UNIVERSITÀ DEGLI STUDI DI VERONA
DIPARTIMENTO DI BIOTECNOLOGIE

DOTTORATO DI RICERCA IN BIOTECNOLOGIE APPLICATE
CICLO XXII

EFFECTS OF *ENOD40* OVEREXPRESSION IN NON-LEGUME PLANTS

Coordinatore: Prof. Massimo Delledonne
Tutore: Prof. Marisa Levi

Dottoranda: Dott. Chiara Francesca Guarnerio

ANNO ACCADEMICO 2009-2010

.

EFFECTS OF *ENOD40* OVEREXPRESSION IN NON-LEGUME PLANTS

Chiara Guarnerio

.

CONTENTS

.

Chapter 1

General Introduction

.

.

ENOD40 is one of the Early Nodulin (*ENOD*) genes. These genes have been identified for the first time as nodule-specific (van Kammen, 1984). They are induced very early upon the interaction of legume plants with symbiotic *Rhizobium* bacteria, indicating that they play a crucial role for the establishing of the symbiosis between legume plants and rhizobia (Kouchi and Hata, 1993; Yang *et al*., 1993; Crespi *et al.*, 1994; Asad, 1994). Among the *ENOD* genes, *ENOD40* is one of the earliest induced (Takeda *et al.*, 2005; Charon *et al.*, 1999). It has been identified in several plant species, both in legume and in non legume plants (Kumagai *et al.*, 2006; Wan *et al*., 2007; Fang and Hirsch, 1998; Kouchi *et al*., 1999; Vleghels *et al.*, 2003), and multiple *ENOD40* copies were found in the genome of many species (Yang *et al*, 1993; Varkonyi-Gasic and White, 2002; Flemetakis *et al*., 2000; Wan *et al.*, 2007). Despite several researches, its precise function is still unclear. To date, only a Mu-transposon *ENOD40*-tagged line has been described in maize, but no phenotypic growth aberration was observed (Compaan *et al.*, 2003).

ENOD40 in legume nodule

The highest expression level of *ENOD40* is during nodule development, therefore the gene has been extensively studied in legume plants. Upon inoculation with *Rhizobium* bacteria, *ENOD40* is highly induced within few hours in the root pericycle cells opposite to the protoxylem pole (Compaan *et al*., 2001; Minami *et al*., 1996) and, subsequently, in the dividing cortical cells and in the nodule primordium (Yang *et al.*, 1993; Fang and Hirsch, 1998; Compaan *et al.*, 2001). In mature nodule, *ENOD40* is expressed in the uninfected cells of the central tissue and in the pericycle of the nodule vascular bundles (Yang *et al.*, 1993).

The spatial and temporal expression patterns of *ENOD40* suggest that this gene is probably implicated in nodule initiation and development. RNAi knock-down in *L. japonicus* (Kumagai *et al.*, 2006) and downregulation due to co-suppression in *M. truncatula* (Charon *et al.*, 1999) lead to a suppression of nodule primordium initiation and subsequently of nodule development, even if bacterial infection process is not inhibited. Recently, a putative second *MtENOD40* gene was identified and a reduced number of nodules was confirmed using gene-specific RNAi knock-down of the two *MtENOD40* genes (Wan *et al.*, 2007). The reduction of the expression of both

.

MtENOD40 genes leads to a higher reduction of nodule initiation, suggesting that the effect of *MtENOD40* on nodule initiation is dose dependent (Wan *et al.,* 2007).
However, overexpression of *ENOD40* in *M. truncatula* leads to changes in the dynamic of nodulation, but not to the formation of more nodules (Charon *et al.,* 1997; Charon *et al.,* 1999). These data show that *ENOD40* is required for nodule primordium formation as well as for subsequent nodule organogenesis.

ENOD40 is not only related to symbiotic interactions

Homologues of *ENOD40* gene have been identified in non legume plants like rice (Kouchi *et al.,* 1999), maize (Compaan *et al.,* 2003), tobacco (Matvienko *et al.,* 1996), tomato (Vleghels *et al.,* 2003), ryegrass and barley (Larsen, 2003) and in many other species belonging to various families across the plant kingdom, such as *Salicaceae, Rosaceae, Rutaceae* and *Poaceae* (Ruttink, 2003),
In non legume plants, *ENOD40* expression has been found associated with vascular tissue development, new organ initiation and in developmental processes in general. In rice, *ENOD40* expression was detected at the early stages of the development of lateral vascular bundles that conjoin the emerging leaf (Kouchi *et al.,* 1999). In tomato, it was found expressed in flower and lateral root development, in germinated seeds and in vascular tissue of stems, young leaves and roots (Vleghels *et al.,* 2003). These expression patterns were also observed in tissues of legume plants, underlying also in this family a role for this gene not related to the symbiosis. Indeed, *ENOD40* transcripts were detected in root and stem vascular tissues (Varkonyi-Gasic and White, 2002) and in non symbiotic meristematic tissues, such as developing lateral roots (Asad, 1994), (Papadopoulou *et al.,* 1996; Fang and Hirsch, 1998), young leaves, stipule primordia (Asad, 1994; Corich *et al.*, 1998) and embryonic tissues (Flemetakis *et al.*, 2000).
Therefore, *ENOD40,* apart from its involvement in root nodule development, plays a role in other developmental processes. It has been suggested that the *ENOD40* gene plays a general role in plant development and, probably, it may have been recruited from another developmental pathway into nodulation during evolution (Szczyglowski and Amyot, 2003).

.

Although putative functions, as transport, organogenesis and regulation of phytohormone status have been proposed (Ruttink, 2003), the function of *ENOD40* gene has not been so far elucidated.

ENOD40 structure

Comparison of *ENOD40* sequences across the plant kingdom reveals that the length of all *ENOD40* cDNA sequences is between 400 and 800 bases, with relatively low overall sequence homology (Ruttink, 2003).

A common feature of *ENOD40* genes is the absence of a long open reading frame (ORF) and the presence of several small ORFs (Sousa *et al.*, 2001). Moreover, at nucleotide level, two regions, designated box1 and box2, are highly conserved among all *ENOD40* genes (Vijn *et al.*, 1995; Kouchi *et al.*, 1999). Box1 is present at 5' end of the gene, whereas box2 is in the central part. The stretches of box1 and box2 span about 30-40 bp and 60-100 bp, respectively (Ruttink, 2003).

The box 1 region has a short ORF (ORF1), which is highly conserved among all species, with the exception of *Casuarina glauca* (Santi *et al.*, 2003), and encodes for a putative peptide of 10-13 amino acids (Compaan *et al.*, 2001; Rohrig *et al.*, 2002; Sousa *et al.*, 2001). The consensus sequence of the peptide is -W-(X4)-HGS (Ruttink, 2003).

Several approaches were carried out to verify the translation and the biological activity of ORF1. Translational fusion between *ENOD40* ORF1 and GFP or GUS have shown that this ORF is recognised by the plant translational machinery (Compaan *et al.*, 2001), (Podkowinski *et al*., 2009). Moreover, *ENOD40* ORF1 translation was shown in *in vitro* assay in *Medicago truncatula* (Sousa *et al.*, 2001) and in *Glycine max* (Rohrig *et al.*, 2002). In particular, *in vitro* translation of *GmENOD40* mRNA revealed that the box1 encodes two peptides of 12 and 24 amino acids (peptides A and B). These peptides are synthesized *de novo* from very short overlapping ORFs and both bind a subunit of sucrose synthase (Rohrig *et al.*, 2002; Rohrig *et al.*, 2004). A role suggested for *ENOD40* peptides A and B was the control of photosynthate use in plants (Rohrig *et al*., 2002), which is congruent with *ENOD40* expression pattern in the vascular bundles of mature soybean nodules (Kouchi and Hata, 1993). In *Medicago truncatula* it has been proposed that the peptide encoded by box1 is responsible for the biological activity of

.

MtENOD40 gene (Wan, 2007). The suggested role of the peptide is the involvement in nodule initiation and at later stages in nodule development (Wan, 2007).
Direct administration of tobacco synthetic peptide to *Arabidopsis thaliana* protoplasts resulted in reduced cell expansion, and the same result was obtained when protoplasts were transiently transfected with *NtENOD40* gene, suggesting the biological activity of the peptide (Guzzo *et al.*, 2005).
These data support the translation and the biological activity of ORF1, but, to date, the peptide has never been biochemically detected *in vivo*.

