IJBA

GENOMIC SURVEY, CHARACTERIZATION AND EXPRESSION PROFILE ANALYSIS OF THE YELLOW STRIP LIKE GENE FAMILY IN RICE AND *ARABIDOPSIS*

RAKESH MENNA*, MAHIMA DUBEY AND GIRISH CHANDEL

Department of Molecular Biology and Biotechnology, Indira Gandhi Krishi Vishwavidyalaya, Raipur, MP, India
*Corresponding author. E-mail: ghchandel@gmail.com

Abstract- Genes in the YSL gene family encode for the metal transporter in plants. In this study, a comprehensive and comparative computational analysis of identified 18 and 8 YSL family genes in rice and *Arabidopsis* respectively have been carried out presenting a complete overview on gene structures, phylogeny, conserved motifs their distribution and relatedness, intron/ exon distribution and expression. The analysis in all indicated that YSL genes are relatively conserved gene family showing similarity on the basis of evolution and indicates their common ancestral origin. The multi-sequence alignment showed primary and secondary level of conservation at sequence level of YSLs in both rice *Arabidopsis*. YSL gene intron/ exon structure and domains were also found to be conserved during the evolution. Characterization of YSLs by sequence tag based methods of expression profiling *viz* ESTs and MPPS signatures showed a total of 833 ESTs and 382 MPSS tags co-localizing with the OsYSLs and 269 ESTs and 219 MPSS tags with AtYSLs. The expression pattern further analyzed digitally, by *in silico* microarrays in 22 tissues libraries, 8 stages of development and under different stimuli revealed assorted and preferential patterns of expression and provide an evidence for spatiotemporal regulation of these genes throughout plant development. OsYSL6 gene of rice showed very high as well as tissue specific expression under iron and phosphorous interaction by digital microarrays which were equally supported by ESTs and MPSS abundance. Similarly in *Arabidopsis* AtYSL1, 3 and 6 showed prominent expression in senescent leaves indicating their role in remobilization of metals.

Keywords- Digital expression analysis, Metal transporter, Phytosiderophore, YSL- gene

Introduction

Iron is a key micronutrient for plants which plays a crucial role in a variety of cellular functions and because plants are the primary source of food for humans, the nutritional value of plants is of central importance to human health as well [1]. The most widespread dietary problem in the world is iron deficiency causing 0.8 million deaths annually [2]. Unfortunately the most important crop on the earth 'Rice' feeding large population of the world is a poor source of essential micronutrients including iron [3]. Iron uptake and translocation in plants are important processes for both plant and human nutrition, but relatively little is known about the molecular mechanisms of iron transport within the plant body and its availability for human consumption through its loading to grains. This lack of knowledge about how nutrients are translocated from vegetative tissues to the seeds referred as the homeostasis mechanism is one of the barriers to the most feasible and viable approach of biofortification [4], resulting in uncertainty about the best genes or pathways to target for modification. In this way, a better understanding of metal homeostasis and localization, as well as the identification of the molecular players that might contribute to the process of metal transport to the seeds is essential [5].

Plant iron homeostasis has been often analyzed by physiological and molecular genetics approaches. More recently, plant genomics also appears as an alternative for the understanding of this aspect of plant nutrition allowing the fast identification of molecular components related to mineral homeostasis [6]. This is possible through database searches for sequences homologous to proteins already characterized. Further the genomics technologies have also led to a paradigm shift in biological experimentation particularly the high-throughput gene expression analysis has become a frequent and powerful research tool in biology because they measure (profile) most or even all components of one class (e.g. transcripts, proteins, etc.) in a highly parallel way [7].

Iron availability is directly correlated with plant productivity but despite its abundance in soils, iron is present as oxihydrates with low bioavailability. To avoid a deficiency, two distinct strategies have been proposed for iron acquisition in plants [8]. In Strategy I, used by dicotyledonous and non graminaceous monocotyledonous plants, Fe2+ transport is coupled to a Fe3+ chelate reduction step. The strategy II employed by graminaceous species involves the release of mugenic

55

acid (MA) derived phytosiderophore in the rhizosphere to bind insoluble Fe (III). The so formed Fe (III)–MA chelates are then reabsorbed by the roots via Fe (III) specific transporters avoiding the reduction steps. These transporters have been identified as *Yellow Stripe1* encoded by genes belonging to yellow stripe family [9, 10]. The maize YSL gene, Zm*YS1* was the first identified molecular component related to the transport of the Fe3+-phytosiderophore complex characterized to date which functions as a proton coupled symporter for phytosiderophore-chelated metals [11]. Research by various groups has led to the identification of eight orthologs of maize YS1 referred as *Yellow Stripe-like* or YSLs in the model genome *Arabidopsis*. The rice genome contains 18 putative YSL family genes nomenclatured as *Oryza sativa Yellow Stripe-like* or OsYSLs [6, 12], among which only three members *viz* OsYSL2, OsYSL15, and OsYSL18 have been characterized in detail [13. [14, 15, & 12]. The genes belonging to YSL family, represents a serious candidate for the transport of NA–metal chelates across plant cell membranes. Although final proof is still missing, experimental evidence points to a role of the YSL proteins in the long distance and intracellular transport of metals, especially Fe, complexed to NA which are needed to be explored further. Thus we have undertaken the study focusing on the structural, phylogenetic, evolutionary analysis. We have also undertaken comprehensive functional characterization of all the YSL family members in rice (OsYSL) and *Arabidopsis* (AtYSL) by their expression profiling using signature tag based EST and MPSS analysis followed by *in slico* microarrays. The purpose of this study was also to contribute to the deeper and comparative understanding of YSL gene family in rice and *Arabidopsis* by analyzing the expression pattern in different tissue types, developmental stages and under different stimuli, motif analysis of the YSL proteins and phylogenetic relationships between them.

MATERIALS AND METHODS
In silico structural analysis of YSL family genes in rice and *Arabidopsis*
To collect all the sequences belonging to YSL gene family in rice and *Arabidopsis,* protein sequences of YSL family genes (26 in total) from both the model species plants were downloaded from SWISSPROT and TrEMBL *http://www.expasy.org/sprot/* databases. The proteome of rice, available in MSU Rice Genome Annotation Project Database release 6.1, KOME database as well as NCBI databases, were searched for the YSL family genes. The presence of YSL family in resulting sequences was confirmed using NCBI Conserved Domain (NCBI-CD) search tool with filter off. YSL family sequences of *Arabidopsis*, eight in all, as defined by Curie *et al.,* 2001, were downloaded and confirmed using The *Arabidopsis* Information Resource TAIR; *http://www.Arabidopsis.org*/ database. The structural details of nucleotide and protein sequences of rice and *Arabidopsis,* YSL family genes were downloaded from

MSU and TAIR databases, respectively. Further the sub-cellular localization of YSL proteins was predicted using three different protein localization prediction software packages including WOLF PSORT *http://wolfpsort.org/*. The 5' UTR, 3'UTR and intron structures were investigated using SeqViewer at TAIR.

Phylogenetic analysis and sequence alignment of YSL family genes in rice and *Arabidopsis*
ClustalW2 on line software Information (Genomique et Structurale, LIRMM, MUSCLE*),* a multiple sequence alignment tool with high accuracy and high throughput available on EMBL-EBI software [16] was used to perform the multiple sequence alignment and generate the unrooted phylogenetic tree based on the protein sequences of OsYSL and AtYSL members with neighbor-joining method [17] and a combined tree with OsYSL and AtYSL proteins was generated. The Neighbor-Joining method was used with the following parameters: pairwise deletion of gaps/missing data; poisson correlation of model; bootstrap 1000 replicates, random seed of phylogeny test. Only clades with the bootstrap value higher than 50 were selected for the bootstrap consensus tree [18] and [19].

Intron/ exon structure and motif analysis
Genomic sequence and CDS (coding DNA sequence) of rice YSL genes were used to derive intron/ exon structure with the online tool Gene Structure Display Server *http://gsds.cbi.pku.edu.cn/chinese.php.,* [20] and for *Arabidopsis* YSL genes, TAIR *http: //Arabidopsis.org* was used for the same. Conserved motif structures within YSL domain for rice and *Arabidopsis* genes were analyzed by MEME 4.3.0 tool (Multiple Expectation Maximization for Motif Elicitation) with the following parameters; distribution of motif occurrences: any number of repetitions; number of different motifs: 20; minimum motif width:6; and maximum motif width: 50 [21] and [22].

