	2.90E-66
204	comp183383_g1_i1	LRC71_HUMAN	2.40E-34
205	comp181188_g1_i3	TIL_ARATH	4.60E-10
206	comp190889_g2_i1	DHKJ_DICDI	5.40E-25
207	comp186944_g1_i2	HCP_GEOMG	2.30E-139
208	comp189726_g1_i7	predicted protein	1.20E-47
209	comp196062_g1_i2	hypothetical protein EMIHUDRAFT_439379	8.40E-65
210	comp189170_g1_i2	hypothetical protein PTSG_08670	6.90E-32
211	comp177214_g4_i4	NLRC3_HUMAN	7.20E-52
212	comp174411_g1_i1	zinc finger DHHC domain-containing	3.70E-20
213	comp192932_g2_i3	E3 ubiquitin- ligase HECTD1 isoform X4	2.00E-17
214	comp176346_g1_i1	FUCM_MOUSE	7.40E-17
215	comp186111_g2_i12	STPG1_DANRE	1.10E-09
216	comp133109_g1_i2	YIW2_SCHPO	3.30E-42
217	comp163866_g1_i1	NUDC_CHICK	4.60E-60
218	comp185375_g2_i1	Eukaryotic translation initiation factor 4E-3	1.10E-19
219	comp179265_g1_i1	pseudouridylate synthase 7 homolog isoform X1	1.60E-79
220	comp186072_g1_i3	kinesin family member 1B	2.30E-27
221	comp195055_g5_i1	adenylate cyclase	5.80E-08
222	comp172315_g1_i2	kinase domain	8.10E-14

223	comp183496_g1_i4	ATR_ORYSI	0.00E+00
224	comp189538_g1_i6	FIXL_RHIME	1.10E-13
225	comp193586_g1_i2	ECHP_DANRE	9.20E-151
226	comp182021_g1_i2	glycosyltransferases (COG1216)	3.10E-19
227	comp194329_g1_i4	Y8013_DICDI	6.40E-11
228	comp180579_g1_i15	D-alanine--D-alanine ligase	3.30E-138
229	comp190285_g1_i3	GCSH_HERA2	2.00E-19
230	comp173360_g1_i3	lysine decarboxylase domain-containing	1.90E-100
231	comp186286_g2_i3	FHIT_DICDI	7.10E-43
232	comp181065_g1_i4	family	5.40E-20
233	comp184430_g3_i4	ATPD_TOBAC	1.40E-23
234	comp163867_g1_i1	SPXS1_DICDI	2.40E-28
235	comp189180_g1_i1	short transient receptor potential channel 4-associated -like isoform X2	2.40E-26
236	comp192778_g1_i11	nicotinate-nucleotide adenylyltransferase	2.90E-65
237	comp178742_g1_i2	insulin-degrading enzyme	2.20E-25
238	comp188298_g2_i5	NLRC3_HUMAN	1.90E-34
239	comp171920_g1_i1	TMEM5_HUMAN	2.40E-12
240	comp186473_g2_i8	PSP1_YEAST	2.70E-18
241	comp184416_g2_i3	EXGA_NEOFI	3.60E-27
242	comp188268_g1_i6	Y528_SYNY3	7.00E-09
243	comp188196_g2_i5	Cold-shock DEAD-box A	1.30E-41
244	comp192090_g2_i2	TNPO1_MOUSE	0.00E+00
245	comp178763_g2_i9	TTD14_DROME	2.30E-34
246	comp187411_g2_i7	LOW PSII ACCUMULATION chloroplastic	8.70E-41
247	comp175614_g1_i1	PTR36_ARATH	1.10E-32
248	comp191570_g1_i1	hypothetical protein PBRA_008009	4.00E-11
249	comp188566_g1_i4	MYCF_MICGR	1.00E-33
250	comp195810_g2_i3	Hypothetical protein, putative	8.80E-16
251	comp180530_g1_i6	SPOT_ECO57	2.80E-76
252	comp183754_g1_i1	DPOLB_BOVIN	5.20E-74
253	comp177139_g1_i1	AAE16_ARATH	9.20E-71
254	comp173764_g1_i5	NIPSNAP-related (ISS)	1.10E-20
255	comp170642_g1_i1	sodium-independent anion transporter	1.90E-62
256	comp179475_g1_i3	NLRC3_HUMAN	9.50E-16
257	comp170754_g2_i3	SMYD3_MOUSE	2.20E-07

258	comp190474_g1_i6	serine threonine- kinase GRIK2-like	1.20E-93
259	comp185767_g1_i11	ODF3B_DANRE	1.80E-14
260	comp190720_g1_i3	NLRC3_HUMAN	1.10E-12
261	comp178257_g1_i2	leucine rich repeat -containing	4.50E-27
262	comp192618_g2_i3	hypothetical protein MNEG_5574	1.90E-09
263	comp183613_g3_i8	IF4E_RAT	1.10E-46
264	comp193624_g1_i1	ADP-ribosyltransferase 1 precursor	1.90E-09
265	comp151007_g1_i3	TRX1_YEAST	3.60E-27
266	comp179626_g6_i2	RTF1_MOUSE	6.40E-13
267	comp191224_g1_i1	glutathione S-transferase C-terminal domain-containing	2.70E-09
268	comp173260_g1_i2	Major Facilitator Superfamily (MFS)	4.10E-31
269	comp181935_g1_i4	CABL2_MOUSE	1.80E-28
270	comp185498_g1_i5	eukaryotic elongation factor 2 kinase isoform X1	1.30E-20
271	comp170160_g1_i2	thioredoxin reductase	0.00E+00
272	comp186705_g1_i6	formate dehydrogenase	1.40E-13
273	comp191564_g1_i3	DUF229 domain containing	7.30E-14
274	comp185489_g1_i2	Hypothetical protein, putative	1.40E-55
275	comp170275_g1_i4	Poly (A) RNA polymerase cid14	5.90E-51
276	comp178860_g1_i5	PARG_DROME	3.10E-47
277	comp176522_g1_i3	PSAA_EUGGR	0.00E+00
278	comp185498_g1_i3	EF2K_RAT	4.60E-20
279	comp188714_g1_i4	FG-GAP repeat	5.20E-18
280	comp161583_g1_i1	ZRAB2_PONAB	4.80E-10
281	comp194688_g1_i6	AB31G_ARATH	4.10E-141
282	comp175153_g1_i2	transporter ,	1.00E-66
283	comp187787_g2_i2	probable plastid-lipid-associated 6, chloroplastic	6.20E-30
284	comp163063_g1_i1	N-formylglutamate amidohydrolase	6.10E-18
285	comp189557_g1_i3	adenylate guanylate cyclase	6.90E-35
286	comp182890_g2_i3	NUCL_XENLA	6.10E-08
287	comp182777_g1_i2	Y103_SYNY3RecName: Full=Uncharacterized protein sll0103	4.90E-08
288	comp127851_g1_i3	VIS_VIBS1	1.70E-07
289	comp193684_g1_i5	GEFF_DICDI	6.80E-15
290	comp195811_g1_i3	IF4E2_MAIZE	2.70E-09
291	comp185006_g1_i1	LVSC_DICDI	2.50E-36
292	comp180440_g1_i3	DXR_NOSS1	2.10E-174