In box 2 region, no ORFs are found that are absolutely conserved. Only 50% of all known *ENOD40* genes contain a translatable ORF (ORF2), that starts in the middle of the box 2 and encodes the MANRQVTKRQ (or similar) peptide motive (Ruttink, 2003). Translational fusions between *NtENOD40*-ORF2 and GFP resulted in fusion protein production only in the absence of ORF1 in cowpea protoplasts (Compaan *et al.*, 2001), and microtargeting experiments show that in alfalfa roots box2 could be translated in a peptide (Sousa *et al.*, 2001). However, box2 is strongly conserved only at nucleotide level and this suggests that its biological activity could be RNA mediated (Ruttink, 2003; Wan, 2007).
Actually, another structural feature of *ENOD40* genes is the presence of regions corresponding to conserved secondary structures of the transcript (Hofacker *et al.*, 2002; Girard *et al.*, 2003; Gultyaev and Roussis, 2007).
Combining a theoretical approach (computer-assisted predictions) and experimental data, six conserved domains, named domains 1-6 (Fig.1.1), were identified in *ENOD40* mRNA. Structure comparison suggests that the domains 2 and 3 are absolutely conserved in all legume and non-legume species and together form the conserved core consensus RNA structure (Gultyaev and Roussis, 2007). These two domains flank the core of box 2, which has the highest level of sequence similarity shared by the *ENOD40* genes (Vijn *et al.*, 1995; Kouchi *et al.*, 1999; Ruttink, 2003). The domain 1, located upstream of the domain 2, is conserved mainly in legumes (Sousa *et al.*, 2001; Hofacker et al., 2002; Girard *et al.*, 2003). Moreover, domain 4 is found only in legume plants developing indeterminate nodules, whereas the domains 5 and 6 are typical of all the legume plants. Enzymatic and chemical probing data support the structure of the

.

domains 1, 3 and 5, and partially of domains 2 and 6 (Girard *et al.*, 2003). The rest of the molecule appears to be less structured (Girard *et al.*, 2003); also the non legume ENOD40 RNAs seem to be less structured compared to those of legumes (Girard *et al.*, 2003).

Recently, a search for RNA domains forming secondary structures similar to those observed in ENOD40 RNA, led to the identification of a number of *ENOD40*-like genes from different angiosperm orders, including the *Myrtales*, *Malvales*, *Brassicales*, *Apiales* and *Gentianales* (Gultyaev and Roussis, 2007), in which sequence homologues were not previuosly found.

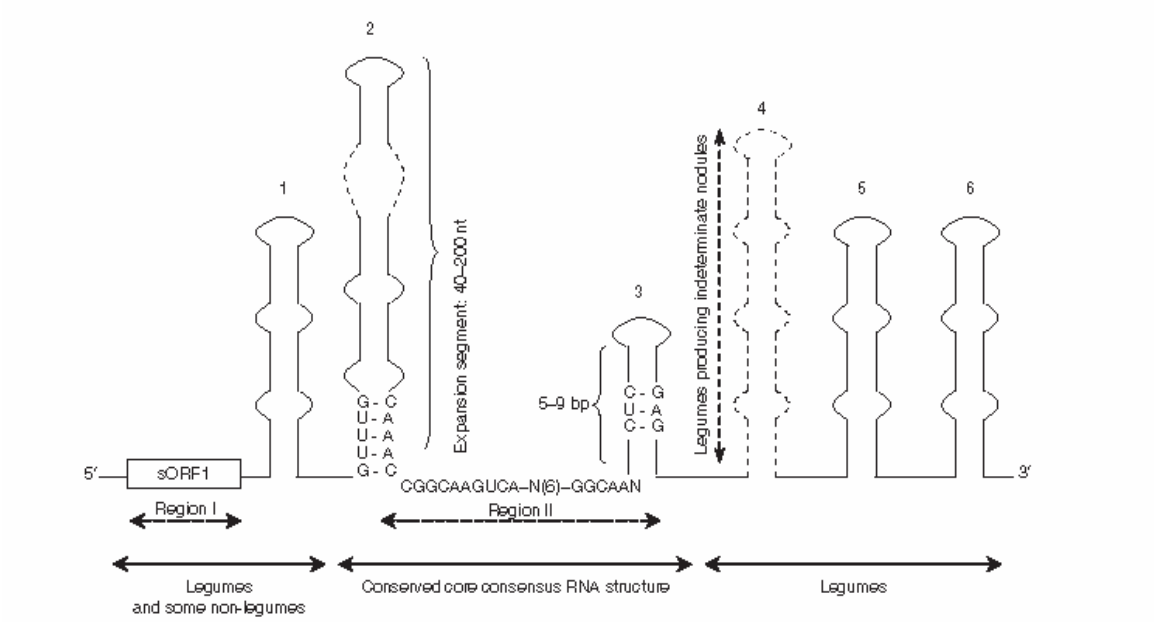


Fig. 1.1 The conserved domains (1-6) identified in ENOD40 RNAs *(Gultyaev et al.,Nucleic Acids Res. 2007)*

The conserved RNA structural domains of *ENOD40* genes suggest that RNA could play a crucial role in the biological activity of *ENOD40.* This hypothesis was supported by the identification of the RNA-binding protein MtRBP1 in *M. truncatula,* which interacts with the *MtENOD40* RNA in the nucleus, and is exported into the cytoplasm during legume nodule development in the same region that expressed *MtENOD40* (Campalans *et al.*, 2004).

.

As shown by the above mentioned literature, if the biological activity of *ENOD40* is mediated by the peptide(s), RNA or both, remains unclear. Probably, for some of *ENOD40* functions the secondary structures of the conserved domains are required, whereas ENOD40 encoded peptide may be responsible for other roles.

ENOD40 in *Arabidopsis thaliana*

Sequence comparisons, based on the highly conserved nucleotide sequences of box 2, were done in the past to identify *ENOD40* homologues in *Arabidopsis thaliana* genome (Ruttink, 2003). Several sequences with homology to box 2 were identified in Arabidopsis. However, these sequences do not span the entire conserved nucleotide sequence of box 2, and ORF1, present in all *ENOD40* known genes, was not recognized upstream of these sequences. Since neither motives were present in Arabidopsis genome, it was concluded that Arabidopsis did not contain an *ENOD40* homologue (Ruttink, 2003).

The analysis of soybean *ENOD40(2)* promoter activity in Arabidopsis shows that it is active and regulated during plant development. The expression pattern is similar in Arabidopsis and legume plants, indicating that regulation mechanisms are conserved among these plants (Mirabella *et al.,* 1999). Moreover, Brassicaceae ancestor still contained an *ENOD40* homologue, suggesting that probably Arabidopsis had lost the gene (Ruttink, 2003).

Recently, it was demonstrated that *Arabidopsis thaliana* contains the RNA structural domains, 2 and 3, conserved in all *ENOD40* genes (Gultyaev and Roussis, 2007). The cDNA sequence obtained by these conserved elements was validated by genomic BLAST comparison, showing a sequence identity in Arabidopsis genome. Thus, *ENOD40* gene is present also in Arabidopsis, even if its nucleotide sequence do not contain the conserved ORF1 typical of all *ENOD40* genes (Gultyaev and Roussis, 2007).

Arabidopsis thaliana was used to study the effect of overexpression of soybean *ENOD40* gene to investigate the role of *ENOD40* in plant development (Guzzo *et al.*, 2005). These studies showed that *ENOD40* overexpression induces a reduction of cell size in various tissues, although the organs maintain the normal size. The morphological cell phenotype was observed both in plants transformed with *ENOD40* gene and in protoplasts transiently expressing the gene (Guzzo *et al.*, 2005).

.

AIM OF THE WORK

ENOD40 gene has been identified in different plants species and its expression has been found also in tissues not related with symbiosis, suggesting a general role of the gene in plant development. Despite several researches, the roles of the *ENOD40* gene has not been so far completely elucidated. Moreover, whether the biological activity should be ascribed to RNA or peptide, or both, is still unclear.

The main goals of the present research are:

1. to investigate the possible presence of the putative peptide encoded by box1 of the *ENOD40* gene in BY-2 *ENOD40* overexpressing cells
2. to investigate the role of *ENOD40* gene in non legume plants, using *Arabidopsis thaliana.*

In detail, in order to biochemically detect the putative peptide encoded by the box1 of *ENOD40* gene, a purification procedure was set up in BY-2 cells overexpressing *NtENOD40* gene. The entire procedure and the mass spectrometry analysis are illustrated in Chapter 2.
Chapter 3 and 4 analyse the metabolic and transcriptional profiles of three Arabidopsis lines overexpressing soybean *ENOD40* (*GmENOD40*), compared to wild type plants. The Arabidopsis lines used in this work were the same used in Guzzo *et al*. (2005), in which a reduction of cell size in various tissues was observed.
In Chapter 5 all data are comprehensively discussed and new hints on the features of the intriguing *ENOD40* gene are proposed.

.