In silico signature tag based expression profiling of YSL family genes
In silico expression profiling of YSL genes was carried out to predict the putative temporal and spatial pattern of expression by analyzing the co-localization of identified ESTs and MPSS tags with the YSL gene sequences. The expression level and pattern were predicted based on the frequency of ESTs and MPSS signatures abundance. The locus identifier of each YSL gene was used as query to search for the ESTs co-localizing with genes and further expression pattern was predicted on the basis of respective tissue expression library information generated by Rice Gene Expression Anatomy Viewer and Digital Northern tools available at TIGR database *http://www.tigr.org/tdb/e2k/osa1/ dnav/*. ESTs corresponding to a tissue library provided information about putative site of expression of the YSL genes. As the MPSS provides more thorough qualitative and quantitative description of gene expression, the further characterization of YSL genes was done by

MPSS signature analysis. The rice MPSS database includes a comprehensive set of libraries which can be accessed at site, *http://mpss.udel.edu/rice*. The tool provides 17 and 20 nucleotide long tags, tag positions, chromosome coordinates *etc*. The sequence of each YSL gene was used as query under 'query by sequence' section of rice MPSS database to identify MPSS tags corresponding to the YSL query sequence as well as their abundance in 22 diverse tissue libraries constructed from various developmental stages, tissue types and tissues treated by various biotic and abiotic stresses. The abundance/ frequency of each tag is expressed in TPM (transcript per million) and the TPM value under 'Norm Abund' category is considered as the measure of expression in a corresponding tissue library. Similarly, *Arabidopsis* MPSS tool was used to extract the MPSS signatures corresponding to each YSL genes in 18 different tissue libraries. *Arabidopsis* MPSS database was used for the MPSS signature of AtYSL family with same parameters.

Expression pattern analysis by *in silico* microarrays

To further investigate and confirm YSLs gene expression, *in silico* microarray analysis was performed in various anatomical tissues, under various developmental stages of plant as well as under different stimuli on Genevestigator Version 3 *https://www.genevestigator.com/gv/index.jsp* server. Rice affymatrix (Os_51K Rice Genome 51K) high quality array platform was employed for analyzing eighteen OsYSL genes [23] while eight AtYSL genes were analyzed by choosing public high quality AtGenExpress ATH1-22k microarray data platform. Further the rice YSL genes were also analyzed separately at Rice array database tool site for differential expression and confirming the tissue specificity under specific treatment of iron and phosphorus [24]. The analysis was based on the *Rice Affymetrix GeneChip* experiment platform. These tools generates Log_2 transformed signal values generated from the average of three biological replicates for each plant organ and performs a heat map of normalized signal intensity values, corresponding to the different organs of the plant, for each gene. This provides a quantitative measure of the transcript of a particular gene and hence its expression. Meta-profile analysis and hierarchical clustering were used to study gene expression at different development stages, in anatomical tissues and under different stimulus [25].

RESULTS AND DISCUSSION

In silico structural analysis of YSL family genes in rice and _Arabidopsis_

Eighteen genes belonging to the YSL family have been reported in rice by Gross *et al.*, 2003 which were supported in further studies by the finding of other researchers [26, 12 and 27]. Similarly eight YSL gene sequences encoding putative transmembrane proteins with founding member of the family have been reported in *Arabidopsis* [11]. When we performed an independent search for YSL family genes in rice genome, all the 18 of

the previously reported YSL proteins could be confirmed using present annotation database (version 6.1) of MSU which were analyzed further. Similarly sequences of all 8 *Arabidopsis* YSL genes were retrieved from TAIR database and verified for the presence of YSL domain. The structural features of rice and *Arabidopsis* YSL family genes including chromosomal position, open reading frame (ORF) length, protein length, molecular weight, iso electric point (pI) and number of introns, position of UTRs etc have been extracted from the respective databases and are presented in table 1 and 2 respectively. It was observed that the protein sequences showed limited variation (554 to 728 amino acid residues) among different members of YSL family in rice while even lesser variation in protein sequence length was observed for AtYSL genes which varied from 664 to 724 amino acids. Since, localization provides an important clue to the function of a protein; the sub-cellular localization of YSL family proteins was also checked using three different online prediction programs. 18 proteins of rice and 8 from *Arabidopsis* were found to exhibit high probability of being localized in plasma membrane (Table 1 and 2).

Phylogenetic analysis and sequence alignment of YSL family genes in rice and _Arabidopsis_

In order to investigate the phylogenetic relationship and functional divergence of YSL members as well as to identify some orthologous genes, a combined phylogenetic tree with OsYSL and AtYSL proteins was established (**Fig. 1**). As a result four subfamilies were formed. Subfamily II and III contained both rice and *Arabidopsis* YSL members while I and IV showed complete absence of AtYSL members. However, most of the members were clustered in species-specific distinct clades, and only genes OsYSL1 and AtYSL1 could be figured out. This result indicate that the main characteristics of YSL family in rice and *Arabidopsis* were formed before the split of monocotyledonous and dicotyledonous plants and then evolved separately in a species-specific manner, due to the difference in the total number of OsYSL and AtYSL which were 18 in rice and eight in *Arabidopsis*.

The multi sequence alignment is the approach to identify the conserved domain for gene function and is of great value for studying functional domains and conserved amino acid residues in genes of organisms especially in the model plant with finished genome sequence. After sequence retrieval from the SwissPort and TAIR database, OsYSL and AtYSL proteins were analyzed by ClustalW2 online tool. Across the whole sequence of OsYSL proteins, several conserved regions with high similarity were observed (represented as different colors as shown in **Table 3 and additional file 1),** primary level conservations colored in black and secondary level conserved regions shown in gray color were observed among YSL proteins. No tertiary level conservation was found. Highly conserved proteins are often required for basic cellular function, stability or reproduction. Conservation of protein sequences is indicated by the

presence of identical amino acid residues at analogous parts of proteins. Conservation of protein structures is indicated by the presence of functionally equivalent, though not necessarily identical, amino acid residues and structures between analogous parts of proteins. Shown below is an amino acid sequence alignment between all 18 OsYSL genes. Conserved amino acid sequences are marked by strings of * on the last line of the sequence alignment. As can be seen from this alignment, these eighteen proteins contain a number of conserved amino acid sequences represented by identical letters aligned between the two sequences. It was also noticeable that OsYSL domains spread across a wide range of protein structure and covers many amino acid residues throughout the whole protein, those were polar non charged, non alphatic residues, most hydrophilic and positively charged residues in all OsYSLs. Although the OsYSL domain was conserved in evolutionary process as evident from the phylogenetic analysis. The phenomena is in contrary to many other genes families such as the TIFY domain of JAZ family, CCCH zinc finger domain of CCCH zinc finger family and the ERF/AP2 domain in ERF family [27, 28] and [29] where a clearly defined region of some specific amino acid residues have been demonstrated.

From the comparison of the protein sequences and sequences alignment, we found three special motifs that were highly conserved in most of the OsYSLs member. Motif 1 (LAACGVMMQIVHTASDLMQDFKTGHLTLTSPRSMFVS QVIGTAMGCVINP) was found between 459 to 561 amino acid from the N terminal of the YSL domain. Motif 2 whose sequence was (VPLRKVMIIDYKLTYPSGTATAHLINSFHTPHGAKQAK KQV) was located between 163 to 253 amino acid region of the OsYSL domain, and motif 3 (GDNCGFHQFPTFGLEAYKHRFYFDFSPTYVGVGMIC PHIVNCS) was identified and located between the 223 to 318 amino acids. The HMM logos of these motifs in OsYSL are shown in (Fig. 2A). The length of motif 1 found in all the 18 OsYSL proteins was 50 amino acid, which was observed as the best conserved sequences, length of motif 2 was 41 AA while motif 3 was 43 AA in length. The similarity score between the motif 1 and 2 was 0.19, between 1and 3 was also 0.19 and between motif 2 and 3 was 0.15. In most AtYSL members, similar conserved motifs could also be figured out (Fig. 2B). However, motif 1 (PGLGWMTGFLFVVSFLGLFSLVPLRKIMIIDYKLTYPS GTATAHLINSFH) was found to be located on and before 200 amino acids, motif 2 (STASDLMQDFKTGHLTLSSPRSMFVSQAIGTAMGCVV APCTFWLFYKAFD) located on and after the 500 amino acids and motif 3 (QFPTFGLKAYQNTFYFDFSMTYVGCGMICPHIVNCSL LLGAILSWGIMWP) was found to be located on and before the 300 amino acids. All the three motifs for the 8 AtYSL had 50 AA in length. The similarity score between motif 1and 2 and between 2 and 3 was 0.15, while this value was 0.16 between motif 2 and 3. The Sequence

screening of motifs in Inter-Proscan and Uniprot databases retrieved only OPT transporter family proteins which is a super family to which YSL members belong. Therefore, it can be said that these motifs were the newly confirmed sequences associated to the plant YSL family.

The multiple sequence alignment correlated with and confirmed the motif finding analysis, where quite diverse motif structures have been found for genes from different subfamilies. Combining the result from sequence alignment and motif finding, it can be stated that YSL genes were conserved in the existence of certain specific motif but had rather diverse motif distributions across different classes. The structural diversity could account for the different biological functions of these genes.