293	comp164653_g1_i3	Hypothetical protein, putative	7.50E-12
294	comp186127_g1_i2	casein kinase	1.30E-57
295	comp183464_g2_i1	IFT88_MOUSE	0.00E+00
296	comp178417_g1_i3	predicted protein, partial	3.40E-25
297	comp191532_g1_i1	PUB1_SCHPO	8.10E-44
298	comp195163_g1_i18	PREDICTED: uncharacterized protein LOC103949034 isoform X1	1.90E-07
299	comp185453_g1_i3	SPOT_ECO57	1.10E-75
300	comp153984_g1_i1	zinc finger with UFM1-specific peptidase domain	3.60E-53
301	comp192177_g1_i6	KAPR_ASPFU	7.70E-08
302	comp193496_g2_i5	CHI11_METAN	1.60E-08
303	comp187682_g1_i1	GLF_MYCTU	1.20E-73
304	comp195449_g3_i1	ANR50_HUMAN	8.40E-09
305	comp180196_g2_i2	CE164_MOUSE	2.30E-25
306	comp189137_g2_i2	ZNFX1_MOUSE	1.90E-60
307	comp180948_g1_i4	Y2800_NOSS1	3.30E-15
308	comp187622_g1_i3	MYCA_BACIU	7.30E-111
309	comp187437_g1_i4	exostosin	1.20E-12
310	comp193946_g1_i3	CLCA_DICDI	8.40E-59
311	comp194383_g1_i8	hypothetical protein PTSG_10764	2.80E-16
312	comp169361_g1_i2	hypothetical protein	4.60E-14
313	comp195327_g2_i1	VP13A_DICDI	5.60E-18
314	comp191424_g2_i1	glutaredoxin-1 isoform X2	1.30E-102
315	comp188826_g2_i1	alpha- kinase 1-like	2.40E-22
316	comp180128_g1_i3	TBC1 domain family member 31	3.30E-28
317	comp190383_g2_i1	kinase domain	3.90E-11
318	comp185561_g1_i5	predicted protein, partial	0.00E+00
319	comp189629_g1_i1	PHF8_HUMAN	3.00E-07
320	comp177185_g1_i2	30S ribosomal S1	1.10E-64
321	comp194938_g1_i1	alpha-L-rhamnosidase	3.30E-62
322	comp248551_g1_i1	RR11_EUGGR	1.70E-54
323	comp185100_g1_i2	ASAT1_ARATH	5.50E-07
324	comp187278_g2_i2	receptor-type adenylate cyclase GRESAG 4,	4.50E-75
325	comp188659_g1_i2	ABC transporter permease	2.00E-25
326	comp195141_g1_i10	SLN1_YEAST	9.00E-14
327	comp149134_g1_i1	adenylate cyclase	4.00E-15

328	comp183695_g2_i4	coproporphyrinogen III oxidase	1.20E-123
329	comp187210_g2_i3	NEK1_MOUSE	2.50E-58
330	comp153984_g1_i2	ZUFSP_BOVIN	6.00E-35
331	comp191893_g1_i2	phospholipase c	1.30E-72
332	comp188826_g1_i7	predicted protein, partial	3.10E-22
333	comp184230_g2_i3	hypothetical protein M078_4933	1.70E-12
334	comp186127_g1_i1	KC1DA_DANRE	2.30E-54
335	comp178942_g1_i3	GALM_ACICA	9.60E-102
336	comp181209_g1_i1	CDC23_ARATH	2.20E-121
337	comp184507_g4_i5	cytosolic carboxypeptidase 1 isoform X3	2.80E-104
338	comp171953_g1_i3	TPL_PASMU	1.50E-119
339	comp153060_g1_i5	#NAME?	3.10E-125
340	comp189924_g2_i2	XRCC6_CHICK	2.20E-48
341	comp189320_g1_i2	e3 ubiquitin- ligase	3.20E-24
342	comp188870_g1_i2	KIN11_ARATH	1.50E-128
343	comp170027_g1_i1	Vesicle-associated membrane -associated Bpredicted protein, partial	2.60E-08
344	comp193939_g1_i8	KRAC_DICDI	2.90E-97
345	comp190038_g1_i4	RTCA_MOUSE	7.70E-75
346	comp189758_g2_i4	AARA_DICDI	2.10E-11
347	comp175440_g1_i1	SAV_SULAC	1.90E-16
348	comp177876_g1_i5	PRP31_DANRE	1.30E-63
349	comp185905_g1_i2	APM1A_ORYSJ	1.20E-25
350	comp170802_g1_i1	HMCN1_HUMAN	4.00E-08
351	comp194572_g1_i5	OHK1_ORYSJ	5.40E-15
352	comp185293_g1_i5	metallophosphoesterase domain-containing 1	5.90E-61
353	comp194387_g1_i2	ACAC_CHICK	0.00E+00
354	comp188583_g1_i3	CY43_TRYBB	2.20E-20
355	comp184602_g1_i4	CRTI_NARPS	0.00E+00
356	comp175093_g1_i4	zinc finger family	3.70E-11
357	comp170001_g1_i2	SMC4_MOUSE	1.60E-147
358	comp195858_g1_i1	proteasome activator complex subunit 4	1.40E-42
359	comp190304_g1_i6	hypothetical protein LOTGIDRAFT_190059	5.40E-07
360	comp187160_g1_i1	HINT3_RAT	1.10E-28
361	comp186463_g1_i2	Y4139_ARATH	2.10E-177
362	comp4325_g1_i1	YCX7_EUGGR	1.20E-169

363	comp184839_g1_i2	TTC28_MOUSE	1.60E-08
364	comp187966_g1_i6	PUB1_SCHPO	2.10E-43
365	comp173942_g1_i2	RAD51_HUMAN	0.00E+00
366	comp194293_g1_i3	IP6K3_MOUSE	2.50E-30
367	comp179921_g1_i1	LAAT1_CAEEL	1.00E-23
368	comp195290_g2_i3	photosystem I assembly Ycf3	6.40E-10
369	comp187787_g2_i3	PAP6_ARATH	2.20E-19
370	comp187467_g1_i3	probable glucuronoxylan glucuronosyltransferase F8H	9.30E-13
371	comp186127_g1_i8	casein kinase	1.00E-57
372	comp170275_g1_i8	Poly (A) RNA polymerase cid14	1.00E-50
373	comp195079_g1_i9	MET22_HUMAN	1.40E-12
374	comp189059_g1_i8	STK11_CHICK	3.90E-53
375	comp192670_g1_i4	Kinesin KIF13A	1.30E-27
376	comp182530_g1_i3	P2C80_ARATH	3.60E-29
377	comp188064_g1_i1	MHCKB_DICDI	4.40E-47
378	comp160211_g1_i4	pseudouridine synthase, family	1.00E-47
379	comp193624_g1_i2	ADP-ribosyltransferase 1 precursor	1.90E-09
380	comp191391_g1_i4	LIVB3_BRUSU	6.40E-07
381	comp176396_g1_i1	BRE1B_HUMAN	4.20E-21
382	comp186529_g1_i5	CHO2_ASPFU	3.20E-101
383	comp182543_g1_i2	DSF1_YEAST	3.00E-107
384	comp185493_g2_i17	hypothetical protein	1.50E-11
385	comp183887_g3_i5	hypothetical protein EMIHUDRAFT_197272	3.80E-16
386	comp196150_g1_i9	CYA1_RHIME	4.10E-37
387	comp169791_g2_i1	membrane-bound adenylyl cyclase	1.40E-115
388	comp190241_g1_i2	leucine-rich repeat-containing 49-like	2.50E-53
389	comp175514_g1_i4	penicillin-binding	5.40E-42
390	comp184647_g2_i7	tRNA pseudouridine synthase A	8.70E-18
391	comp166448_g1_i2	60S ribosomal L26	1.50E-20
392	comp192395_g1_i2	QOR_PSEAE	2.20E-92
393	comp192028_g1_i8	unnamed protein product	5.10E-28
394	comp179170_g1_i3	DNA2_ACAPL	1.40E-34
395	comp181111_g1_i5	MYCBP-associated -like isoform X1	5.00E-38
396	comp182711_g1_i6	TRK2_YEAST	6.00E-38
397	comp182569_g3_i5	NUCL_MACFA	1.40E-14