REFERENCES

Asad (1994) Isolation and characterization of cDNA and genomic clones of *MsENOD40*; Transcripts are detected in meristematic cells of alfalfa. Protoplasma 183: 10-23

Campalans A, Kondorosi A, Crespi M (2004) Enod40, a short open reading frame-containing mRNA, induces cytoplasmic localization of a nuclear RNA binding protein in Medicago truncatula. Plant Cell 16: 1047-1059

Charon C, Johansson C, Kondorosi E, Kondorosi A, Crespi M (1997) enod40 induces dedifferentiation and division of root cortical cells in legumes. Proc Natl Acad Sci U S A 94: 8901-8906

Charon C, Sousa C, Crespi M, Kondorosi A (1999) Alteration of enod40 expression modifies *Medicago truncatula* root nodule development induced by sinorhizobium meliloti. Plant Cell 11: 1953-1966

Compaan B, Ruttink T, Albrecht C, Meeley R, Bisseling T, Franssen H (2003) Identification and characterization of a *Zea mays* line carrying a transposon-tagged *ENOD40*. Biochim Biophys Acta 1629: 84-91

Compaan B, Yang, W.-C., Bisseling, T., and Franssen, H. (2001) *ENOD40* expression in the pericycle precedes cortical cell division in *Rhizobium*-legume interaction and the highly conserved internal region of the gene does not encode a peptide. Plant Soil 230: 1–8

Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M (1998) Patterns of *ENOD4*0 gene expression in stem-borne nodules of *Sesbania rostrata*. Plant Mol Biol 37: 67-76

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. Embo J 13: 5099-5112

Fang Y, Hirsch AM (1998) Studying early nodulin gene *ENOD40* expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol 116: 53-68

Flemetakis E, Kavroulakis N, Quaedvlieg NE, Spaink HP, Dimou M, Roussis A, Katinakis P (2000) Lotus japonicus contains two distinct *ENOD40* genes that are expressed in symbiotic, nonsymbiotic, and embryonic tissues. Mol Plant Microbe Interact 13: 987-994

Girard G, Roussis A, Gultyaev AP, Pleij CW, Spaink HP (2003) Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean. Nucleic Acids Res 31: 5003-5015

Gultyaev AP, Roussis A (2007) Identification of conserved secondary structures and

.

expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res 35: 3144-3152

Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by *ENOD40* in *Arabidopsis thaliana*. J Exp Bot 56: 507-513

Hofacker IL, Fekete M, Stadler PF (2002) Secondary structure prediction for aligned RNA sequences. J Mol Biol 319: 1059-1066

Kouchi H, Hata S (1993) Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. Mol Gen Genet 238: 106-119

Kouchi H, Takane K, So RB, Ladha JK, Reddy PM (1999) Rice *ENOD40*: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J 18: 121-129

Kumagai H, Kinoshita E, Ridge RW, Kouchi H (2006) RNAi knock-down of *ENOD40s* leads to significant suppression of nodule formation in *Lotus japonicus*. Plant Cell Physiol 47: 1102-1111

Larsen K (2003) Molecular cloning and characterization of a cDNA encoding a ryegrass (*Lolium perenne*) ENOD40 homologue. J Plant Physiol 160: 675-687

Matvienko M, vande Sande, K., Pawloski, K., van Kammen, A., and, Bisseling T (1996) *Nicotiana tabacum* SR1 contains two *ENOD40* homologs. Biology of Plant Microbe Interactions.: 387-391

Minami E, Kouchi H, Cohn JR, Ogawa T, Stacey G (1996) Expression of the early nodulin *ENOD40*, in soybean roots in response to various lipo-chitin signal molecules. Planta J 10: 23-32

Mirabella R, Martirani L, Lamberti A, Iaccarino M, Chiurazzi M (1999) The soybean *ENOD40(2)* promoter is active in Arabidopsis thaliana and is temporally and spatially regulated. Plant Mol Biol 39: 177-181

Papadopoulou K, Roussis A, Katinakis P (1996) *Phaseolus ENOD40* is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. Plant Mol Biol 30: 403-417

Podkowinski J, Zmienko A, Florek B, Wojciechowski P, Rybarczyk A, Wrzesinski J, Ciesiolka J, Blazewicz J, Kondorosi A, Crespi M, Legocki A (2009) Translational and structural analysis of the shortest legume *ENOD40* gene in *Lupinus luteus*. Acta Biochim Pol 56: 89-102

Rohrig H, John M, Schmidt J (2004) Modification of soybean sucrose synthase by S-thiolation with *ENOD40* peptide A. Biochem Biophys Res Commun 325: 864-870

.

Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M (2002) Soybean *ENOD40* encodes two peptides that bind to sucrose synthase. Proc Natl Acad Sci U S A 99: 1915-1920

Ruttink (2003) *ENOD40* affects phytohormone cross-talk. the Netherlands

Santi C, von Groll U, Ribeiro A, Chiurazzi M, Auguy F, Bogusz D, Franche C, Pawlowski K (2003) Comparison of nodule induction in legume and actinorhizal symbioses: the induction of actinorhizal nodules does not involve ENOD40. Mol Plant Microbe Interact 16: 808-816

Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M (2001) Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Mol Cell Biol 21: 354-366

Szczyglowski K, Amyot L (2003) Symbiosis, inventiveness by recruitment? Plant Physiol 131: 935-940

Takeda N, Okamoto S, Hayashi M, Murooka Y (2005) Expression of *LjENOD40* genes in response to symbiotic and non-symbiotic signals: *LjENOD40-1* and *LjENOD40-2* are differentially regulated in Lotus japonicus. Plant Cell Physiol 46: 1291-1298

van Kammen A (1984) Suggested nomenclature for plant genes involved in nodulation and symbiosis. Plant Mol. Biol. Rep. 2: 43-45

Varkonyi-Gasic E, White DW (2002) The white clover enod40 gene family. Expression patterns of two types of genes indicate a role in vascular function. Plant Physiol 129: 1107-1118

Vijn I, Yang WC, Pallisgard N, Ostergaard Jensen E, van Kammen A, Bisseling T (1995) *VsENOD5*, *VsENOD12* and *VsENOD40* expression during *Rhizobium*-induced nodule formation on *Vicia sativa* roots. Plant Mol Biol 28: 1111-1119

Vleghels I, Hontelez J, Ribeiro A, Fransz P, Bisseling T, Franssen H (2003) Expression of *ENOD40* during tomato plant development. Planta 218: 42-49

Wan X (2007) Analysis of Nodule Meristem Persistence and *ENOD40* functioning in *Medicago truncatula* nodule formation. Wageningen

Wan X, Hontelez J, Lillo A, Guarnerio C, van de Peut D, Fedorova E, Bisseling T, Franssen H (2007) *Medicago truncatula ENOD40-1* and *ENOD40-2* are both involved in nodule initiation and bacteroid development. J Exp Bot 58: 2033-2041

Yang WC, Katinakis, P., Hendriks, P., de Vries, F., Spee, J., van Kammen,, A. B, T., and Franssen, H. (1993) Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J. 3: 573-585

.

Chapter 2

Looking for ENOD40 putative peptide

.

.

ABSTRACT

A remarkable feature of all *ENOD40* genes is the lack of a long open reading frame and the presence of two conserved regions called box 1 and box 2. Box 1 seems to encode for a putative peptide of 10-13 amino acids. Several experiments have demonstrated the translation of the box1 region, but no one have revealed biochemically the putative peptide.

In this work, a purification procedure consisting of membrane cut- off, ion exchange chromatography, solid exchange extraction, HPLC-DAD and mass spectrometry was set up to search for the putative peptide in BY-2 tobacco cells.

INTRODUCTION

Comparison of *ENOD40* sequences across the plant kingdom reveals that *ENOD40* genes exhibit several intriguing structural features, such as the absence of a long open reading frame, the presence of two highly conserved regions, named box 1 and box 2 respectively (Vijn *et al.*, 1995), and of regions corresponding to conserved secondary structures of the transcript (Girard *et al*., 2003; Gultyaev and Roussis, 2007). In the box1 region, a conserved ORF is present, that can be translated into a peptide of 10-13 amino acids (Compaan *et al.*, 2001, Rohrig *et al.*, 2002; Sousa *et al.*, 2001) which contains -W-(X4)-HGS as consensus sequence (Ruttink, 2003). Otherwise, the ORF spanning box2 region is present in only about 50% of species and the amino acids sequence of the peptide is less conserved compared to its strong conservation at nucleotide level (Wan, 2007). This suggest that the box2 could more likely function at RNA level, even if ballistic micro-targeting experiments showed that both box1 and box2 could be translated in a peptide (Sousa *et al*., 2001).

If the biological activity of *ENOD40* is mediated by peptide or RNA or both, is still unclear. The ORF1 translation was demonstrated using several approaches. The ENOD40 ORF1 of *Glycine max* and *Medicago truncatula* is translated in *in vitro* assay (Rohrig *et al.*, 2002; Sousa *et al.*, 2001); moreover, translational fusions between *NtENOD40* ORF1-GFP (Compaan *et al.*, 2001) and between *LlENOD40* ORF1-GUS (Podkowinski *et al*., 2009) resulted in fusion protein production, demonstrating that this ORF is recognised by the plant translational machinery. Moreover, the activity of the ENOD40 peptide was shown in *Arabidopsis thalian*a protoplasts, where the transfection

.

with full-length *ENOD40* and the treatment with the synthetic ENOD40 peptide induced the same phenotype (Guzzo *et al.*, 2005). The above observations support the hypothesis of *ENOD40* activity mediated, at least in part, by the peptide.

Up to now, the peptide has never been biochemically detected *in vivo*. Immunological approaches were used to search for the presence of the peptide in alfalfa nodule, but the immunological signal disappeared very rapidly. The peptides were rapidly degraded in nodules and *in vitro* translation extracts, suggesting that immunological approaches may not be adequate to their biochemical detection (Sousa *et al.*, 2001).