Intron/ exon structure of YSL genes

In rice the intron/ exon structure of YSL genes were found to be different for each gene and could be divided into several groups as shown in (Fig. 3). The numbers of introns were quite different across the whole family ranging from one intron to seven introns in one gene. Only single gene, YSL17 was found to be intronless. For example, the OsYSL 8 had only one intron and OsYSL6 had seven introns. The OsYSL, 3, 4, 5, 10, 11, 12, 13, 14 and 18 all showed the presence of five introns, OsYSL 2, 9, 15 and 16 showed six introns and OsYSL 7 and 17 had three introns; while OsYSL 1 had only four introns. Similarly in *Arabidopsis* the intron/ exon structure of YSL genes were different for each gene and can also be divided into several groups (Fig. 6). The numbers of introns ranged from three introns to eight introns in one gene. The AtYSL 7 and AtYSL1 had only three introns and *AtYSL*3 showed the presence of eight introns. The AtYSL2, 5, and 8 all had five introns and AtYSL 4 and 6 were found to have six introns. A clear correlation between the groups based on intron/ exon structures and the classes of YSL genes based on their phylogenetic relationship (Fig. 1) was observed. The YSLs falling in a particular subfamilymade on the basis of their evolutionary relationship showed more or less same number of introns. This could be probably due to the recent expansion of YSL genes in each class of YSL genes. On the other side, the YSL gene intron/ exon structure has certain level of stability during the evolution and intron gain/ loss would have played a role in the early stage evolution of YSL genes. Correlation between the intron/ exon structure and the motif structure was also observed, which is reflected from the distinct patterns found in each subfamily of YSL family genes. It is therefore expected that the gene birth due to the intron insertion or intron loss happened earlier during the evolution. Both the gene length and intron phase correlate with the gene family classification and intron numbers to a certain degree. Intron phase 0, 1, and 2 referred to the splicing occurred after the first, second, and third nucleotide of the codon, respectively. As shown in Figures 6 and 7, genes with similar intron/ exon structures and gene length also had conserved splicing phase patterns [31] and [21]. The results indicated that

the YSL genes in rice and *Arabidopsis* may have experienced fewer intron birth events as compared to the rice. Overall, intron/ exon structure, and the conserved domain all correlate well with the phylogenetic analysis and relatedness of the genes [29, 30 and [32].

In silico signature tag based expression profiling of YSL family genes

In silico expression profiling of YSL genes was carried out to predict the putative temporal and spatial pattern of expression. The expression pattern was predicted based on the frequency of ESTs and MPSS signatures co-localizing with the YSL gene sequences. A total of 833 ESTs were identified which expressed in 19 different tissues libraries, EST number among 18 YSL genes in rice ranged from 0-384 ESTs per gene and 0-254 ESTs in each tissue library. OsYSL7 gene encoding metal-nicotianamine transporter showed the highest number of ESTs (384) followed by OsYSL*6* which showed the presence of 150 co- localized ESTs. Some of the genes were found to show preferential expression in certain tissue libraries. OsYSL7 expressed preferentially in panicle tissue type with EST count of 214 followed by mixed and flower tissue types. Similarly OsYSL6 expressed specifically in the shoot tissues (Fig.5). Overall, considering all the eighteen OsYSLs, high level expression was observed in panicle while significant expressions were also observed in other tissue libraries including mixed, flower, callus, shoot and root tissues. [33] described the criteria for transcriptome analysis of stress modulated genes by digital northern technique in model genome *Arabidopsis.* As per their criteria, the genes were classified to three different categories based on the number of EST matches to a gene. Minimally expressed category was assigned if the EST matches to a gene were <7, relatively highly expressed for 7-200 EST matches and highly expressed for > 200 EST matches to a gene sequence. A similar criterion was followed by [34] for characterization of NBS-LRR genes in the model tree species Medicago by *in silico* expression analysis using EST library. Based on these two combined criteria, the YSLs were sorted to minimally expressed, relatively highly expressed and highly expressed category. Based on this, 11out of 18 YSLs present in rice showed EST matches of 7 and above and thus were in relatively highly expressed category while, a single gene OsYSL7 was in the highly expressed category.

MPSS signature analysis of YSL genes in rice and *Arabidopsis*

We also performed *in silico* expression analysis of 18 OsYSLs and 8 AtYSLs by measuring the abundance of MPSS signatures co-localizing with the gene sequences. The abundance of MPSS tags identified in each sequence is depicted by its TPM value (transcript per million), which is an exact digital representation of number of copies of the transcript in a tissue which indicates expression level of the corresponding gene quantitatively. Great variation in TPM values of MPSS

tags were observed for YSL gene ranging from 0 to 6,427. But only those MPSS tags having TPM > 15 in at least one tissue library were considered. The TPM value below 15 is indicative of very low and basal levels of expression [35, 36]. A total of 719 MPSS tags (17 bp) were found corresponding to 18 YSLs genes and the tag number ranged from 75-217. High TPM tags corresponding to YSLs genes were found with MPSS tag sequence GATCTGGAGTGTTCCAT corresponding to OsYSL6 showing the highest cumulative TPM value of 6427 in all the libraries. Strong expression levels based on TPM value were observed for OsYSL6, OsYSL7, OsYSL12, OsYSL13 and OsYSL16 whereas, OsYSL5 and osYSL15 showed higher levels of expression (TPM> 500, [36]. From the tissue library wise expression of OsYSLs*,* it was observed that these genes expressed to higher levels in libraries consisting of pathogen (*M. grisea* and *X. oryzae*) challenged tissues. This was a bit unexpected finding in regard to metal transporter genes, but the literature reporting similar observation for zinc uptake regulator (*zur*) gene supports our findings. Zur is a functional member of the Zur regulator family that controls zinc and iron homeostasis and have been found to be involved in maintenance of virulence of the pathogen *Xanthomonas oryzae* [37]. These findings suggest the role of YSLs in plant defense and resistance mechanisms as well. Among other tissue types, significantly higher expressions were observed in mature pollen, young and mature roots, young and mature leaves etc.

Similarly a total of 220 MPSS tags were found corresponding to 8 AtYSLs genes present on the five chromosomes of *Arabidopsis* and the tag number ranged from 1 to 85. Out of 8 AtYSL genes, AtYSL1 showed highest MPSS TPM value of 828 followed by AtYSL6 and AtYSL7. In terms of signature tag sequence, the tags GATCTGTTCCATTTCCG and GATCTAGATGTGTTGAC corresponding to AtYSL1 showed highest TPM value of 372. These tags expressed in almost all the tissue libraries. The AtYSLs with high TPM value tags showed comparatively higher expression in reproductive parts like inflorescence, silique and other parts like leaves, callus, and root. Preferential tissue expression could not be worked out clearly for the YSLs in *Arabidopsis* as based on EST analysis (Fig. 6 and 7).

Differential expression pattern of YSL genes in rice and *Arabidopsis*

In order to understand the YSLs gene functions and their relevance to gene evolution, we investigated the gene expression level of YSL genes in rice and *Arabidopsis* digitally in different tissues, development stages and under different stimuli conditions using Rice array database and Genevestigator version 3.0 tools for rice and *Arabidopsis* respectively. Developmental stage and tissue-specific expression data were analyzed by hierarchical clustering as shown in (Fig. 8A and 8B), whilst gene expression patterns under stimulus conditions were analyzed and shown using meta-profile

analysis in (Fig. 8C). Twenty two development stages were surveyed for the digital gene expression analysis in both the model species. It was observed that, YSL genes showed significant variations for gene expression in terms of both the expression levels and presence at different conditions (Fig. 8A). From this analysis it was observed that the genes OsYSL1, 3, 4, 7, 8, 10, 11, 17 and 18 expressed in similar pattern with very high expression levels in stamen and anther tissues. While the same genes have shown negligible expression in female reproductive organs like pistil, stigma, ovary, seed, embryo and vegetative tissues like root, shoot, leaf etc indicating their male reproductive organ specific expression. A very good correlation could be worked out in regard of the response of many these genes by *in silico* microarrays and our MPSS signature analysis. Role of YSL in the development of reproductive organs has been well demonstrated with AtYSL1 and 3 involved in the pollen development activity [38]. Moreover current reports on the characterization of *Arabidopsis* YSL knockouts point to an involvement of the YSL transporters in the development of pollen grains and seeds. This is consistent with the strong expression of the YSL genes observed in these organs and with the fact that pollen and seeds represent the main sink for Fe in *Arabidopsis* [10]. OsYSL6 showed very good expression in seedling, leaf, blade (lamina), rhizome and root tissues. This finding confirms our findings of ESTs and MPSS signature analysis where, higher expression of OsYSL6 gene in these tissue types (shoot, root and leaf) were observed. This finding from the three types of analysis indicates that this particular gene is a potential long route transporter showing its active involvement in the plant activities right from the uptake/ absorption of iron by roots to its loading to grains via intermediate shoot tissues. [39], reported uniformly high expression of YSL6 gene among 12 diverse rice genotypes in the root and shoot tissues as analyzed by semi quantitative RT-PCR.

OsYSL15 showed significant expression in root and rhizomes type of tissues indicating its role in the uptake/ absorption of iron from the source. This finding is consistent with the findings of Inoue *et al.,* 2009) where the essentiality of Iron-regulated Iron (III)-deoxymugineic acid transporter of rice, OsYSL15 and its role in transport of Fe (III)–DMA from the rhizosphere to roots, has been demonstrated. The gene OsYSL2 showed preferential expression in the leaf tissues suggesting that OsYSL2 functions as transporter responsible for the phloem transport of iron. Similar findings have also been reported where OsYSL2 expression has been found to be strongly induced in iron-deficient leaves with particularly strong expression in phloem cells of the leaves and leaf sheaths [14].