398	comp192042_g1_i2	COG5_MOUSE	2.70E-55
399	comp184084_g3_i1	kinase domain	1.30E-12
400	comp102129_g1_i2	ADP-ribosyltransferase 1 precursor	8.90E-29
401	comp179846_g1_i2	tankyrase isoform X2	2.50E-12
402	comp192786_g2_i8	PAS domain-containing sensor histidine kinase	2.80E-13
403	comp161194_g1_i4	FACT complex subunit SSRP1-like	1.10E-44
404	comp182369_g4_i2	feline leukemia virus subgroup C receptor-related 2	3.90E-29
405	comp195636_g1_i4	transient receptor potential cation channel subfamily V member 6-like isoform X2	5.80E-56
406	comp185293_g1_i7	metallophosphoesterase domain-containing 1	3.10E-61
407	comp174058_g1_i1	thioredoxin	2.40E-28
408	comp193557_g2_i2	RINI_PIG	8.40E-08
409	comp180579_g1_i6	D-alanine--D-alanine ligase	1.40E-138
410	comp173647_g1_i4	TR164_ORYSJ	5.80E-14
411	comp192456_g3_i7	CORA_THEMA	5.60E-28
412	comp190947_g2_i1	cation-chloride cotransporter family	0.00E+00
413	comp185532_g2_i3	membrane-associated protein, putative	2.50E-11
414	comp157268_g1_i3	RMGL_PICST	8.20E-22
415	comp171450_g1_i1	YCX2_EUGGR	0.00E+00
416	comp175741_g1_i1	deSI At4g17486	1.20E-18
417	comp191711_g3_i1	flavincontaining amine oxidasedehydrogenase,	1.80E-174
418	comp192486_g1_i13	alpha subunit	1.60E-19
419	comp177726_g1_i2	kinase domain	8.30E-16
420	comp176696_g1_i1	KAPS_SCHPO	2.90E-82
421	comp188453_g1_i10	alpha-L-rhamnosidase	1.70E-13
422	comp186164_g4_i11	Armadillo betacatenin-like repeat domain containing	1.10E-24
423	comp188728_g1_i5	FANCM_HUMAN	1.60E-122
424	comp182900_g1_i2	GTP cyclohydrolase I	2.30E-68
425	comp189693_g2_i5	CYAA_TRYEQ	6.20E-31
426	comp195012_g1_i1	AT133_HUMAN	3.80E-32
427	comp185411_g2_i4	FER_GLEJA	1.30E-10
428	comp184942_g1_i7	AB37G_ARATH	8.10E-13
429	comp192353_g1_i6	MCES_SCHPO	1.90E-09
430	comp195323_g1_i3	ANM7_XENLA	2.20E-18
431	comp171128_g1_i1	ABCA1_DICDI	1.50E-10

432	comp174809_g1_i1	unnamed protein product	3.00E-11
433	comp177898_g1_i3	duf676 domain containing hydrolase	8.80E-21
434	comp176287_g1_i2	DGDG1_LOTJA	4.10E-113
435	comp177703_g3_i1	tubulin epsilon chain	4.50E-157
436	comp188033_g2_i7	SAMC2_ARATH	7.70E-22
437	comp153204_g1_i2	hypothetical protein Naga_100030g15	2.00E-09
438	comp176727_g1_i4	phage shock A	7.00E-68
439	comp192676_g1_i9	Carbohydrate-binding and sugar hydrolysis	1.80E-08
440	comp189726_g1_i1	predicted protein	1.50E-47
441	comp193424_g1_i5	transporter	1.80E-37
442	comp163635_g1_i1	amino acid transporter	1.00E-13
443	comp186014_g2_i1	6GAL_HYPRU	3.10E-27
444	comp188033_g1_i9	S-adenosylmethionine carrier 1, chloroplastic mitochondrial	1.40E-23
445	comp179244_g1_i3	ESIB_ECOL6	2.00E-18
446	comp182670_g1_i5	kinase domain	1.80E-15
447	comp176704_g1_i1	predicted protein	2.70E-49
448	comp150619_g1_i1	REEP6_DANRE	4.90E-11
449	comp166967_g1_i9	GP157_MOUSE	1.90E-08
450	comp173304_g1_i1	histone RNA hairpin-binding	3.80E-19
451	comp158868_g1_i3	MBF1A_ARATH	3.00E-39
452	comp179480_g2_i2	ALG14_YARLI	4.70E-41
453	comp182634_g1_i7	CACP_CANTR	7.10E-37
454	comp192128_g4_i2	CMGC CDKL kinase	2.40E-71
455	comp182313_g2_i1	glycosyltransferase aer61	2.30E-41
456	comp165751_g1_i3	TM165_HUMAN	4.20E-48
457	comp181809_g1_i3	anaphase-promoting complex subunit 4	1.20E-11
458	comp181589_g1_i10	3-phosphoinositide dependent kinase-1	3.50E-86
459	comp175354_g1_i5	armadillo repeat-containing 2-like	2.60E-32
460	comp173214_g1_i2	C2 domain-containing	2.50E-09
461	comp191699_g1_i6	F2612_CAEEL	4.00E-62
462	comp181326_g1_i4	L-asparaginase I	1.20E-92
463	comp160021_g1_i2	plastid terminal oxidase	1.50E-92
464	comp186126_g5_i2	NLRC3_MOUSE	3.30E-13
465	comp184700_g1_i1	PBUO_BACSU	2.30E-51
466	comp168095_g1_i1	TTLL1_TETTS	3.90E-56