In this chapter, a purification procedure to look for the putative peptide encoded by ORF1 in BY-2 tobacco cell is described. Synthetic peptide and an overexpressing *NtENOD40* BY-2 line (Nt-S) were used to set up an attempt of purification procedure, that consists of a previous filtration with low cut-off membrane , followed by ion exchange chromatography, solid exchange extraction, HPLC-DAD (High Performance Liquid Chromatography-Diode Array detector) and mass spectrometry analysis.

MATERIALI E METODI

BY-2 cell lines

Wild type and transformed BY-2 tobacco cell lines were kindly provided by Henk Franssen (Molecular Biology Laboratory, Wageningen University). The Nt-S line was transformed with *NtENOD40-1* gene (470 bp) driven by 35S promoter.

The cell lines were grown in solid MS medium (Murashige and Skoog 1962) supplemented with sucrose 2%, 2.4D 0.2 mg/l and agar 0.7% (pH 5.9) with a 16 h light/ 8 h dark photoperiod and 60 $\mu E/m^2s$ irradiance at 25° C.The solid medium of Nt-S cells was supplemented with ticarciline 0.2 mg/l and hygromycin 0.04 mg/l. Calli were subcultured in fresh medium every 3-4 weeks.

Calli grown for 3 weeks, were ground with liquid nitrogen in a fine powder and stocked at – 80°C.

.

Primer design

All the primers were designed using Primer 3 (http://frodo.wi.mit.edu/).

Primer Name	Primer Sequence (5'-3')
UBI forward	GTCACCTTGTCCTCCGTCTC
UBI reverse	ACAAAGCACATCACGACCAC
NtENOD40 forward	TCCAGAAAATGCAGCAAAAA
NtENOD40 reverse	ATTGCCGTTTCGTGACTTG

RNA Extraction, DNAse Treatment and cDNA synthesis

RNA was extracted using TRIzol® reagent (Invitrogen). After RNA isolation, samples were treated with RQ1 RNase-Free DNase (Promega); 3 μg of RNA were treated with 3 μl RQ1 DNase in total volume of 30 μl to remove any contaminating DNA. The cDNA was synthesized using the Improm-II^{TM} Reverse Transcription System kit (Promega); 1μg of RNA treated with DNase was reverse-transcribed, whereas the RNA not reverse-transcribed was used as negative control in subsequent Real-Time PCR experiments.

Real-Time PCR

Real-time PCR was used to confirm the expression of *NtENOD40*.

PCR reaction mix contained: 5 μl of cDNA diluted 1:10, 12.5 μl of Platinum ® SYBR ® Green qPCR SuperMix UDG with ROX (Invitrogen), 0.5 μl of each primer (stock 20 μM), and sterile double-distilled water to a final volume of 25 μl.

Real-time PCR amplification was performed on an ABI PRISM 7000 sequence detection system (Applied Biosystems) programmed for initial denaturing at 95°C for 10 min, and 40 cycles of denaturing at 95°C for 15 sec, annealing at 55°C for 30 sec and extension at 72°C for 30 sec. The relative level of the expression was calculated in accord to $2^{-\Delta\Delta Ct}$ method as described by (Livak and Schmittgen, 2001).

Chemicals

All chemicals and water LC-MS grade were supplied by Sigma-Aldrich.

.

BY-2 cell extraction

Frozen callus powder was extracted for 4 hours on ice with extraction buffer (protease inhibitor, PMSF and benzamidine, at concentration of 0.2 mM and 2.5% w/v of polyvinylpolypyrrolidone (PVPP) in water). Most experiments were performed using 15 gr of frozen callus powder and 30 ml of buffer solution.

Cell extracts were centrifuged at 17100 x *g* for 25 minutes at 4°C (Sorvall RC 5C Plus centrifuge, rotor SS-34) to eliminate cell debris, and subsequently filtrated with a filter paper. Cell extracts were stored in tubes at -80 °C and thawed on ice just prior to be further processed.

Synthetic peptide

The synthetic peptide, deduced from highly conserved box1 region of *NtENOD40* (MQWDEAIHGS), was synthesized by Primm (Milan, Italy). The peptide has a molecular weight of 1173 Da and contains two residues, aspartic and glutamic acids, that confer to the peptide acidic properties. To set up the attempt purification procedure the synthetic peptide was added to the extract at final concentration of 10^{-4} M or 10^{-6} M.

Peptide purification strategy

The thawed cell extract was further filtrated through a membrane with a cut-off of 5000 Da (VIVASPIN 15R, Sartorius Stedim Biotech) by centrifugation at 3000 x *g* for 2 h at 4°C (Eppendorf 5804 R centrifuge, rotor A-4-4).

The obtained fraction, rich in small oligopeptides (< 5000 Da), was adjusted to 0.01M Tris-HCl (pH 7.5) and applied to a weak anion exchange chromatography column (Discovery® DSC-NH_2, Supelco) equilibrated in Tris-HCl 0.01M (pH 7.5).

The elution was carried out with a discontinuous salt gradient at pH 7.5. The concentrations of NaCl used were 0.1M, 0.3M, 0.5M and 1M. The fraction 0.3M of NaCl, in which the synthetic peptide eluted, was desalted and concentrated by SPE-C18 cartridge (Waters). The cartridge was first conditioned with acetonitrile 80% (v/v) and trifluoroacetic acid (TFA) 0.05% (v/v) in water and subsequently with TFA 0.05% (v/v) in water. The extract was acidified with TFA 0.05% (v/v) and then applied to the cartridge. The fractions were eluted with water containing respectively 10%, 20%, 30%,

.

40% and 80% (v/v) acetonitrile and TFA 0.05% (v/v).
The fractions were evaporated to dryness in a Speed Vac vacuum concentrator and then dissolved in 200 μl deionised water prior to HPLC-DAD analysis or LC-ESI-MS.

HPLC-DAD analysis

HPLC-DAD analysis was performed on a Beckman Coulter System Gold instrument equipped with a 126 solvent module, Diode Array Detector (System Gold 168) and Autosampler (System Gold 507). Chromatografic data were collected and processed using 32 Karat Software version 7.0 (Beckman Coulter Inc. Fullerton, CA).
Solvents were (A): 0.5% (v/v) formic acid and 5% (v/v) acetonitrile in water, and (B): 100% acetonitrile. Gradient profile was 0-35 min: linear 50% solvent B; a 15 min equilibration time followed each sample analysis. The column used was Alltima HP C18 3 μm 150 mm x 2.1 mm, with a guard column 7.5 mm x 2.1 mm (Alltech Associates, Inc., Derfield, IL). The injection volume was 20 μl. and the flow rate was 200 μl/min.

LC-ESI-MS analysis

LC-ESI-MS was performed using an HPLC system (Beckman Coulter System Gold 127, Solvent Module) coupled on-line with the Bruker ion trap mass spectrometer Esquire 6000, equipped with an ESI source. Data were collected using the Esquire Control 5.2 software, and processed using the Data Analysis 3.2 software (Bruker Daltonics Esquire 5.2, Bruker Daltonik GmbH). Due to the negative charge of the expected peptide, the analysis was performed in negative mode. Chromatography conditions were the same of the HPLC-DAD analysis.

Negative ion mass spectra were recorded in the range of 50-1500 m/z (full scan mode, 13,000 m/z s^{-1}). Nitrogen was used as nebulising gas (pressure: 50 psi, temperature 350°C) and drying gas (12 l min^{-1}). Helium was used as collision gas.
For tandem mass spectrometry analysis, ms/ms and ms3 spectra were recorded, in the range of 50–1500 m/z, with a fragmentation amplitude of 1 V.
For Multiple Reaction Monitoring (MRM) analysis, the pattern of the peptide

.

fragmentation was preliminary determined by direct off-line injection of the synthetic peptide into the mass spectrometer. Then, based on these previous observations, the isolation of a fragment of m/z=1172, corresponding to the m/z of the peptide, and its following fragmentation and isolation of the m/z= 1153 fragment, corresponding to the main product of the synthetic peptide, were selected. The other mass spectrometer parameters were set up as follows.

Source: capillary +4000 V, end plate offset -500 V; skimmer: -40 V; cap exit -121 V; Oct.1DC: -12 V; Oct.2DC: -1.70 V; lens 1: 5 V; lens 2: 60 V.