In terms of expression in various developmental stages, differential expression of YSLs has been observed. We found that the *OsYSL6* gene showed maximum expression in tillerning stage of rice growth and high gene expression in seedling followed by stem elongation stage, germination and dough stages while, medium

expression in booting stage was observed. The OsYSL5 and OsYSL13 gene showed medium gene expression in germination, seedling, tillering, dough stage and high expression in stem elongation stages of development and differential gene expression patterns, too (Fig. 8B). The YSL genes were also found responding to the biotic and abiotic stimulus treatment quite differently (Fig. 8C). For example, OsYSL2 is highly up-regulated under stress as cold, arsenate, whilst it is down-regulated under some salt stress conditions. However, other YSL genes did not show similar expression pattern under these treatments.

Similarly in *Arabidopsis,* it was observed that most of the genes were only moderately expressed in different tissues while some showed tissue specific expression. AtYSL1, 3 and 6 showed very good expression in some common tissue types including cauline leaf and senescent leaf tissues. Marked increase of expression of AtYSL1 and AtYSL3 in senescent leaves have been experimentally demonstrated by strong induction of their specific promoters which supports the concept of role of the YSLs in metal remobilization from senescent leaves laid down by Curie *et al.,* 2009. Besides this, AtYSL1 also showed higher expression levels in petal and sepals and AtYSL6 in seed tissues. From the perspective of developmental stage specific expression, it was observed that AtYSL6 showed very strong expression in germinating seeds, flower and mature silique. Higher level expression for AtYSL1 was also observed in flower and siliques and developed flowers. This finding is consistent with the findings of Curie et al., (2009) reporting higher expression of this gene in these tissues as evidenced by the strong induction of AtYSL1 promoter in flower and silique. On the other hand gene AtYSL3 expressed highly in majority of the developmental stages. Medium to low level expressions were recorded for rest of the genes. Various AtYSL genes also responded differently to treatment of different stimuli. AtYSL7 responded maximally to various stimuli among all AtYSL genes with higher expression under various biotic stresses including *P. syringae*, wounding, abiotic stress including drought, salt, cold, hormone treatment (Zeatin, GA$_3$ and ACC), treatment with elicitor molecules like salicylic acid, methyl jasmonate, and supplementation with nutrients and nitrate starvation. This indicates its involvement in wide array of physiological and metabolic activities in plants as revealed from the signal intensity displayed in microarray data format (Fig. 9A, B and C).

The *in silico* microarray gene expression pattern observed for rice YSL genes were further analyzed by performing the same analysis on the rice array database to analyze their tissue specificity under a specifically defined condition of iron phosphorus interaction studies in rice seedlings (Fig. 10). This analysis revealed that two genes namely OsYSL 6 and 12 have high expression in roots of 7 days seedling, mature leaf and young leaf while they showed very low level expression in young inflorescence P1 (upto 3 cm) to P5 (15 - 22 cm) (P1); 3-10 cm, meiotic stage (P2 and P3); 10-15 cm,

young microspore stage (P4); 15-22 cm, vacuolated pollen stage (P5); 22-30 cm,) and medium expression in seed also shown by the scale. On the other hand these genes showed very high expression in root_+Fe+P and shoot_+Fe+P both while a little lesser level in root_-Fe+P and shoot_-Fe+P type treatments (Fig.12) Overall, the gene expression pattern indicated that YSL genes are involved in diverse biological functions and most of the YSL genes evolve new functions after the gene duplication. These all expression of different YLS genes in different tissues has confirmed the findings by the EST and MPSS signature analysis, for example the YSL6 has high number of EST and MPSS signature in all tissues library. These observations imply a distinct role for each sub-class of YSL. Either they function in different pathways of metal transport, or they transport in different directions such as loading versus unloading of the vessels. Alternately YSL transporters could have different substrate specificities in terms of the metal or the ligand involved in the transported complex. The OsYSLs genes are known as components of Strategy II of metal transport found in cereals, encoding oligopeptide phytosiderophore transporter proteins [6, 40 & 10].

CONCLUSION

Comprehensive characterization of YSL family genes in model representative species of monocots and dicots reveals their diversity as well as dynamicity in plant mechanisms. Various studies have been carried out to characterize YSL members in rice and *Arabidopsis* but only few of them have been characterized in detail. Our *in silico* characterization covering the complete list of YSL family genes based on structure as well as function in the two model systems provides deeper and comparative understanding of YSL genes. YSL family members are relatively underexplored genes particularly from the expression perspective, our expression analysis studies using high throughput transcriptional profiling *in silico* tools over a range of tissue types, plant developmental stages and under various stimuli conditions will help to understand their role better. The huge compilation of data presented in the study not only deeply describes the role of YSL family genes in metal transportation but also provides insight into their role in other plant mechanisms as well. The will serve as the single platform for selection of the YSL family genes for further characterization and manipulation of the mechanism of metal transport in plants.

ACKNOWLEDGEMENTS

References:

[1] Grusak M.A., DellaPenna D. **(**1999) *Ann Rev Plant Physiol Plant Mol Biol*, 50, 133 –161.

[2] WHO. (2002) World Health Report (2002) *World Health Organization, Geneva, Switzerland*.

[3] Bouis H. E., Welch R. M. (2010) *Crop Sci.* 50:S20–S32.

[4] Colangelo E. P., Guerinot M.L. (2004) *Plant Cell*, 16, 3400–3412.

[5] Raul A. Sperotto., Tatiana Boff., Guilherme L., Duarte Lívia S.Santos., Michael A. Grusak., Janette P. Fett. (2010) *Journal of Plant Physiology* 167, 1500–1506.

[6] Gross J., Stein J.S., Fett-Neto A.G., Fett J.P. (2003) *Genet Mol Biol,* 26, 477-497.

[7] Sacha Baginsky, Lars Hennig, Philip Zimmermann, Wilhelm Gruissem (2010) *Plant Physiology* 152, 402-410.

[8] Marschner H., Romheld V., Kissel M. (1986) *J Plant Nutr*, 9, 3-7.

[9] Takagi S. (1976) *Soil Sci Plant Nutr* 22, 423–433.

[10] Curie C., Cassin G., Couch D., Divol F., Higuchi K., Le Jean M., Misson J., Schikora A., Czernic P., Mari S. (2009) *Ann Bot (Lond)* 103, 1–11.

[11] Curie C., Panaviene Z., Loulergue C., Dellaporta SL., Briat J.F., Walker E.L. (2001) *Nature* 409, 346-349.

[12] Koike S., Inoue H., Mizuno D., Takahashi M., Nakanishi H., Mori S., Nishizawa N.K. (2004) *Plant J* 39, 415-424.

[13] Takahiro Aoyama, Takanori Kobayashi, Michiko Takahashi, Seiji Nagasaka,Kanako Usuda, Yusuke Kakei, Yasuhiro Ishimaru, Hiromi Nakanishi, Satoshi Mori, Naoko K. Nishizawa (2009) *Plant Mol Biol* 70, 681–692.

[14] Inoue H., Kobayashi T., Nozoye T., Takahashi M., Kakei Y., Suzuki K., Nakazono M., Nakanishi H., Mori S., Nishizawa N.K. (2009) *J Biol Chem.* 284, 3470-3479.

[15] Ishimaru Y., Suzuki M., Tsukamoto T., Suzuki K., Nakazono M., Kobayashi T., Wada Y., Watanabe S, Matsuhashi S., Takahashi M., Nakanishi H., Mori S., Nishizawa N.K. (2006) *Plant J* 45, 335-346.

[16] Dereeper A., Guignon V., Blanc G., Audic S., Buffet S., Chevenet F., Dufayard J.F., Guindon S., Lefort V., Lescot M., Claverie J.M., Gascuel O. (2008) *Nucleic Acids Res.* 1:36(Web Server issue):W465-9.

[17] Saitou N., Nei M. (1987) *The. Mol Biol Evol* 4: 406-425.

[18] Li X., Duan X., Jiang H., Sun Y., Tang Y., Yuan Z., Guo J., Liang W., Chen L., Yin J. (2006) *Plant physiology*, 141(4):1167-1184.

[19] Soltis PSSa D.E. (2003) *Statist Sci* 18 (2):256-267.

[20] Guo AYZQ., Chen X., Luo J. C. GSDS. (2007) *Yi Chuan*, 29 (8):1023-1026.

[21] Bailey T.L., Elkan C. (1994) *ISMB* 2:28-36.

[22] Yanhui C., Xiaoyuan Y., Kun H., Meihua L., Jigang L., Zhaofeng G., Zhiqiang L., Yunfei Z., Xiaoxiao W., Xiaoming Q. (2006) *Plant molecular biology*, 60(1):107-124.

[23] Zimmermann P., Hirsch-Hoffmann M., Hennig L., Gruissem W. (2004) *Plant physiology* 136(1):2621-2632.