467	comp165023_g1_i3	FDHC_METTF	1.30E-19
468	comp176383_g1_i3	hypothetical protein GUITHDRAFT_162439	9.90E-08
469	comp186560_g5_i9	XYLB_MOUSE	3.10E-115
470	comp151802_g1_i2	transmembrane 56 isoform X1	6.80E-21
471	comp185679_g1_i2	nicotinate-nucleotide adenylyltransferase	5.40E-52
472	comp180616_g3_i2	THI4_CHLRE	2.60E-94
473	comp159242_g1_i2	major facilitator superfamily	4.20E-90
474	comp188423_g1_i2	-glutamine gamma-glutamyltransferase 4	7.40E-73
475	comp190947_g1_i6	cation-chloride cotransporter family	0.00E+00
476	comp177753_g1_i5	XRP2	7.40E-49
477	comp186810_g1_i7	phosphatidylinositol class b	1.30E-52
478	comp172015_g1_i4	WD repeat-containing 43-like	1.50E-08
479	comp195384_g1_i2	DUG3_YEAST	9.00E-64
480	comp187406_g1_i4	hypothetical protein	2.20E-08
481	comp172656_g1_i2	TBCD9_MOUSE	1.30E-07
482	comp174941_g1_i2	ASM1_CAEEL	1.00E-08
483	comp194968_g1_i2	hypothetical protein	3.20E-11
484	comp175049_g1_i3	Mg(2+) transporter ALR2	2.30E-25
485	comp184547_g2_i3	F-box-like WD repeat-containing TBL1X	3.00E-152
486	comp187393_g7_i2	lysophospholipase	3.60E-60
487	comp190928_g3_i3	phosphohydrolase	3.30E-48
488	comp162171_g1_i1	SPEF1_XENLA	1.20E-34
489	comp165323_g2_i2	PSBA_EUGGR	3.50E-53
490	comp184944_g1_i3	Low temperature viability	1.10E-12
491	comp186411_g2_i6	rhodanese-like domain-containing	4.20E-16
492	comp180251_g1_i4	MHCKA_DICDI	2.20E-41
493	comp177170_g1_i3	Phosphoglycerate mutase family isoform 1	4.20E-51
494	comp174963_g1_i1	hypothetical protein PBRA_006659	9.90E-10
495	comp193820_g1_i8	FPRA1_CLOAB	9.60E-79
496	comp184319_g1_i2	HNRPR_HUMAN	2.10E-08
497	comp165740_g2_i1	homoisocitrate dehydrogenase	2.90E-81
498	comp188579_g1_i1	transmembrane protein, putative	1.60E-16
499	comp196279_g2_i5	ADCY1_MOUSE	7.10E-12
500	comp191441_g2_i3	MTP2_ORYSJ	1.20E-41
501	comp195722_g2_i1	EXPR_XANCP	9.00E-48

502	comp181872_g2_i5	Os01g0896500predicted protein, partial	3.50E-10
503	comp194643_g1_i3	calpain-like cysteine peptidase	1.60E-08
504	comp190525_g2_i1	GRIK2_ARATH	3.70E-46
505	comp193585_g2_i4	TRA2A_MOUSE	1.70E-07
506	comp187555_g2_i2	MHCKA_DICDI	6.60E-37
507	comp181260_g1_i2	S2540_MOUSE	8.00E-58
508	comp178308_g1_i5	Tetratricopeptide repeat	2.30E-24
509	comp188676_g3_i4	SPAG1_DANRE	1.10E-18
510	comp195431_g4_i1	Y1796_ARATH	2.50E-82
511	comp171902_g1_i1	TRPV6_HUMAN	3.00E-08
512	comp182564_g5_i1	RN165_HUMAN	4.70E-09
513	comp193673_g2_i1	ACSF4_MOUSE	8.20E-40
514	comp185640_g2_i2	ADP-ribosyltransferase 1 precursor	5.30E-37
515	comp190280_g1_i2	PRIPO_DANRE	8.70E-20
516	comp183649_g2_i5	glycosyl transferase	3.00E-19
517	comp191051_g1_i2	adenylate cyclase type 8 isoform X1	7.40E-15
518	comp183610_g1_i5	serine threonine kinase	2.00E-39
519	comp176007_g1_i4	predicted protein	2.60E-08
520	comp184179_g2_i1	RFWD2_MOUSE	2.40E-07
521	comp184586_g1_i4	ubiquitin- ligase	4.20E-54
522	comp162052_g1_i2	PIN1_ARATH	1.40E-37
523	comp184397_g6_i7	membrane-associated ,	5.20E-138
524	comp190165_g1_i6	Plastid lipid-associated fibrillin conserved domain	1.50E-117
525	comp196132_g1_i3	ef hand family	2.70E-36
526	comp192756_g1_i1	BAT1_ARATH	2.40E-77
527	comp175533_g2_i1	S35F5_PONAB	5.10E-41
528	comp186187_g1_i5	CHO2_ASPFU	2.90E-101
529	comp188721_g3_i2	DIOXL_ARATH	2.50E-42
530	comp177371_g1_i8	sodium hydrogen exchanger	1.60E-57
531	comp186072_g1_i8	KIF1C_HUMAN	1.90E-26
532	comp178846_g1_i1	nuclear lim interactor-interacting factor	5.80E-19
533	comp194917_g2_i7	F184A_HUMAN	1.20E-11
534	comp192503_g1_i3	ITPR3_BOVIN	1.60E-19
535	comp180128_g1_i8	TBC1 domain family member 31	3.00E-28
536	comp195833_g1_i4	ALA3_ARATH	1.30E-162

537	comp157236_g1_i2	ITPA_SORBI	1.80E-86
538	comp179437_g1_i1	DUF4436 domain-containing	4.60E-17
539	comp171677_g1_i2	IFT74_MOUSE	6.80E-74
540	comp178442_g1_i6	CRTY_ARTPT	2.90E-95
541	comp166606_g1_i2	PIRL5_ORYSJ	1.10E-12
542	comp186797_g2_i1	microtubule-associated 4 isoform X13	3.50E-24
543	comp190117_g3_i2	AK1_DICDI	9.90E-47
544	comp192852_g1_i4	RBBP6_HUMAN	2.50E-21
545	comp176474_g1_i3	amidohydrolase	1.60E-72
546	comp189526_g1_i1	S18B1_MOUSE	1.20E-07
547	comp192121_g2_i7	flagellar associated	1.20E-13
548	comp191532_g1_i9	PUB1_SCHPO	2.60E-43
549	comp178959_g1_i1	hypothetical protein PBRA_003427	7.10E-101
550	comp187129_g1_i2	ATPG_ODOSI	8.20E-77
551	comp193577_g1_i3	SPXS1_DICDI	4.60E-45
552	comp142981_g1_i3	PNC1_ARATH	6.90E-25
553	comp178430_g1_i6	acetyl- acetyltransferase	5.60E-90
554	comp193627_g1_i7	EF-hand calcium-binding domain-containing 6	5.50E-12
555	comp195386_g1_i3	hypothetical protein AURANDRAFT_67841	3.80E-17
556	comp191891_g5_i1	Kynurenine 3-monooxygenase and related flavo monooxygenases (ISS)	8.70E-48
557	comp187790_g1_i13	PYR1_SQUAC	0.00E+00
558	comp186214_g1_i4	CG30118-PApredicted protein, partial	1.50E-74
559	comp195305_g1_i10	peptidyl-prolyl cis-trans isomerase FKBP62-like isoform X2	3.40E-23
560	comp154889_g1_i1	YRHG_BACSU	1.00E-19
561	comp180250_g1_i8	ubiquitin family	6.30E-92
562	comp182174_g1_i2	maleylacetoacetate isomerase isoform X1	7.10E-56
563	comp194710_g1_i12	MTFP1_CAEEL	7.80E-11
564	comp177871_g1_i3	TB10C_MOUSE	8.10E-07
565	comp189120_g1_i3	DYH10_HUMAN	7.20E-08
566	comp181898_g1_i1	hypothetical protein ACA1_171730	2.60E-21
567	comp187028_g2_i10	YHM3_SCHPO	9.80E-08
568	comp193816_g1_i1	ferric reductase	1.60E-36
569	comp194614_g3_i6	VASH1_HUMAN	5.50E-46
570	comp186072_g2_i4	RA51C_CRIGR	3.70E-23
571	comp158874_g1_i2	bacterial leucyl aminopeptidase	1.20E-07