MALDI-TOF analysis

MALDI-MS measurements were performed using a MALDI-TOF Ultraflex (Bruker Daltonics) operating in reflectron positive ion mode. Ions were formed by a pulsed UV laser (λ=337 nm) beam. The pulsed ion extraction conditions were as follows: IS1, 25 kV; IS2, 21.70 kV; reflectron, 26.30 kV; delay time, 150 ns. The signal of 100 shots was accumulated to obtain a spectrum. The matrix used was α-cyano-4-hydroxycinnamic acid (HCCA, LaserBioLabs). The dry purified extract was dissolved in 20 μl of 0.1% TFA in deionised water and then the extract was furthermore desalted and purified by ZipTip C18 pipette tips (Millipore), following the procedure described in the ZipTip user's guide. 1 μl of the sample, diluted 1:20 with the matrix solution, was deposited on the stainless steel sample holder, and allowed to dry before introduction into the mass spectrometer. External mass calibration was done using the Peptide Calibration Standard, based on the monoisotopic values of $[M+H]^+$ of Angiotensin II, Angiotensin I, Substance P, Bombesin, ACTH clip (1-17), ACTH clip (18-39) and Somatostatin 28 at m/z 1046.5420, 1296.6853, 1347.7361, 1619.8230, 2093.0868, 2465.1990 and 3147.4714, respectively.

Post source decay (PSD) was also used. PSD technique is based on collisional activation; during MALDI analysis a large fraction of the desorbed analyte ions undergoes delayed fragmentation reactions (occurring during the flight) and the *m*/*z* values of related decomposition ions can be determined by a reflectron time-of-flight (RETOF) analyzer (Spengler 1992). Ion selection and mass calibration for the PSD experiments were performed using the FAST® method.

.

RESULTS

Analysis of Nt-S transformed line for the expression of *NtENOD40*

The expression of *NtENOD40* gene in the transformed cells (Nt-S line) was confirmed by Real-Time PCR using NtENOD40 specific primers. RNA extracted from Nt-S line was reverse-transcribed and cDNA was employed as template for PCR (Fig.2.1). Ubiquitin gene was used as a reference to normalize gene expression.

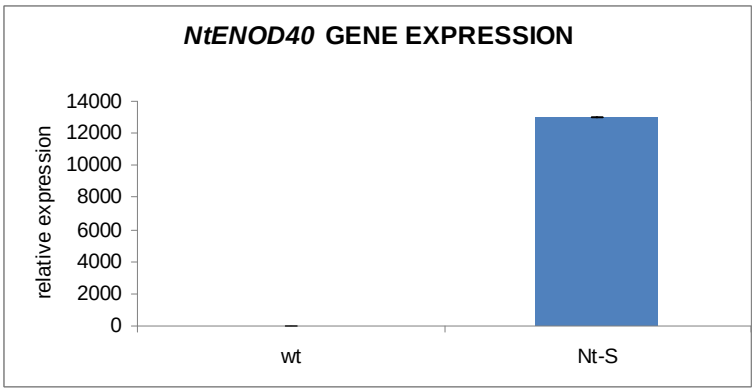


Fig.2.1 Real-Time RT-PCR for the expression o f *NtENOD40* in Nt-S line of BY-2. Bar represents mean ± SE of RNA samples extracted from three different calli. In the wild type *NtENOD40* expression could not be detected, indicating a very low expression level of the endogenous *NtENOD40* gene.

Purification procedure for NtENOD40 putative peptide

Physiological amount of ENOD40 putative peptide in plants is estimated to be very low. Protoplasts directly treated with different concentrations of ENOD40 synthetic peptide showed the strongest effect on cell size reduction at a concentration of 10^{-6} M; however, small effects were also revealed at concentration of 10^{-7} M and 10^{-8} M (Guzzo *et al.*, 2005).

In order to set up a purification strategy, the synthetic peptide NtENOD40 (MQWDEAIHGS) was used as reference standard. The analytical procedure was initially set up using the synthetic peptide. Afterward, a temptative purification procedure was set up using Nt-S cell extracts additioned with ENOD40 peptide from 10^{-6} to 10^{-4}M. A four-step purification procedure, achieved using a previous filtration

.

through a membrane with low cut-off, followed by ion exchange chromatography, solid-phase extraction and reverse phase HPLC-DAD was set up.

HPLC-DAD was used to analyse all the fractions of each purification step, in order to detect the fraction in which the synthetic peptide eluted. The detailed procedure involved the following steps: Nt-S extracts, with and without the synthetic peptide, were filtrated through a membrane with 5000 Da cut off (VIVASPIN 15R, Sartorious) to obtain a fraction rich in small oligopeptides.

The filtrated samples were pre-fractionated by an anion exchange column (DSC-NH_2) using a discontinuous elution and increasing the concentration of Cl^- ions. Several concentrations of NaCl (0.1M, 0.3M, 0.5M and 1M) were tested to elute the synthetic peptide in a single fraction. The synthetic peptide was collected in the fraction eluted using 0.3M NaCl, and the recovery efficiency was around 92 %. This fraction was further purified and concentrated by a C18 solid-phase extraction cartridge. Using this support, that retains differentially various kind of molecules depending on hydrophobic interactions, the peptide was eluted, as demonstrated by HPLC-DAD analysis, with acetonitrile 30% and trifluoroacetic acid 0,05% in water. After this last step of procedure the recovery efficiency was around 86 %.

Figure 2.2 shows HPLC-DAD chromatogram of the Nt-S extract additioned with the synthetic peptide 10^{-4} M after the above purification procedure. One peak showed retention time and UV-VIS spectrum comparable to those of the synthetic peptide.

.

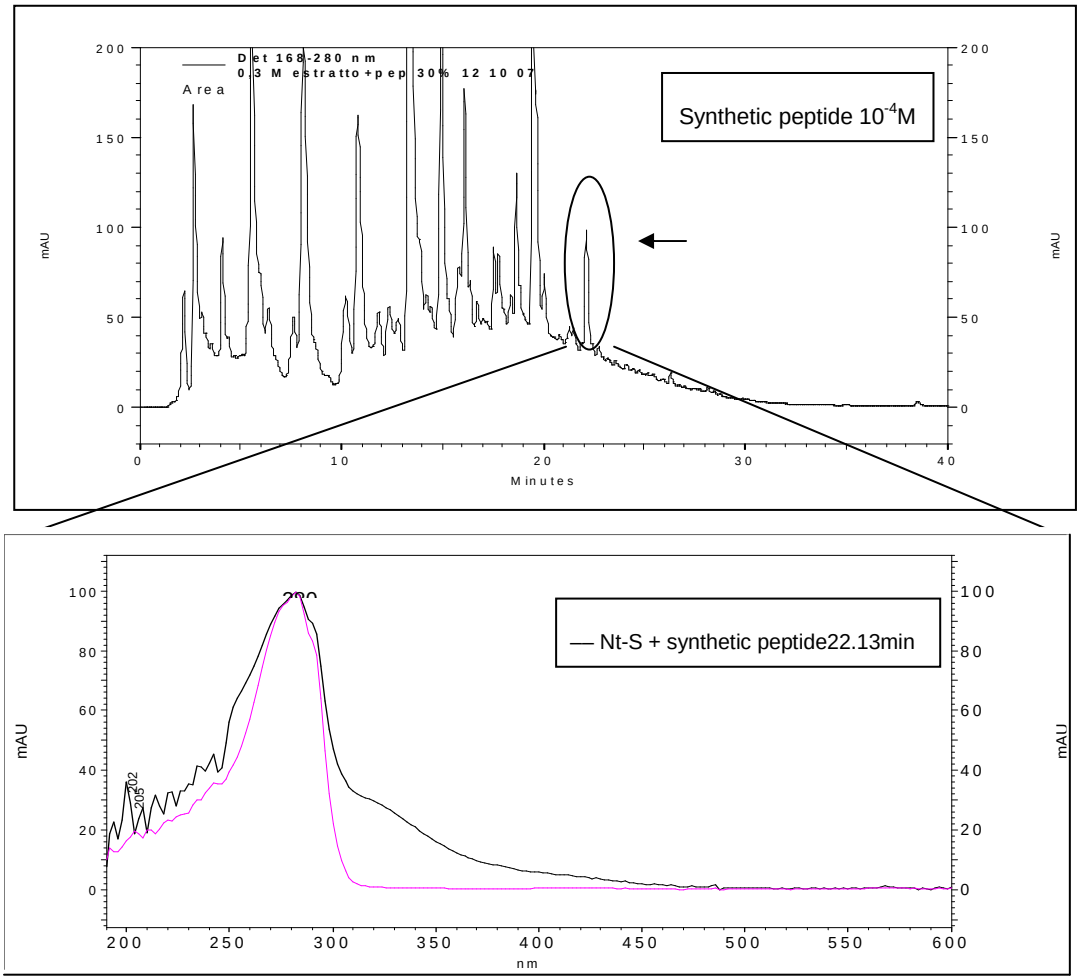


Fig 2.2 Upper panel: HPLC-DAD chromatogram. The marked peak corresponds to the synthetic peptide added to Nt-S cell extract before the beginning of the purification procedure (cut-off, DSC-NH_2 and SPE-C18 cartridge). After the SPE-C18 cartridge, the fraction in which the synthetic peptide eluted was evaporated to dryness, resuspended in sterile water and loaded on C18 reverse-phase HPLC column. The solvent system used was (A): 0.5% (v/v) formic acid and 5% (v/v) acetonitrile in water, and (B): 100% acetonitrile. The gradient profile was 0-35 min linear 50% solvent B and the flow rate was 200 µl/min. The synthetic peptide eluted after 22 minutes and its maximum absorbance peak was 280 nm. Lower panel: the absorbance spectra of the synthetic peptide (pink line) and the synthetic peptide added to Nt-S extract (black line) are reported.