[24] Jung K.H., Dardick C., Bartley LE., Cao P., Phetsom J., Canlas P., Seo Y.S., Shultz M., Ouyang S., Yuan Q., Frank B. C., Ly E., Zheng L., Jia Y., Hsia A.P., An K., Chou H. H., Rocke D., Lee G. C., Schnable P. S., An G., Buell C.R., Ronald P. C. (2008) *PLoS ONE*, 3(10):e3337. PubMed

[25] Owens D. K., Alerding A. B., Crosby K. C., Bandara A. B., Westwood J. H., Winkel B. S. (2008) *Plant physiology* 147(3):1046-1061.

[26] Schaaf G., Ludewig U., Erenoglu B.E., Mori S., Kitahara T., von Wirén N. (2004) *J Biol Chem* **279**: 9091-9096.

[27] Raymond J., DiDonato Jr., Louis A., Roberts., Tamara Sanderson, Robynn Bosler Eisley, Elsbeth L., Walker. (2004) *The Plant Journal* 39, 403–414.

[28] Thines B., Katsir L., Melotto M., Niu Y., Mandaokar A., Liu G., Nomura K., He S.Y., Howe G. A., Browse J. (2007) *Nature* 448 (7154), 661-665.

[29] Nakano T., Suzuki K., Fujimura T., Shinshi H. (2006) *Plant physiology* 140 (2), 411-432.

[30] Wang D., Guo Y., Wu C., Yang G., Li Y., Zheng C. (2008) *BMC genomics* 9:44.

[31] Long M., Rosenberg C., Gilbert W. (1995) *Proceedings of the National Academy of Sciences of the United States of America* 92 (26), 12495-12499.

[32] Yuan J. S., Wang D., Stewart C. N. (2008) *Biotechnology journal* 3(1), 112-123.

[33] Mahalingam R., Gomez-Buitrago A., Eckardt N., Shah N., Guevara A., Day P., Raina R., Fedoroff N.V. (2003) *Genome Biology* 4: R20.1-R20.14

[34] Ameline-Torregrosa C.A., Wang B.B., Majesta S. O'Bleness., Deshpande S., Zhu H., Roe B., Young N.D., Cannon S.B. (2008) *Plant Physiol* 146, 5-21

[35] Meyers B.C., Galbraith D.W., Nelson T., Agarwal V. (2004a) *Plant Physiol* 135, 637-652.

[36] Meyers B.C., Tej SS., Vu T.H., Haudenschild C.D., Agrawal V., Edberg S.B., Ghazal H., Decola D. (2004b) *Genome Res* 14, 1641-1653.

[37] Yang W., Liu Y., Chen L., Gao T., Hu B., Zhang D., Liu F. (2007) *Curr Microbiol*. 54 (4), 307-14.

[38] Waters B., Chu H., DiDonato R., Roberts L., Eisley R., Lahner B., Salt D. and Walker E. (2006) *Plant Physiol*. 141, 1446-1458.

[39] Banerjee S., Sharma DJ., Verulakr SB., Chandel G. (2010) *Indian Journal of Biotechnology* 9, 203-212.

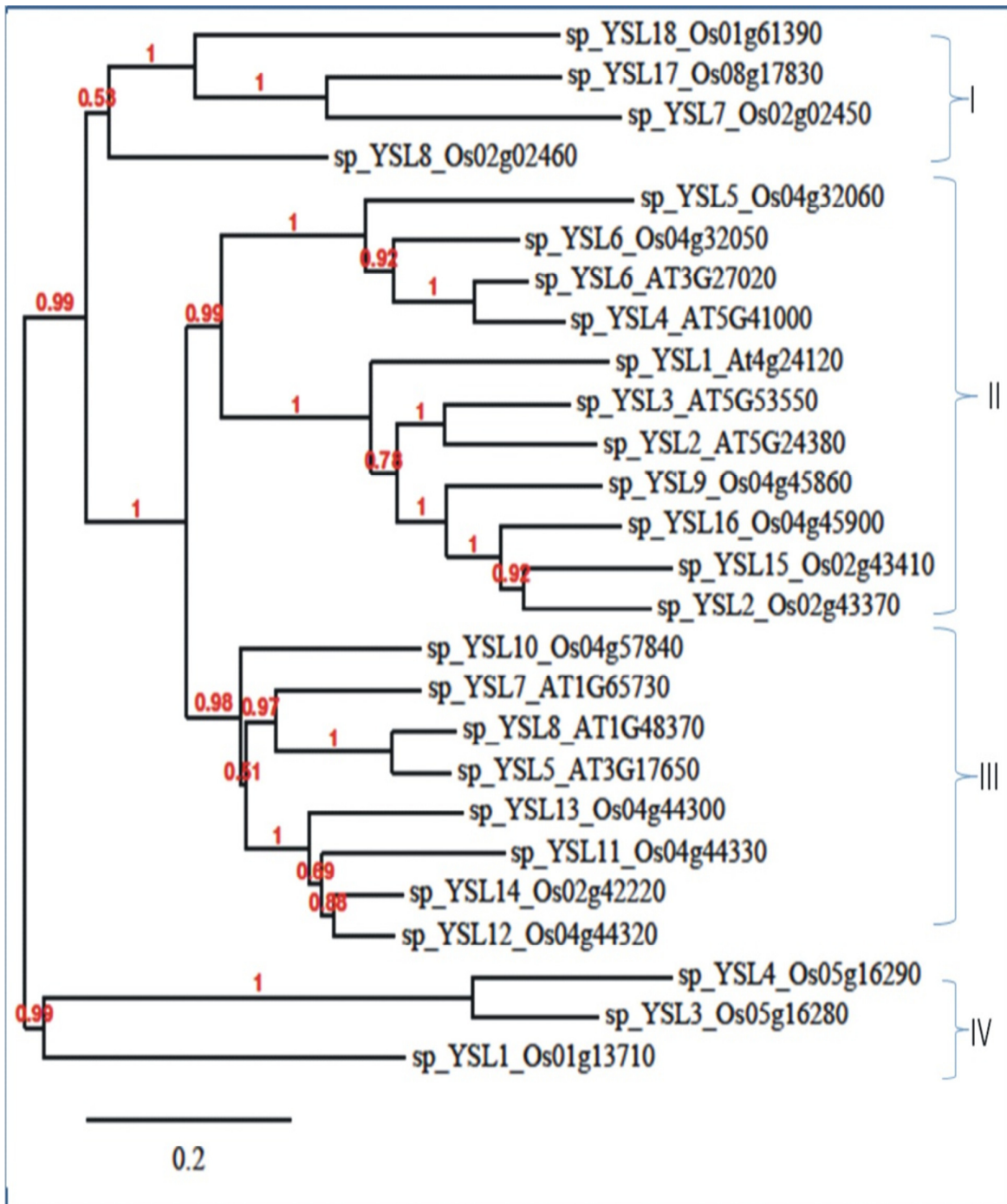[40] Yen M.R., Tseng Y.H., Saier Jr M.H. (2001) *Microbiology* 147, 2881-2883.