572	comp180043_g1_i1	GPAN1_CAEEL	2.60E-10
573	comp194598_g2_i2	unnamed protein product	9.00E-11
574	comp140373_g1_i1	PSBK_EUGGR	2.50E-15
575	comp177776_g1_i1	ALKB8_XENTR	1.50E-13
576	comp191631_g7_i1	membrane-associated protein, putative	4.80E-09
577	comp184360_g1_i2	unnamed product	5.90E-21
578	comp183454_g1_i1	THLA_PEPD6	1.00E-98
579	comp187994_g4_i3	HCNC_PSEAE	1.00E-14
580	comp187660_g1_i10	predicted protein	5.60E-14
581	comp179939_g1_i3	DDL_BORPD	2.10E-13
582	comp195456_g1_i1	kinase domain	7.00E-28
583	comp181661_g2_i4	phosphatase 1 regulatory subunit 42	2.80E-33
584	comp181083_g2_i2	EXD1_MOUSE	3.30E-10
585	comp180783_g1_i2	sulfite oxidase, mitochondrial	1.30E-156
586	comp186376_g1_i5	AT74_ARATH	9.90E-23
587	comp191510_g1_i2	enhancer of polycomb group	5.20E-43
588	comp183142_g1_i2	predicted protein	1.50E-73
589	comp179938_g2_i2	hypothetical protein SDRG_13758	1.70E-09
590	comp175716_g1_i2	Tetratricopeptide-like helical	2.50E-46
591	comp194522_g3_i1	NOXA_DICDI	2.90E-24
592	comp182725_g1_i6	PABP5_ARATH	3.30E-15
593	comp193782_g2_i2	PGL1B_ARATH	6.60E-09
594	comp164194_g1_i2	PDLI5_RAT	4.90E-15
595	comp183911_g1_i3	mitochondrial carrier family	1.90E-72
596	comp161585_g1_i3	TBCA_ARATH	7.00E-08
597	comp181796_g1_i4	RSH1C_ARATH	3.40E-56
598	comp188762_g1_i1	PREDICTED: uncharacterized protein LOC106175552	5.90E-42
599	comp187053_g1_i2	6-phosphofructo-2-kinase fructose-2,6-bisphosphatase 4-like isoform X1	1.20E-58
600	comp174154_g1_i2	CAPTC_DICDI	2.50E-31
601	comp182075_g1_i6	TRXF_ORYSJ	7.70E-19
602	comp182476_g1_i3	F-box only 11	6.40E-48
603	comp173974_g2_i1	enoyl- hydratase	4.60E-60
604	comp193939_g1_i5	KRAC_DICDI	2.10E-97
605	comp182083_g2_i8	PSR2_YEAST	5.50E-15
606	comp186705_g1_i4	formate dehydrogenase	1.60E-13

607	comp187105_g1_i2	POLH_ARATH	4.70E-101
608	comp189464_g2_i1	FPG_ARATH	2.40E-52
609	comp184814_g1_i6	TTD14_DROME	6.60E-35
610	comp166574_g1_i2	MGDP1_HUMAN	4.40E-40
611	comp189594_g1_i6	SGNH hydrolase- partial	4.90E-99
612	comp194349_g1_i12	kinase domain	1.20E-19
613	comp189812_g1_i1	PREDICTED: uncharacterized protein LOC106806144	5.30E-20
614	comp167689_g1_i6	sulfate transporter , chloroplastic-like	4.40E-81
615	comp180408_g1_i3	omega-amidase NIT2	1.90E-88
616	comp182521_g1_i2	probable phospholipase YOR022C	8.20E-19
617	comp186139_g1_i6	methyltransferase family	2.00E-29
618	comp182403_g2_i3	ALA3_ARATH	0.00E+00
619	comp184599_g1_i12	bifunctional metallophosphatase 5 -nucleotidase	7.10E-94
620	comp188286_g1_i6	PK2_DICDI	2.30E-90
621	comp195308_g1_i1	cytoplasmic dynein 1 heavy chain 1	0.00E+00
622	comp190541_g1_i2	MFS-type transporter SLC18B1-like	1.90E-10
623	comp192327_g1_i3	YG64_SCHPO	4.10E-44
624	comp176060_g1_i9	glycosyl transferase	2.00E-17
625	comp182848_g1_i6	SNARE associated Golgi	6.50E-50
626	comp181605_g1_i2	Hypothetical protein, putative	9.40E-42
627	comp192951_g1_i1	AChain A, Crystal Structure Of Engineered Northeast Structural Genomics Consortium Target	1.80E-13
628	comp55141_g1_i2	unnamed product	7.40E-12
629	comp189051_g1_i8	macrocin O-methyltransferase	6.00E-35
630	comp183611_g1_i5	SDRD_STAA8	2.80E-10
631	comp191923_g1_i5	STK11_CHICK	4.20E-53
632	comp152452_g1_i1	RH5_ORYSJ	4.40E-136
633	comp178651_g1_i6	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1 isoform X2	1.70E-84
634	comp188049_g1_i4	PHSA_ARATH	1.10E-40
635	comp180251_g1_i8	myosin heavy chain kinase	7.90E-42
636	comp184482_g1_i3	STT3A_ORYSJ	1.30E-21
637	comp189895_g1_i1	RPE_HUMAN	2.20E-82
638	comp171707_g1_i3	AK1_DICDI	5.40E-25
639	comp153483_g1_i3	hypothetical protein EMIHUDRAFT_247554	2.60E-36