Unfortunately, the described procedure allowed us to detect the synthetic peptide additioned to cell extract only if the initial concentration was 10^{-4} M. The synthetic peptide 10^{-5} M and 10^{-6} M was not detected by HPLC-DAD after the described procedure.

.

LC-ESI-MS analysis

Since HPLC-DAD does not have a very high sensitivity, the extract additioned with ENOD40 synthetic peptide was analysed by HPLC-Reverse-phase liquid chromatography coupled to ESI mass spectrometry (ESI-LC-MS) after the purification procedure. Also in this case the peptide was not detected. In order to understand if the failure in the peptide detection depended on matrix effect-ion suppression effects (suppression of m/z signal due to the specific matrix or to co-eluting ion species), the cell extracts was purified as described above, and the synthetic peptide 10^{-6} M was added just before the LC-MS-analysis. The peptide was well detected and fragmented also if the peptide was mixed to the processed cell extract, indicating that the problem in the peptide detection was due to the insufficient recovery during the purification procedure (not shown). This obviously prevented any peptide detection when the Nt-S cell extract alone was processed and analysed.

MALDI analysis

To overcome the limit of LC-ESI-MS for the expected low concentration of the putative ENOD40 peptide, a MALDI-TOF mass spectrometer was employed (in collaboration with Dr. Pietro Traldi, CNR-ISTM, Padova) to detect the peptide in Nt-S extracts.

The synthetic peptide at a final concentration of 10^{-6}M was used to set up the analysis and each step of the purification procedure was tested and optimized for MALDI analysis. Polyvinylpolypyrrolidone (PVPP), used during the extraction phase as described in "Material and Methods" section, was not used for the purification of the samples analysed by MALDI, because in preliminary experiments PVPP caused a high background noise level that hid the signals and prevented the detection of the synthetic peptide (data not shown). Therefore, zip-tip purification was also included before MALDI analysis.

In order to detect the putative peptide in Nt-S extracts, the mass spectra of the synthetic peptide and Nt-S purified fraction were compared. The ENOD40 peptide molecular ion, *m/z* =1173, was not revealed in Nt-S spectra, however two signals at *m/z* = 1155 and *m/z* = 1102 typical in the synthetic ENOD40 peptide spectrum, were detected also in the purified cell extract spectrum (Fig.2.3).

.

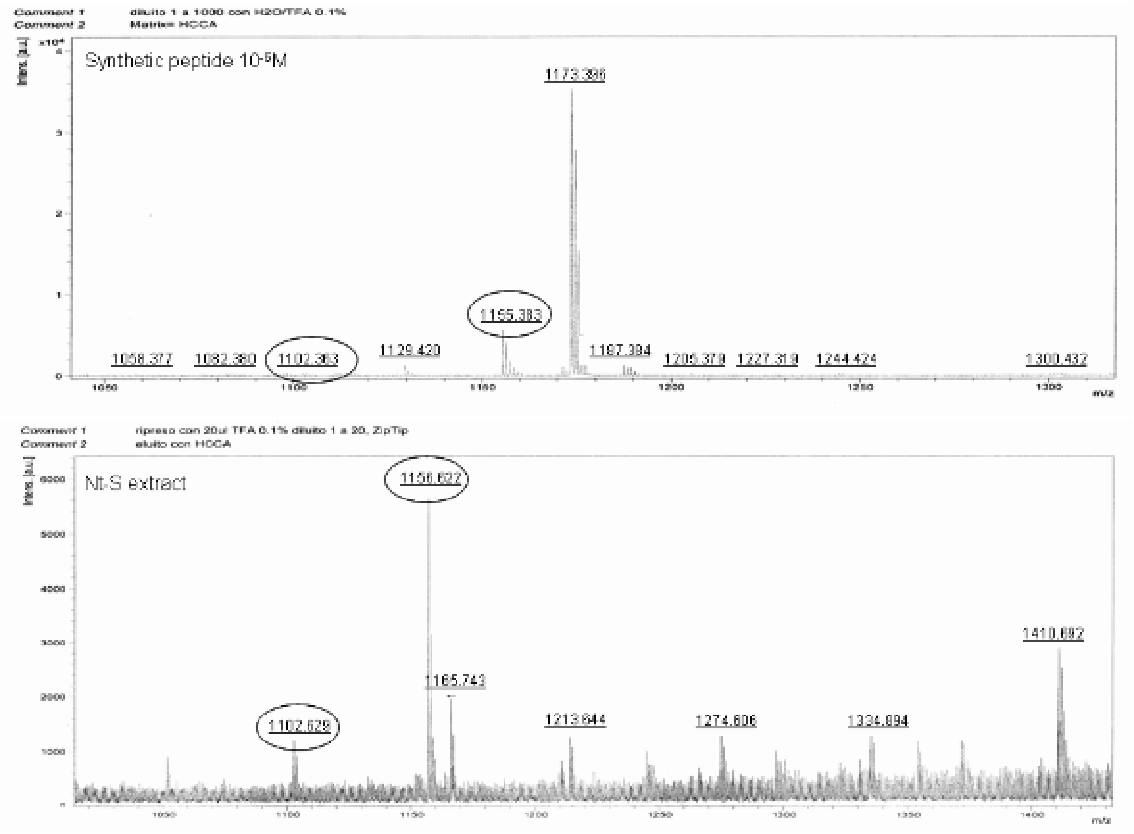


Fig.2.3 Synthetic peptide (10^{-6}M) and Nt-S purified fraction analysed by MALDI. In the black rings the two signals at m/z=1155/6 and m/z = 1102 identified in both the spectra.

Therefore, for further elucidation, MALDI-TOF MS was extended to post source decay (PSD) fragmentation. This technique allows the selection of a precursor molecule in a specific mass window and the subsequent analysis of its fragments. The signal at *m/z* =1156 was the precursor molecule selected. The intensities of fragment signals were very low in the PSD spectrum. However, three of the signals present in the PSD spectra and typical of the synthetic peptide were observed with very low intensity also in Nt-S extract, at *m/z* =1010, *m/z* = 741 and *m/z* = 632. (Fig. 2.4), suggesting the presence of ENOD40 peptide in the cell extracts.

.

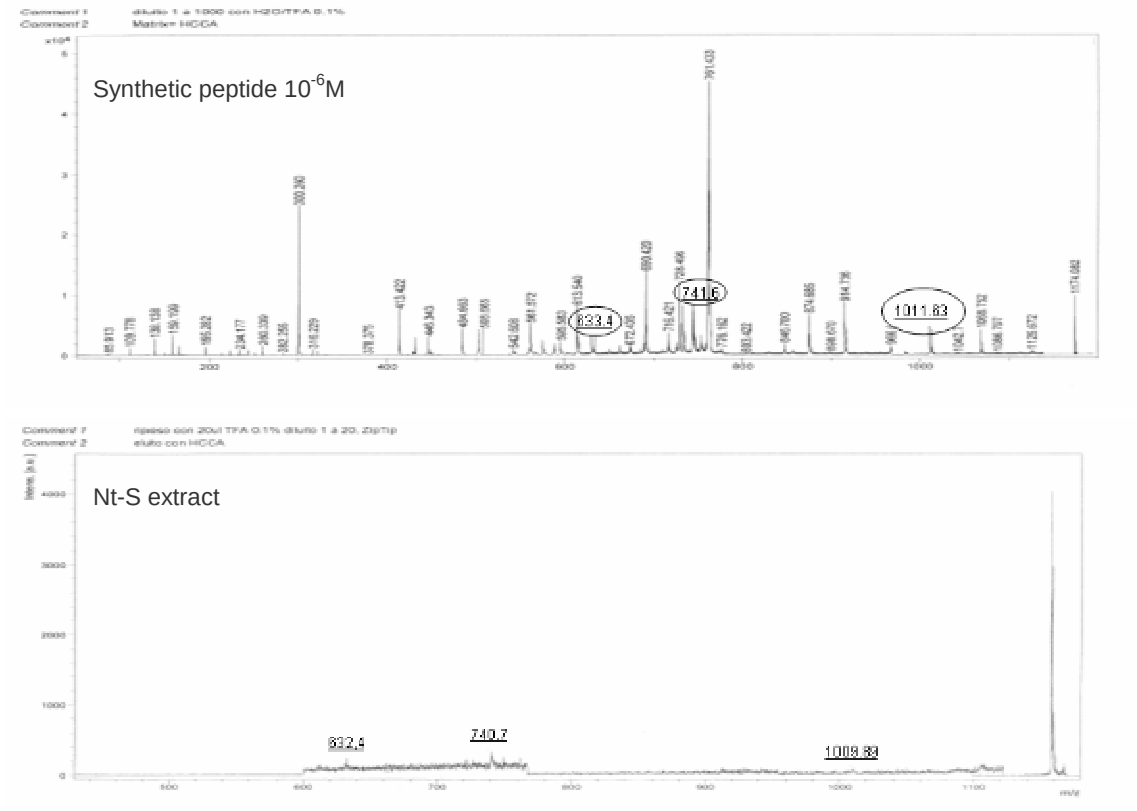


Fig. 2.4 PSD mass spectra of the synthetic peptide and of Nt-S extract.