Fig. 1- Combined phylogenetic relationship between OsYSL and AtYSL proteins. Bootstrap values from 1000 replicate are indicated at each node. Scale bar represents 0.2 amino acid substitution per site. The protein on the tree can be divided into four distinct subfamilies (I-IV).
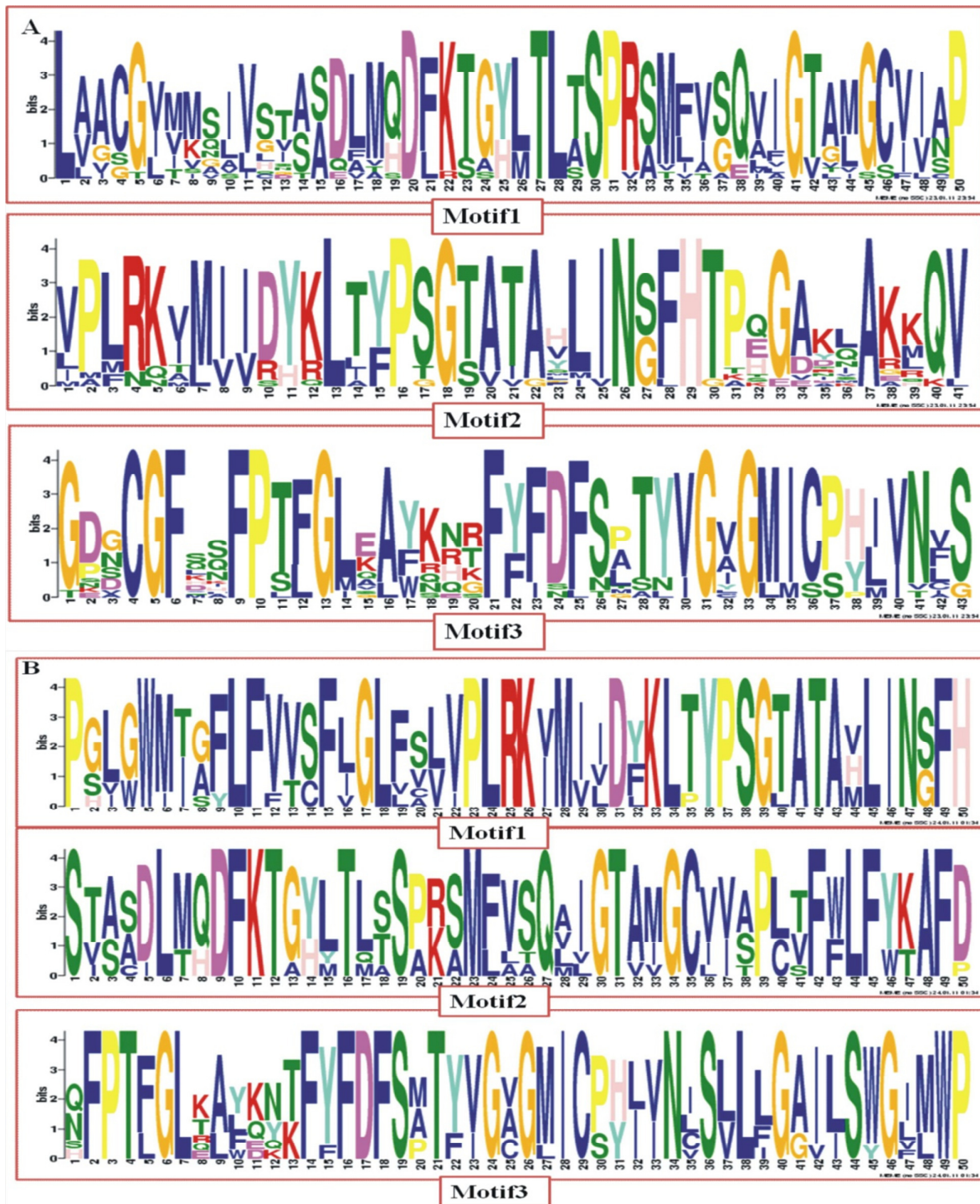
Fig. 2- HMM logos of conserved motifs in OsYSL and AtYSL proteins. X- axis indicate the content sequence of the motif, Y-axis is scale of relative entropy which reflects the rate of each amino acid. The names of the motifs are marked under graphics A: Logos of three motifs OsYSL, B: logos of three motifs in AtYSL.
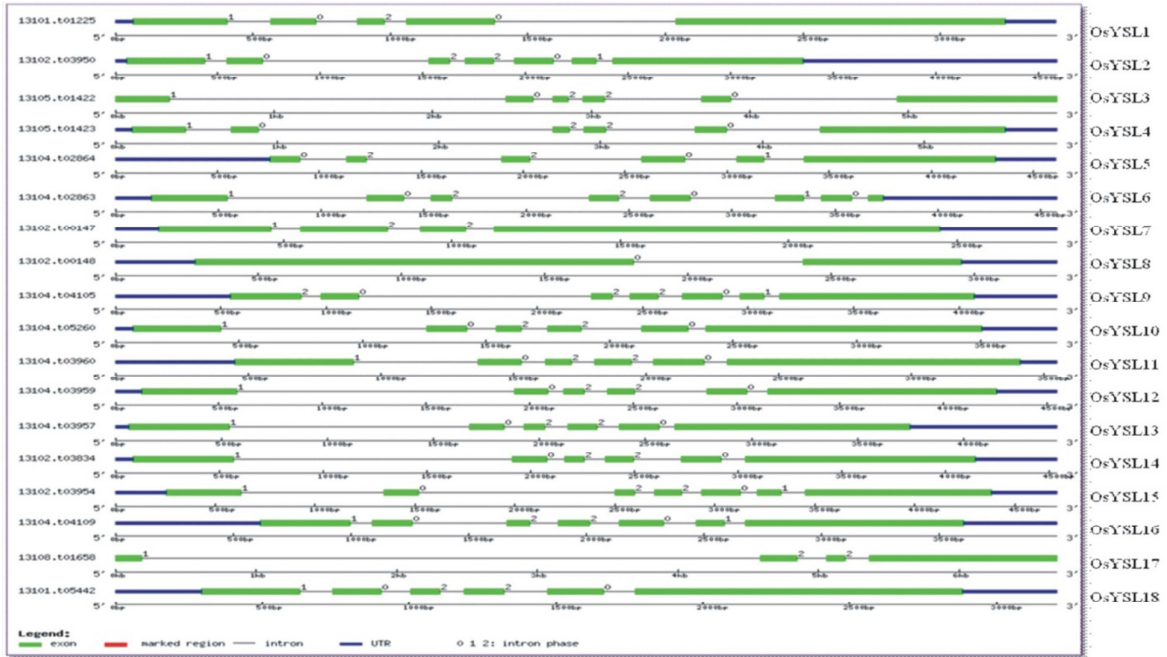
Fig. 3- Intron/ exon structures in conjunction of OsYSL genes in rice. Gene structures were drawn using online tool GSDS. As shown in the legend, green boxes stand for exons, black lines stand for introns, blue boxes are UTR regions, numbers at the exon-intron joints are intron phases.
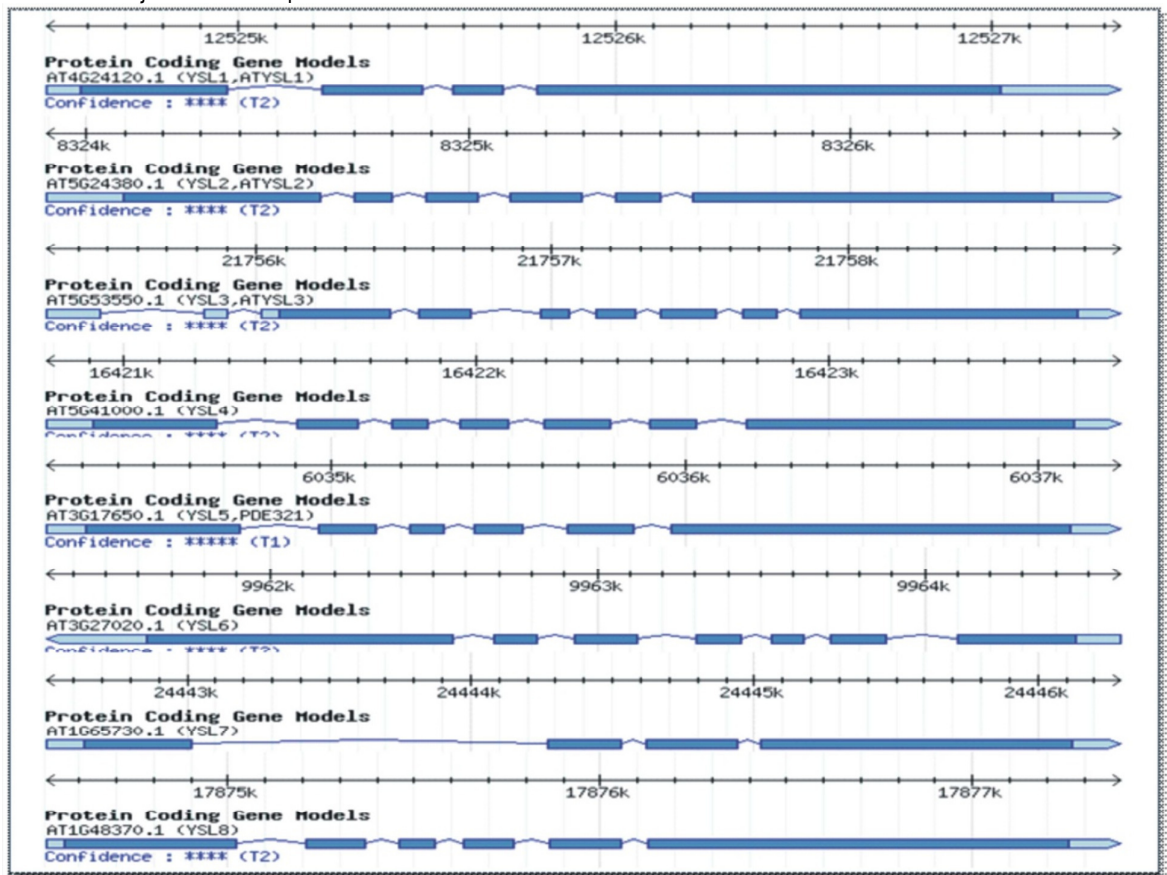
Fig. 4- Intron/exon structures in conjunction of YSL genes in *Arabidopsis*. Gene structures were drawn using online tool NCBI. As shown the blue boxes stand for exons, blue lines stand for introns and light blue boxes are UTR regions.
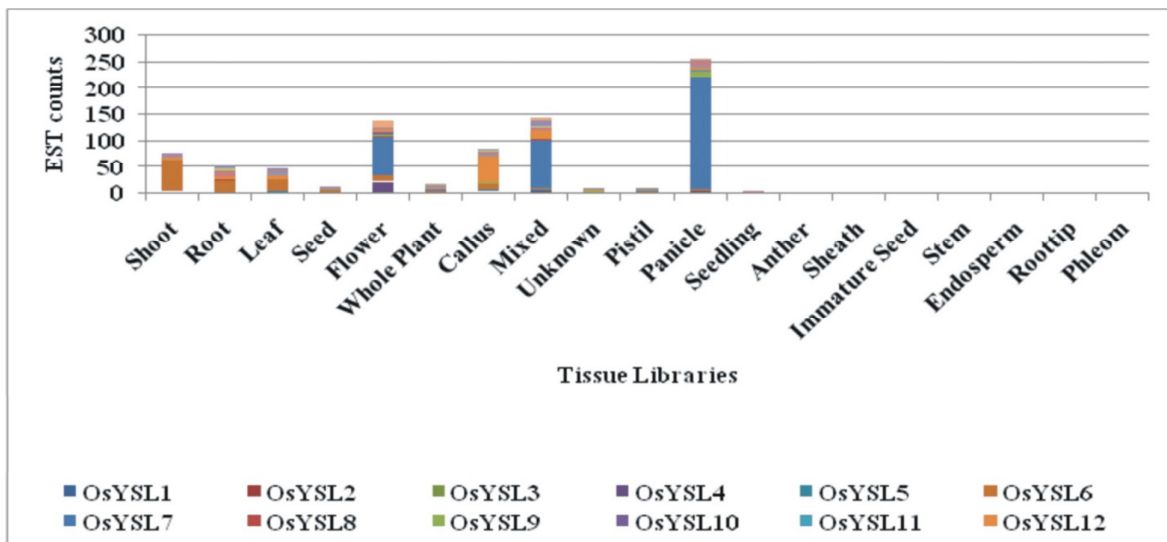
Fig. 5- EST count distribution in different tissue libraries concerned with different YSL genes in rice.
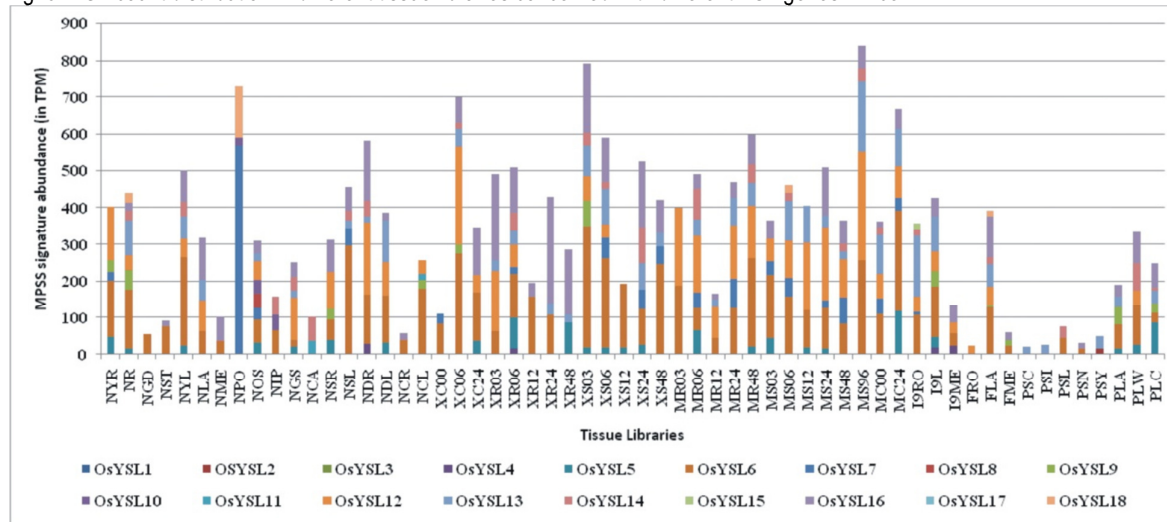
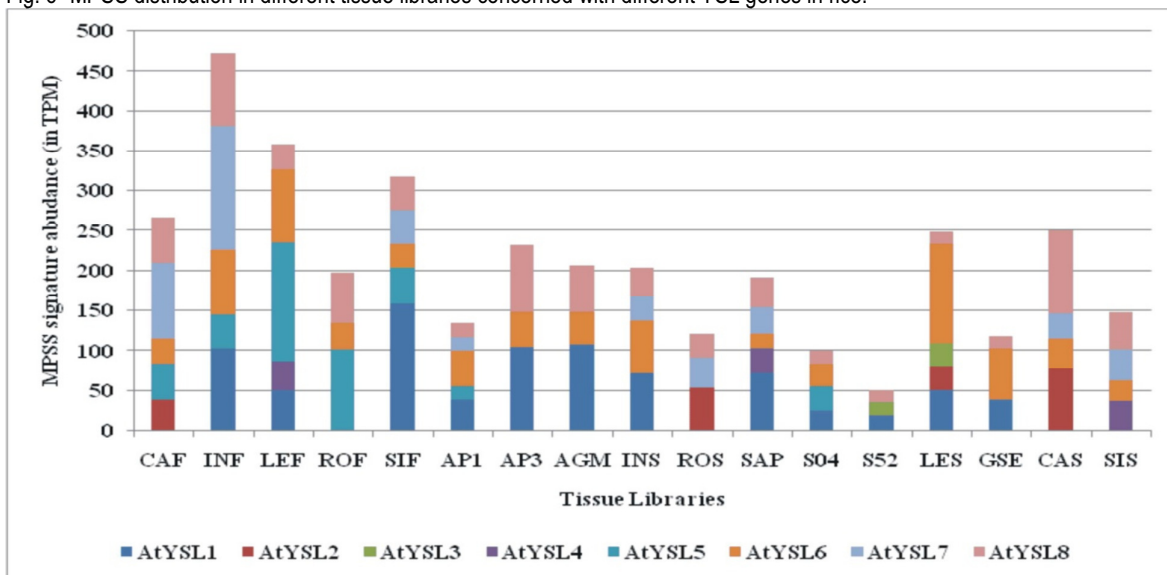Fig. 6- MPSS distribution in different tissue libraries concerned with different YSL genes in rice.