640	comp187624_g1_i7	hypothetical protein COCSUDRAFT_45614	3.30E-14
641	comp190574_g1_i1	hypothetical protein EMIHUDRAFT_214712	6.50E-116
642	comp175946_g1_i2	SIR2 family transcriptional regulator	3.80E-97
643	comp190475_g1_i1	e3 ubiquitin- ligase	1.90E-27
644	comp185973_g1_i3	TMN7_ARATH	0.00E+00
645	comp174852_g1_i10	nuclear poly(A) polymerase 1-like	1.50E-63
646	comp190847_g2_i1	YHM3_SCHPO	1.40E-07
647	comp184457_g1_i1	E3 ubiquitin- ligase rififylin-like	1.00E-08
648	comp192919_g1_i3	karyopherin beta	5.90E-104
649	comp182908_g2_i2	unnamed product	3.90E-45
650	comp190203_g1_i7	MTP9_ARATH	1.10E-37
651	comp189571_g1_i1	AMP deaminase	2.30E-164
652	comp185751_g2_i8	Y1621_SYNY3	2.20E-73
653	comp187558_g1_i2	predicted protein	1.10E-09
654	comp166556_g1_i8	Os03g0719500predicted protein, partial	5.40E-158
655	comp195127_g1_i11	probable beta-1,3-galactosyltransferase 19	2.80E-16
656	comp186041_g1_i8	MAOX_VITVI	1.10E-134
657	comp185341_g1_i1	ERGI3_BOVIN	2.60E-108
658	comp173480_g1_i1	Endonuclease exonuclease phosphatase family	9.30E-27
659	comp170991_g1_i6	GSTT1_ARATH	1.70E-20
660	comp189196_g2_i4	methyl transferase	4.00E-29
661	comp192291_g1_i3	kinase A regulatory subunit	2.40E-67
662	comp192577_g2_i4	ZDHC1_ARATH	3.20E-22
663	comp185523_g1_i1	phospholipid-transporting ATPase IA	1.30E-33
664	comp191406_g1_i4	SRFR1_ARATH	6.10E-09
665	comp193571_g1_i1	BAT1_ARATH	2.30E-77
666	comp167689_g1_i4	sulfate transporter , chloroplastic-like	3.60E-81
667	comp162363_g3_i2	VPS39_HUMAN	2.30E-57
668	comp193855_g1_i1	kinase domain	2.40E-12
669	comp187098_g2_i1	AP4AT_MOUSE	2.30E-09
670	comp189522_g2_i4	MFS transporter	3.50E-09
671	comp184534_g1_i1	CYAB_STIAU	2.00E-11
672	comp175900_g1_i5	NLRC3_HUMAN	5.30E-44
673	comp190196_g1_i1	isoform A	2.00E-19
674	comp177152_g1_i5	Y575_SYNY3	9.80E-23

675	comp149586_g1_i1	AB1F_ARATH	0.00E+00
676	comp194729_g1_i5	SULT2_CHLRE	1.10E-151
677	comp192934_g1_i2	hypothetical protein NAEGRDRAFT_78864	2.20E-10
678	comp186384_g1_i5	SEC_ARATH	4.50E-21
679	comp192944_g1_i4	GRDP1_ARATH	5.30E-47
680	comp180401_g1_i2	N-acetylglucosaminyl transferase component	1.80E-29
681	comp178860_g1_i4	PARG_DROME	1.90E-47

Appendix Source code

The fold changes logarithm value.R

The fold changes logarithm value of Program

```
#Description :
#The fold changes logarithm value of TF families transcripts and other transcripts for BLSOM
        #loading ggplot2
        library(ggplot2)
        #shown The fold changes logarithm value of TF families transcripts and other #transcripts for BLSOM.
a=read.table("BLSOM.txt")
# aerobic and anaerobic condition
ggplot(a, aes(V2, V3)) +
   geom_point(aes(colour = V4), show.legend =TRUE) +
   xlab(NULL) + scale_x_continuous(breaks=seq(0,350,50))+
   ylab(NULL) + scale_y_continuous(breaks=seq(0,192,16))+
   labs(title = "aerobic and anaerobic",colour = "fold changes")+
   scale_colour_gradient2("fold changes",low ="blue",high ="red",limits=c(-15,15))+
   theme(panel.background = element_rect(fill = "black"),
            plot.title = element_text(size = 20),
            axis.text = element_text(color = "blue", size = 12),
            legend.title = element_text(size = 15),
            legend.text = element_text(size = 10))+
   guides(colour = guide_colourbar(direction="horizontal",
                                                   title.theme   =   element_text(size=15,   colour   =   "red",   angle   =   0),
title.position = "top"))
# wild type and bleached type
ggplot(a, aes(V2, V3)) +
   geom_point(aes(colour = V5), show.legend =TRUE) +
   xlab(NULL) + scale_x_continuous(breaks=seq(0,350,50))+
   ylab(NULL) + scale_y_continuous(breaks=seq(0,192,16))+
   labs(title = "wild type and bleached type",colour = "fold changes")+
   scale_colour_gradient2("fold changes",low ="blue",high ="red",limits=c(-15,15))+
   theme(panel.background = element_rect(fill = "black"),
            plot.title = element_text(size = 20),
```

```
          axis.text = element_text(color = "blue", size = 12),
          legend.title = element_text(size = 15),
          legend.text = element_text(size = 10))+
  guides(colour = guide_colourbar(direction="horizontal",
                                  title.theme = element_text(size=15, colour = "red", angle = 0),
title.position = "top"))
# midlogarithmic growth phase and late logarithmic growth phase
ggplot(a, aes(V2, V3)) +
  geom_point(aes(colour = V6), show.legend =TRUE) +
  xlab(NULL) + scale_x_continuous(breaks=seq(0,350,50))+
  ylab(NULL) + scale_y_continuous(breaks=seq(0,192,16))+
  labs(title = "midlogarithmic growth phase and late logarithmic growth phase",colour = "fold changes")+
  scale_colour_gradient2("fold changes",low ="blue",high ="red",limits=c(-15,15))+
  theme(panel.background = element_rect(fill = "black"),
          plot.title = element_text(size = 20),
          axis.text = element_text(color = "blue", size = 12),
          legend.title = element_text(size = 15),
          legend.text = element_text(size = 10))+
  guides(colour = guide_colourbar(direction="horizontal",
                                  title.theme = element_text(size=15, colour = "red", angle = 0),
title.position = "top"))
```

Input Data:

a BLSOM map for BLSOM software

Data format:

seqname	x	y	Fold change value of aerobic and anaerobic	Fold change value of wild type and bleached type	Fold change value of different growth phase
comp178174_g1_i1	0	3	-2.67476	-7.10421	-7.11449
comp178026_g3_i2	280	90	3.621236	3.006516	2.127862

Output result:

1. The fold changes logarithm value of TF families transcripts and other transcripts between

aerobic and anaerobic condition (Fig 3.14(a))

2. The fold changes logarithm value of TF families transcripts and other transcripts between wild type and bleached type (Fig 3.14(b))
3. The fold changes logarithm value of TF families transcripts and other transcripts midlogarithmic growth phase and late logarithmic growth phase. (Fig 3.14(c))