DISCUSSION

A common feature of *ENOD40* genes is the absence of a long open reading frame and the presence of two highly conserved regions, box 1 and box 2. Only the ORF within the box 1 is conserved among all species: it seems to encode a putative peptide of 10-13 amino acids that seems responsible for *ENOD40* activity. However, the putative peptide has not so far been detected *in vivo* and it has been supposed that it could be present in a very low amount in the plant cell (Sousa *et al*., 2001).

In this work, we try to set up a purification procedure, to detect ENOD40 putative peptide in ENOD40 overexpressing BY-2 cells. The tobacco synthetic peptide (MQWDEAIHGS) was used to set up the analytical procedure, while the the *NtENOD40* overexpressing line Nt-S additioned with the synthetic peptide were used to set up the following purification procedure. The recovery of the well detectable synthetic peptide guided the purification procedure set-up.

A four-step purification procedure was set up, and using this procedure, it was possible to detect the synthetic ENOD40 peptide 10^{-6} M added to the cell extract with the very

.

sensitive MALDI-TOF technique.

However, despite several efforts spent to set up the purification procedure and the different and sensitive techniques used for the analysis, only with MALDI-TOF PSD analysis, three very weak signals detected in Nt-S spectra could possibly be ascribed to the synthetic peptide. This could represent an initial clue of the presence of the peptide in ENOD40 overexpressing BY2 cells, although purification and analytical techniques should be further improved for an unequivocal detection. It should be noted that the overexpression of the gene increases the amount of the transcripts, but not necessarily the amount of the peptide, since the accumulation of the peptide could be prevented by a high degradation rate.

Previous experiments performed on nodules (Sousa *et al*., 2001) demonstrated that immunological approaches were not adequate to detect the putative peptide biochemically since neither signals of synthetic peptide were detected in some experiments. However, western blot analysis performed on synthetic peptide added to nodule extracts showed that the peptide could bind to a subcellular structure or a protein larger than 10 kDa. This could mask the peptide, preventing both an efficient immunological detection as well as the purification and the detection by mass spectrometry. Moreover, *in vitro* translation analysis showed that the peptide could be rapidly degraded (Sousa *et al*., 2001).

.

REFERENCES

Compaan B, Yang, W.-C., Bisseling, T., and Franssen, H. (2001) *ENOD40* expression in the pericycle precedes cortical cell division in Rhizobium-legume interaction and the highly conserved internal region of the gene does not encode a peptide. Plant Soil 230: 1–8

Girard G, Roussis A, Gultyaev AP, Pleij CW, Spaink HP (2003) Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean. Nucleic Acids Res 31: 5003-5015

Gultyaev AP, Roussis A (2007) Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res 35: 3144-3152

Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by enod40 in Arabidopsis thaliana. J Exp Bot 56: 507-513

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408

Podkowinski J, Zmienko A, Florek B, Wojciechowski P, Rybarczyk A, Wrzesinski J, Ciesiolka J, Blazewicz J, Kondorosi A, Crespi M, Legocki A (2009) Translational and structural analysis of the shortest legume *ENOD40* gene in Lupinus luteus. Acta Biochim Pol 56: 89-102

Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M (2002) Soybean *ENOD40* encodes two peptides that bind to sucrose synthase. Proc Natl Acad Sci U S A 99: 1915-1920

Ruttink T (2003) *ENOD40* affects phytohormone cross-talk. Wageningen University and Research Center, the Netherlands

Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M (2001) Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Mol Cell Biol 21: 354-366

Vijn I, Yang WC, Pallisgard N, Ostergaard Jensen E, van Kammen A, Bisseling T (1995) VsENOD5, VsENOD12 and VsENOD40 expression during Rhizobium-induced nodule formation on Vicia sativa roots. Plant Mol Biol 28: 1111-1119

Wan X (2007) Analysis of Nodule Meristem Persistence and ENOD40 functioning in *Medicago truncatula* nodule formation. Wageningen.

.

Chapter 3

Metabolic profiling of *Arabidopsis thaliana* plants overexpressing *ENOD40*

.

.

ABSTRACT

ENOD40 gene plays a key role in nodule formation, but the presence of *ENOD40* homologues in non-legume plants suggests that this gene is involved in processes common to all plants. However, up to now little is known about its function in non-legumes. *Arabidopsis thaliana* plants overexpressing soybean *ENOD40* were used to investigate *ENOD40* role.

In this work, metabolite profiles of wild type and three different transgenic lines of *A. thaliana* plants were analysed by LC-ESI-MS followed by multivariate analysis techniques. Biomarkers characterizing transformed and wild type plants were identified.

INTRODUCTION

ENOD40 is one of the Early Noduline genes that is known to play a key role in nodule formation in response to interaction of legume plants with symbiotic *Rhizobium* bacteria. Homologues of *ENOD40* gene have been identified in monocots like maize, rice and sorghum, and dicots such as tomato, tobacco, citrus and several leguminous species.

The highest expression levels of *ENOD40* genes are associated with the formation and development of radical nodules (Asad, 1994; Crespi *et al.*, 1994; Wan *et al.*, 2007). However, low *ENOD40* expression has also been localized in tissues not related to symbiotic interactions, but associated with developmental processes, such as formation of lateral roots and shoots (Papadopoulou *et al.*, 1996; Corich *et al.*, 1998). Similar expression pattern has also been found in non-legume plants (Kouchi *et al.*, 1999) in which *ENOD40* homologues have been identified.

The expression of *ENOD40* in non-symbiotic tissues and its presence in non-legume plants suggest that the gene is involved in many processes common to all plants. Although extensive investigation have been done in the past, the precise function of *ENOD40* remains unknown.

In *A. thaliana* plants *ENOD40* overexpression induces a morphological cell phenotype, since plants transformed with *ENOD40* and protoplasts transiently expressing *ENOD40* showed a reduced cell size (Guzzo *et al.*, 2005).

In the present research the effect of the overexpression of *ENOD40* on *A. thaliana* metabolome was investigated.

.

Chemical characterization of the phenotype of an organism can be done at the level of macromolecules or low molecular weight compounds. The analysis of the latter molecules offers the most direct measurement of a cell's physiological activity and of changes after experimental treatments (Beecher, 2002). Therefore metabolomics analysis, defined as the qualitative and quantitative survey of all the metabolites of a cell, tissue, organ or organism (Verpoorte, 2008), is a powerful tool to define the phenotype of an organism from the chemical point of view.

The metabolome consists of two types of compounds, primary metabolites and secondary metabolites. Primary metabolites are compounds involved in the basic functions of the living cell, such as respiration and biosynthesis of the amino acids and other compounds needed for a living cell. Secondary metabolites are species specific compounds that play a role in the interaction of a cell with its environment, i.e. other cells in the organism, external organisms or abiotic factors (Verpoorte, 2007).

The different chemical properties of cell metabolites make the analysis of all metabolites in a cell, tissue, organ or organism in reality impracticable with the technology available so far, so that, usually, the analysis is focused on sub-metabolome fractions (for istance, the sub-metabolome extractable with a certain solvent) (Ceoldo *et al.*, 2009). Among the technologies available for analyzing a metabolome, liquid chromatography (LC) coupled to mass spectrometry (MS) offers higher sensitivity and selectivity and hence an increased identification of metabolites.

In this chapter, an untargeted metabolomics approach followed by bioinformatics and statistical analysis was applied to wild type and to three different lines of *A. thaliana*, overexpressing the *ENOD40* gene.

The high number of variables generated in the context of metabolomics requires bioinformatics and statistical tools for extracting and analysing information from the results. Here LC-MS dataset was preprocessed with MZmine software, and then subjected to principal component analysis (PCA) and an Orthogonal Projections to Latent Structures-Discriminant Analysis (O2PLS-DA).

Furthermore, biomarkers characterizing transformed and wild type plants were identified by two-class O2PLS-DA analysis.

.

MATERIALS AND METHODS

Plant Material

Three different lines, 1.1, 3.1 and 4.4, produced by three independent transformation events, and wild type plants of *A. thaliana* ecotype Wassilewskjia, were used in this work.

Arabidopsis thaliana plants transformed with the first 448 bp of soybean *ENOD40* driven by 35S were kindly supplied by Henk Franssen (Molecular Biology Laboratory, Wageningen University). The insert contained both box 1 and box 2 regions.

The three homozygous transformed lines were selected afterwards the segregation analysis of F1, F2 and F3 generations. Each line contains one copy of the gene.

Arabidopsis thaliana plants were grown in short day photoperiod (10 h light/ 14 h dark) with 70% of humidity and a day/night temperature of 24°C and 22°C respectively. The light intensity was 100-120 μmol quanta m^2/s.

Six weeks after germination, all the rosette leaves were collected, grinded with liquid nitrogen in a fine powder and lyophilized for metabolomic and transcriptomic analysis. Three biological replicates were analysed, each one composed by a pool of rosette leaves from eight plants.