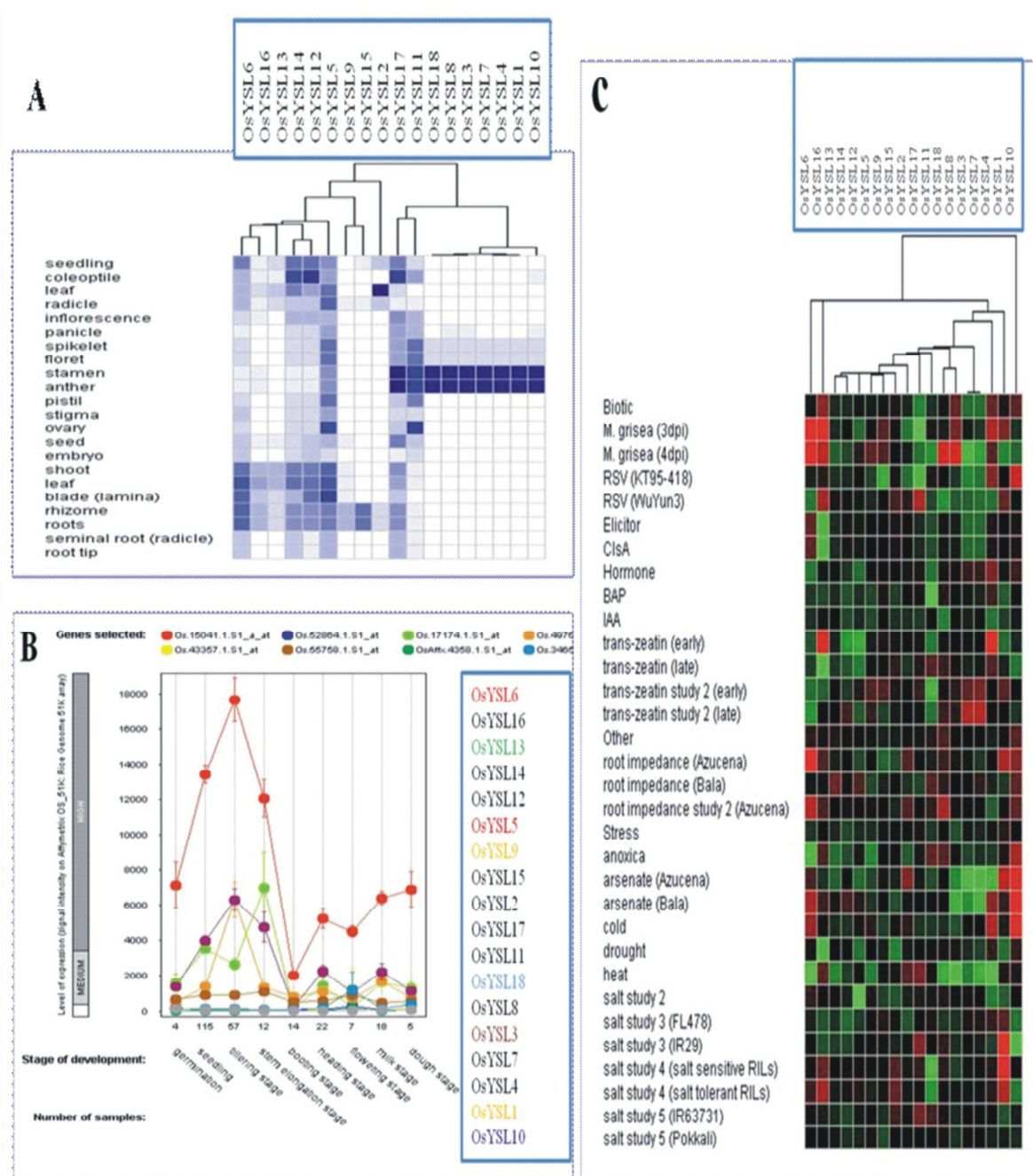Fig. 7- MPSS distribution in different tissue library concerned with different YSL genes in *Arabidopsis.*