TF families transcripts and other gene transcripts.R

TF families transcripts and other gene transcripts of Program

```
#Description :
# Analysis of TF families transcripts and other gene transcripts for BLSOM.
        #loading ggplot2
        library(ggplot2)
# Analysis of TF families transcripts and other gene transcripts for BLSOM.
a=read.table("BLSOM.txt")
b=readlines("transcription factor.txt")
for (m in b$V2)
   {
          for (n in a$V1) {if (m==n) {a$V4=b$V1}}
}
ggplot(a, aes(V2, V3)) +
   geom_point(data = transform(a, V4 = NULL), aes(V2,V3), colour = "grey85")+
   geom_point(data = a,aes(V2,V3,colour = V4),na.rm = TRUE, show.legend =TRUE) +
   facet_wrap(~V4, nrow = 6) +
   theme_bw()+
   xlab(NULL) + scale_x_continuous(breaks=seq(0,350,350))+
   ylab(NULL) + scale_y_continuous(breaks=seq(0,192,192))+
   labs(title = "Transcription factors families and other genes",colour = "Transcription factors families")+
   theme(plot.title = element_text(size = 20),
            axis.text = element_text(color = "blue", size = 12),
            legend.title = element_text(size = 15),
            legend.text = element_text(size = 10),
            legend.key.size = unit(1, "cm"),
            legend.position = "bottom",
            legend.box = "horizontal",
            strip.text.x = element_text(size = 15))+
   guides(colour = guide_legend(nrow = 3,
                                              direction="horizontal",
                                              title.theme = element_text(size=15, colour = "red", angle = 0), title.position =
"top"))
```

Input Data:

1. BLSOM map

Data format:

seqname	x	y	Fold change value of aerobic and anaerobic	Fold change value of wild type and bleached type	Fold change value of different growth phase
comp178174_g1_i1	0	3	-2.67476	-7.10421	-7.11449
comp178026_g3_i2	280	90	3.621236	3.006516	2.127862

2. TF families transcripts annotation file

Data format:

Seqname	TF family
comp184981_g1_i1	MYB
comp184133_g1_i3	NAC

Output result:

Distribution of different TF families transcripts (Fig 3.15).

Co-expression patterns analyses.py

Co-expression patterns analyses of program

```
#Description :
# Co-expression patterns analyses for BLSOM.
    file1 = open("BLSOM.txt", "r")
    f1 = file1.readlines()
    file1.close()
    file2 = open("expression area.txt",'w')
    for f in f1:
        temp1 = f[:-1]
        temp2 = temp1.split('\t')
        a1= float(temp2[3])
        a2= float(temp2[4])
        a3= float(temp2[5])
        if a1>2:
            if a2>2:
                if a3>2:
                    temp3=temp1+'\t'+'1'
                elif a3<-2:
                    temp3=temp1+'\t'+'2'
                else:
                    temp3=temp1+'\t'+'3'
            elif a2<-2:
                if a3>2:
                    temp3=temp1+'\t'+'4'
                elif a3<-2:
                    temp3=temp1+'\t'+'5'
                else:
                    temp3=temp1+'\t'+'6'
            else:
                if a3>2:
                    temp3=temp1+'\t'+'7'
                elif a3<-2:
```

```
            temp3=temp1+'\t'+'8'
        else:
            temp3=temp1+'\t'+'9'
elif a1<-2:
    if a2>2:
        if a3>2:
            temp3=temp1+'\t'+'10'
        elif a3<-2:
            temp3=temp1+'\t'+'11'
        else:
            temp3=temp1+'\t'+'12'
    elif a2<-2:
        if a3>2:
            temp3=temp1+'\t'+'13'
        elif a3<-2:
            temp3=temp1+'\t'+'14'
        else:
            temp3=temp1+'\t'+'15'
    else:
        if a3>2:
            temp3=temp1+'\t'+'16'
        elif a3<-2:
            temp3=temp1+'\t'+'17'
        else:
            temp3=temp1+'\t'+'18'
else:
    if a2>2:
        if a3>2:
            temp3=temp1+'\t'+'19'
        elif a3<-2:
            temp3=temp1+'\t'+'20'
        else:
            temp3=temp1+'\t'+'21'
    elif a2<-2:
        if a3>2:
            temp3=temp1+'\t'+'22'
```

```
                elif a3<-2:
                    temp3=temp1+'\t'+'23'
                else:
                    temp3=temp1+'\t'+'24'
            else:
                if a3>2:
                    temp3=temp1+'\t'+'25'
                elif a3<-2:
                    temp3=temp1+'\t'+'26'
                else:
                    temp3=temp1+'\t'+'27'

        cn += 1
        file2.write(temp3+'\n')
    file2.close()
    print(file2)
    print(cn)
```

Input Data:

BLSOM map

Data format:

seqname	x	y	Fold change value of aerobic and anaerobic	Fold change value of wild type and bleached type	Fold change value of different growth phase
comp178174_g1_i1	0	3	-2.67476	-7.10421	-7.11449
comp178026_g3_i2	280	90	3.621236	3.006516	2.127862

Output result:

Expression patterns areas.

Data format:

seqname	x	y	Fold change value of aerobic and anaerobic	Fold change value of wild type and bleached type	Fold change value of different growth phase	area
comp178174_g1_i1	0	3	-2.67476	-7.10421	-7.11449	14
comp178026_g3_i2	280	90	3.621236	3.006516	2.127862	1

Co-expression patterns displays.R

Co-expression patterns displays of Program

```
#Description :
# Co-expression patterns analyses for BLSOM.
     # Loading ggplot2
     library(ggplot2)
# Analysis of TF families transcripts and other gene transcripts for BLSOM.
     a=read.table("expression area.txt")
     #Co-expression patterns analyses of TF families transcripts and other gene #transcripts for BLSOM
     ggplot(a, aes(V2, V3)) +
       geom_point(data = transform(a, V7 = NULL), colour = "grey85")+
       geom_point(aes(colour = factor(V7)), show.legend =TRUE) +
       theme_bw()+
       facet_wrap(~V7, scales = "fixed", nrow = 4) +
       xlab(NULL) + scale_x_continuous(breaks=seq(0,350,50))+
       ylab(NULL) + scale_y_continuous(breaks=seq(0,192,32))+
       labs(title = "co-expression patterns",colour = "colour code of co-expression patterns areas")+
       theme(plot.title = element_text(size = 20),
             axis.text = element_text(color = "blue", size = 12),
             legend.title = element_text(size = 20),
             legend.text = element_text(size = 15),
             legend.key.size = unit(1, "cm"),
             legend.position = "bottom",
             legend.box = "horizontal",
             strip.text.x = element_text(size = 15))+
       guides(colour = guide_legend(ncol = 20,
                                    direction="horizontal",
                                    title.theme = element_text(size=20, colour = "red", angle = 0),
     title.position = "top"))
```

Input Data:

Expression area

Data format:

seqname	x	y	Fold Change value of aerobic and anaerobic	Fold Change value of wild type and bleached type	Fold Change value of different growth phase	area
comp178174_g1_i1	0	3	-2.67476	-7.10421	-7.11449	14
comp178026_g3_i2	280	90	3.621236	3.006516	2.127862	1

Output result:

Drawing co-expression patterns areas. (Figure 3.16)

Appendix hierarchical clustering (HC)

Partitioning algorithms are based on specifying an initial number of groups, and iteratively reallocating objects among groups to convergence. In contrast, hierarchical algorithms combine or divide existing groups, creating a hierarchical structure that reflects the order in which groups are merged or divided. In an agglomerative method, which builds the hierarchy by merging, the objects initially belong to a list of gene sets $X = \{x_1, x_2, x_3, ..., x_n\}$. Then a cost function is used to find the pair of sets $\{x_i, x_j\}$ from the list that is the "cheapest" to merge. Once merged, x_i and x_j are removed from the list of sets and replaced with $x_i \cup x_j$. This process iterates until all objects are in a single group. Different variants of agglomerative hierarchical clustering algorithms may use different cost functions. Complete linkage, average linkage, and single linkage methods use maximum, average and minimum distances between the members of two clusters, respectively.

Following is the pseudo code of the hierarchical clustering algorithm to explain how it works:

Let $X = \{x_1, x_2, x_3, ..., x_n\}$ be the set of gene data points.

1. Begin with the disjoint clustering having level $L(0) = 0$ and sequence number $m = 0$.
2. Find the least distance pair of clusters in the current clustering, say pair r, s, according to

$$d[(r),(s)] = \min_{\substack{i=1,2,..,n \\ j=1,2,...,n}} d[(i),(j)] \tag{1}$$

   where the minimum is over all pairs of clusters in the current clustering.

3. Increment the sequence number:

$$m = m+1 \tag{2}$$

Merge clusters r and s into a single cluster to form the next clustering m. Set the level of this clustering to

$$L(m) = d[(r),(s)] \qquad r,\ s = 1,\ 2,\ \ldots,\ n \tag{3}$$

4. Update the distance matrix D, by deleting the rows and columns corresponding to clusters r and s and adding a row and column corresponding to the newly formed cluster. The distance between the new cluster, denoted r,s and old cluster k is defined in this way:

$$d[(k),(r,s)] = \min_{\substack{k=1,2,\ldots,n \\ r=1,2,\ldots,n \\ s=1,2,\ldots,n}} (d[(k),(r)], d[(k),(s)]) \tag{4}$$

5. If all the data points are in one cluster then stop, else repeat from step 2.

Appendix non-negative matrix factorization (NMF)

Description of NMF Method:

The data set is consisting of the expression levels of N genes in M samples (which may represent distinct tissues, experiments, or time points). For gene expression studies, the number N of genes is typically in the thousands, and the number M of experiments is typically <100. The data are represented by an expression matrix A of size $N \times M$, whose rows contain the expression levels of the N genes in the M samples.

The goal is to find a small number of metagenes, each defined as a positive linear combination of the N genes. The gene expression pattern of samples can be approximated as positive linear combinations of these metagenes.

Mathematically, this corresponds to factoring matrix A into two matrices with positive entries, $A \approx WH$. Matrix W has size $N \times k$, with each of the k columns defining a metagene; entry w_{ij} is the coefficient of gene i in metagene j. Matrix H has size $k \times M$, with each of the M columns representing the metagene expression pattern of the corresponding sample; entry h_{ij} represents the expression level of metagene i in sample j.

Given a factorization $A \approx WH$, matrix H can be use to group the M samples into k clusters. Each sample is placed into a cluster corresponding to the most highly expressed metagene in the sample; that is, sample j is placed in cluster i if the h_{ij} is the largest entry in column j.

There is a dual view of decomposition $A \approx WH$, which defines metasamples (rather than metagenes) and clusters the genes (rather than the samples) according to the entries of W.

NMF Algorithm:

Given a positive gene matrix A of size $N \times M$ and a desired rank k, the NMF algorithm iteratively computes an approximation $A \approx WH$, where W and H are nonnegative matrices with respective sizes $N \times k$ and $k \times M$. The method starts by randomly initializing matrices W and H, which are iteratively updated to minimize a divergence functional. The functional is related to the Poisson likelihood of generating A from W and H, $D = \Sigma_{i,j} A_{i,j} \log(A_{i,j}/(WH)_{i,j}) - A_{i,j} + (WH)_{i,j}$, $i, j = 1, 2, \ldots, M$.

At each step, W and H are updated by using the coupled divergence equations:

$$H_{au} = H_{au} \frac{\sum_i W_{ia} A_{iu}/(WH)_{iu}}{\sum_k W_{ka}} \quad (1)$$

$$W_{ia} = W_{ia} \frac{\sum_u H_{au} A_{iu}/(WH)_{iu}}{\sum_v H_{av}} \quad (2)$$

Appendix principal component analysis (PCA)

The object of study is an *n* by *d* gene expression matrix, *X*, giving the expression of *N* genes in *M* experiments. The gene expression matrix has a dual nature: one can cluster either genes or experiments; to express this duality, *X* can be refered as

$$\mathrm{X} = \begin{bmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{bmatrix} \tag{1}$$

$g_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iM})$ are the expression levels of gene *i* across all experiments, or as $e_j = (x_{1j}, , \ldots, x_{ij}, \ldots, x_{Nj})'$ are the expression levels in experiment *j* across all genes. The genes are clustered, i.e., the *N* patterns g_N. When clustering experiments, substitute *e* and *g* in what follows. We make a distinction between a variable, which is each one of the *N* variables that make up g_N, and a feature, which denotes a combination of variables.

The PCs are *q* orthogonal directions that can be defined in several equivalent ways. They can be defined as the *q* leading eigenvectors of the covariance matrix of *Y*. The eigenvalue associated with each vector is the variance in that direction. Thus, PCA finds a set of directions that explains the most variance. For Gaussian data, the PCs are the axes of any equiprobability ellipsoid. A low-dimensional representation of a data set is obtained by projecting the data on a small number of PCs.

The PCs can be defined as the *q* leading eigenvectors of the experiment–experiment covariance matrix:

$$Cov(Y)_{ij} = \frac{1}{N}(e_i - < e_i >)'(e_j - < e_j >), \quad i,j = 1, \ldots., M \tag{2}$$

in which $<e_j> = 1/\mathrm{N}\sum_{i=1}^{N} x_{ij}\,[1 \quad \cdots \quad 1]_{1\times M}$ is a M-dimensional vector with the mean expression value for experiment *j*. Alternatively, the PCs can be defined through the correlation matrix:

$$Cov(Y)_{ij} = \frac{1}{N}\frac{(e_i - <e_i>)'}{\sigma(e_i)}\frac{(e_j - <e_j>)}{\sigma(e_j)}, \quad i,j = 1, \ldots., M \tag{3}$$

in which $\sigma(e_i)$ is the vector of estimated standard deviation in experiment i. Equivalently, one can consider the PCs as the eigenvectors of the matrix *XX'* when applying first a normalization stage of centering:

$$e_i \leftarrow e_i - < e_i > \tag{4}$$

or standardization:

$$e_i \leftarrow e_i - < e_i > / \sigma(e_i) \tag{5}$$

The first corresponds to PCA relative to the covariance matrix, and the second to PCA relative to the correlation matrix. To distinguish between the two, we denote them by centered PCA and standardized PCA, respectively. One can also consider PCs relative to the second moment matrix, i.e., without any normalization. In this case, the first PC often represents the mean of the data, and the larger the mean, the larger this component relative to the others.