Fig. 8- Digital expression pattern analysis of YSL genes in Rice. Based on public microarray data, digital expression pattern was performed by Genevestigator. Rice YSL family genes expression patterns at different development stages, in anatomical tissues and under different stimulus are shown. Hierarchical clustering was played in data analysis as shown in figure A, B and C.
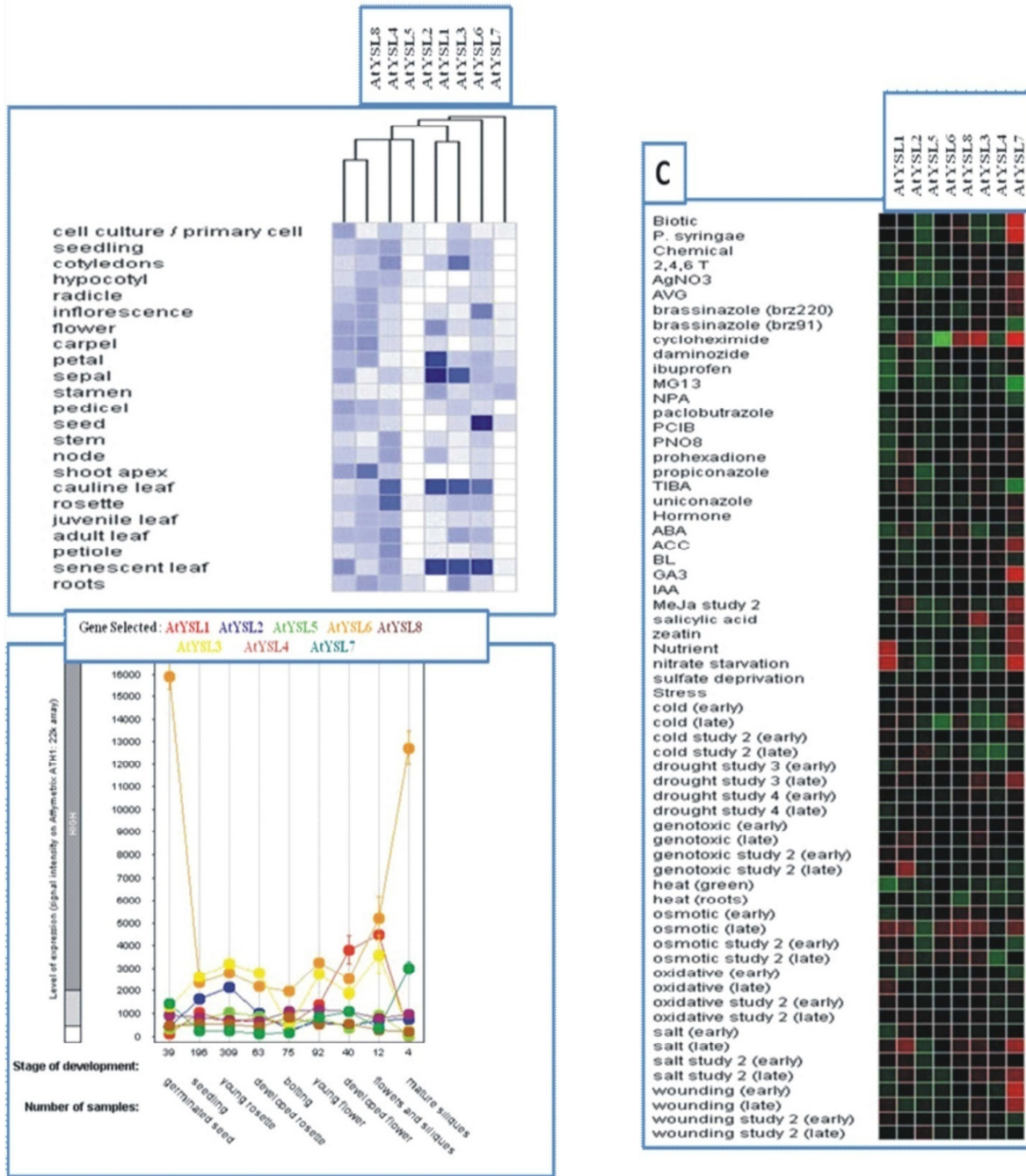
Fig. 9- Digital expression pattern analysis of YSL genes in *Arabidopsis*. Based on public microarray data, digital expression pattern was performed by Genevestigator. *Arabidopsis* YSL family genes expression patterns at different development stages, in anatomical tissues and under different stimulus are shown. Hierarchical clustering was played in data analysis as shown in figure A, B and C.
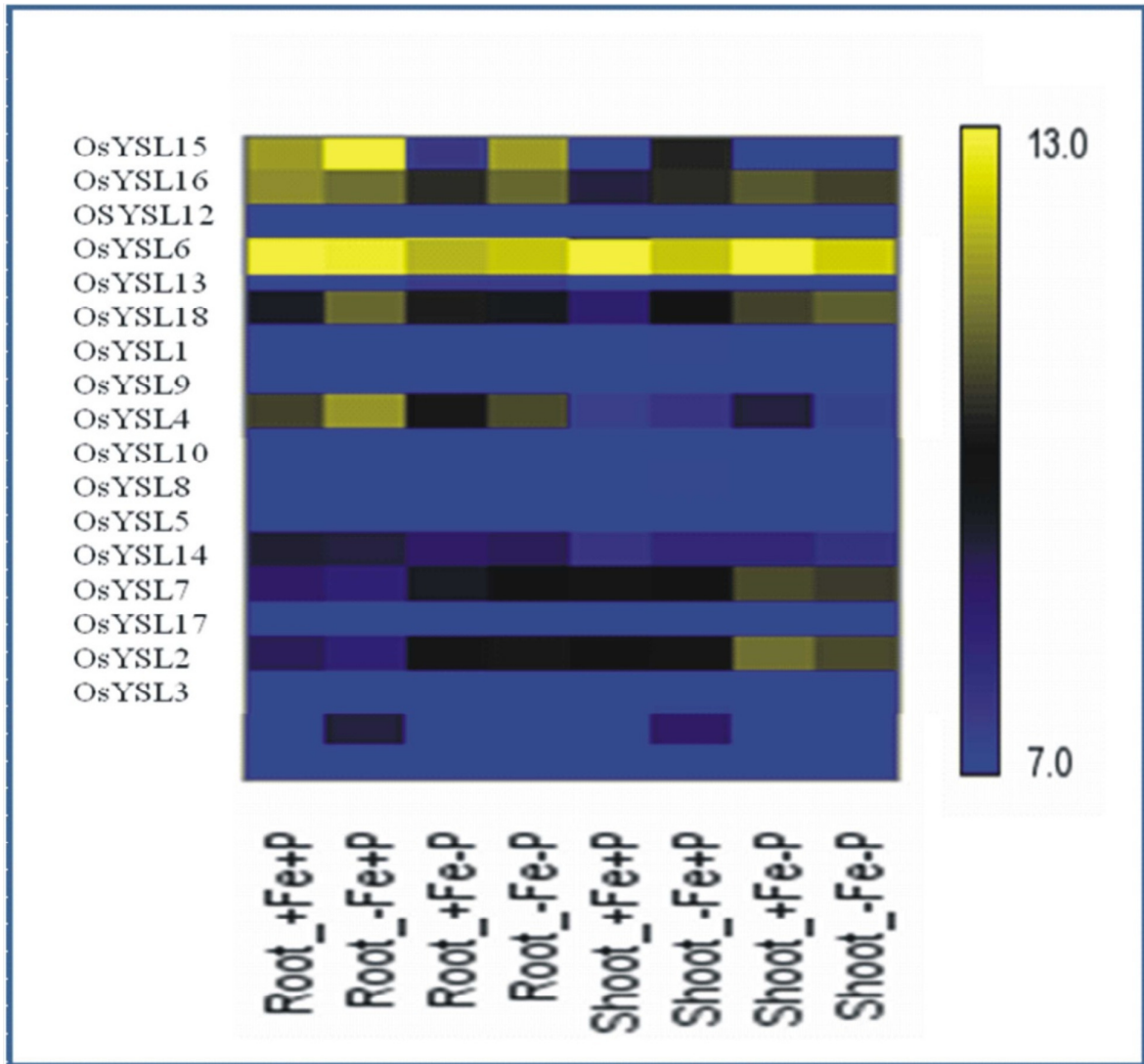
Fig.10- Expression of OsYSL genes under the iron and phosphorous interaction in root and shoot tissues.