Primer Design

All the primers were designed using Primer 3 (http://frodo.wi.mit.edu/).

Primer Name	Primer Sequence (5'-3')
Actin1 forward	CCGAGCGTGGTTACTCTTTC
Actin1 reverse	GAGCTGGTTTTGGCTGTCTC
GmENOD40 forward	GGGTGCTCACTCCTCACACT
GmENOD40 reverse	TCCGCCACTCAAGAAAGAAT

.

RNA Extraction, DNAse Treatment and cDNA synthesis

RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich). After RNA isolation the samples were treated with RQ1 RNase-Free DNase (Promega); 3 μg of RNA were treated with 3 μl of RQ1 DNase in total volume of 30 μl to remove any contaminating DNA. The cDNA was synthesized using the Improm-II™ Reverse Transcription System kit (Promega); 1μg of RNA treated with DNase was reverse-transcribed, whereas the RNA not reverse-transcribed was used as negative control in subsequent Real-Time PCR experiments.

Real-Time PCR

Real-time PCR was used to confirm the expression of the *GmENOD40*.

PCR reaction mix contained: 5 μl of cDNA diluted 1:10, 12.5 μl of Platinum ® SYBR ® Green qPCR SuperMix UDG with ROX (Invitrogen), 0.5 μl of each primer (stock 20 μM), and sterile double-distilled water to a final volume of 25 μl.

Real-time PCR amplification was performed on an ABI PRISM 7000 sequence detection system (Applied Biosystems) programmed for initial denaturing at 95°C for 10 min, and 40 cycles of denaturing at 95°C for 15 sec, annealing at 55°C for 30 sec and extension at 72°C for 30 sec. The relative expression was calculated according to $2^{-\Delta\Delta Ct}$ method as described by Livak and Schmittgen (2001).

Chemicals

All chemicals and water LC-MS grade were supplied by Sigma-Aldrich. Standard for LC-MS were from Lab-Service (Extrasynthese).

Secondary metabolite extraction

For LC-ESI-MS, fresh Arabidopsis leaves were grinded with liquid nitrogen in a fine powder and then lyophilized. To extract the metabolites, 100 mg of lyophilized material were resuspended in 5 ml of methanol/water (9:1;v/v), sonicated for 20 minutes in an ultrasonic bath (Falc Ultrasonic Bath) at room temperature at maximum frequency (40 kHz) and centrifuged at 1200 x *g* for 10 minutes at 4° C (Eppendorf 5804 R centrifuge, rotor A-4-4) to eliminate cell debris. Cell extracts were stored at -20 °C, and

.

the supernatant was filtered just prior the LC-MS injection through Minisart RC 4 0,2 µm filters (Sartorius Stedim Biotech).

To test the glucosinolates content, wild type leaves of different age, young and old (fig 3.1), were collected. The metabolites were extracted from 50 mg lyophilized material with 2.5 ml of methanol/water (9:1;v/v). The remaining steps of extraction were those described in the previous paragraph.

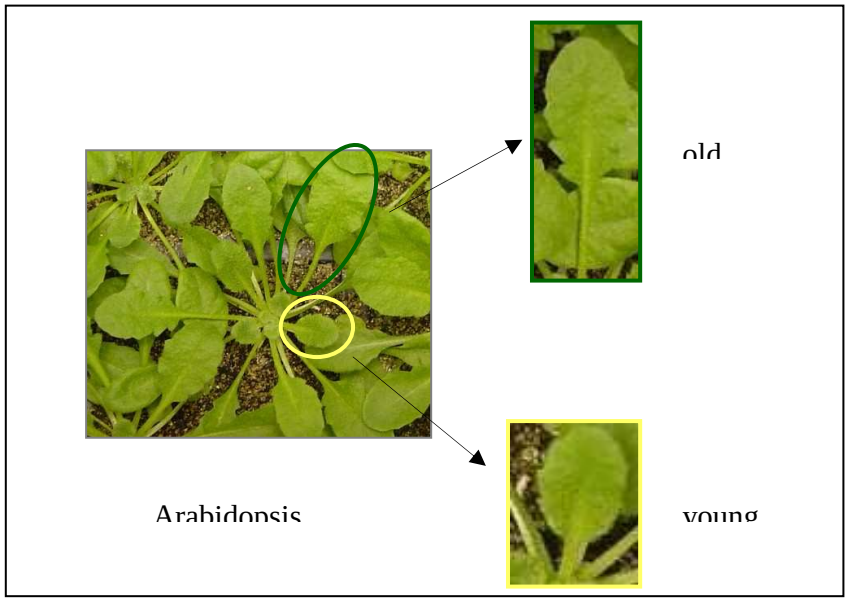


Fig. 3.1 *A. thaliana* leaves were collected to test glucosinolates concentration. The figure show an example of young and old leaves selected for the analysis.

LC-ESI-MS analysis

Metabolomic analysis was performed by LC-MS, using an HPLC system (Beckman Coulter System Gold 127, Solvent Module) coupled on-line with the Bruker ion trap mass spectrometer Esquire 6000, equipped with an ESI source. The data were collected using the Esquire Control 5.2 software, and processed using the Data Analysis 3.2 software (Bruker Daltonics Esquire 5.2, Bruker Daltonik GmbH).

The extracts were diluted 1:10 with water before chromatographic analysis. Each biological replicate was analyzed twice. Solvents were (A): 0.5% (v/v) formic acid and

.

5% (v/v) acetonitrile in water, and (B): 100% acetonitrile. Gradient profile: 0–5 min: linear 10% B; 5-20 min: linear 10-20 % B; 25-30 min: linear 20-25 % B; 30-45 min: linear 25-70 %; the flow rate was 200 µl/min. A 15 min equilibration time followed each sample analysis. The column used for the analysis was a 150 mm x 2.1 mm Alltima HP C18 3 µm, with a guard column 7.5 mm x 2.1 mm (Alltech Associates, Inc., Derfield, IL). The injection volume was 20 µl.

Negative and positive ion mass spectra were recorded in the range of 50-1500 m/z (full scan mode, 13,000 m/z s^{-1}). Nitrogen was used as nebulizing gas (pressure: 50 psi, temperature 350°C) and drying gas (12 l min^{-1}). Helium was used as collision gas.

For tandem mass spectrometry analysis, ms/ms and ms3 spectra both in negative and positive mode were recorded, in the range of 50–1500 m/z, with a fragmentation amplitude of 1 V. The other mass spectrometer parameters were set up as follows.

Source: capillary +4000 V, end plate offset -500 V; skimmer: -40 V; cap exit -121 V; Oct.1DC: -12 V; Oct.2DC: -1.70 V; lens 1: 5 V; lens 2: 60 V.

The putative identification of the metabolites was performed through comprehensive analysis of their m/z and fragmentation pattern (tandem mass spectrometry) as well as through the comparison of their fragmentation pattern, absorbance spectra and retention times with those reported in literature and with those obtained by commercial standards, when available.

Data Analysis

LC-MS data were transformed in net.cdf format with the Bruker Daltonics Esquire 5.2-Data analysis 3.2 software and processed by MZmine 1.94 software, (http://mzmine.sourceforge.net). This software performed: noise filtering, centroid detection and alignment of the signals (= centroids).

The LC-MS data matrix was imported in SIMCA-P software (v. 12.0, Umetrics). Principal Component Analysis (PCA) and Bidirectional Orthogonal Projections to Latent Structures Discriminant Analysis were performed on the whole data matrix (signals obtained with positive and negative mode) using Unit Variance scaling. PCA was used to get an overview of the metabolite profiles.

O2PLS-DA (Discriminant Analysis), which is an extension of O2PLS analysis (Trygg

.

et al., 2007), was used in order to obtain a clearer and more straightforward interpretation of the data compared to PCA (Bylesjo *et al.*, 2007).

PCA and O2PLS-DA were performed on twenty-four observations: three biological replicates for each line (each constituted by a pool of 8 plants) and two technical replicates for each sample. Furthermore, SUS-plot was used to improve the visualization and interpretation of O2PLS-DA models and allow the identification of biomarkers. The SUS-plot combines the correlation profiles, p(corr), from two models. The variables considered as biomarkers have $p(corr) \geq \pm 0.7$.

Cross-validation with CV-Anova ($p < 0.05$ significant) and a permutation test (200 permutations) were performed in order to estimate the strength of the models and ensure that the models were not random and not over-fitted.

.

RESULTS

Analysis of *A. thaliana* transformed lines for the expression of *GmENOD40*

The expression of the transgene was confirmed by Real-Time PCR using *GmENOD40* specific primers. RNA was extracted from the three different lines, 1.1, 3.1 and 4.4, produced by three independent transformation events, and from wild type plants. Extracted RNA was reverse-transcribed and cDNA was employed as template for Real-Time PCR. Wild-type sample was used as negative control and no soybean *ENOD40* expression was detected, as expected (Fig.3.2).

Actin gene, which is supposed to be equally expressed in all samples, was used as a reference to normalize gene expression.

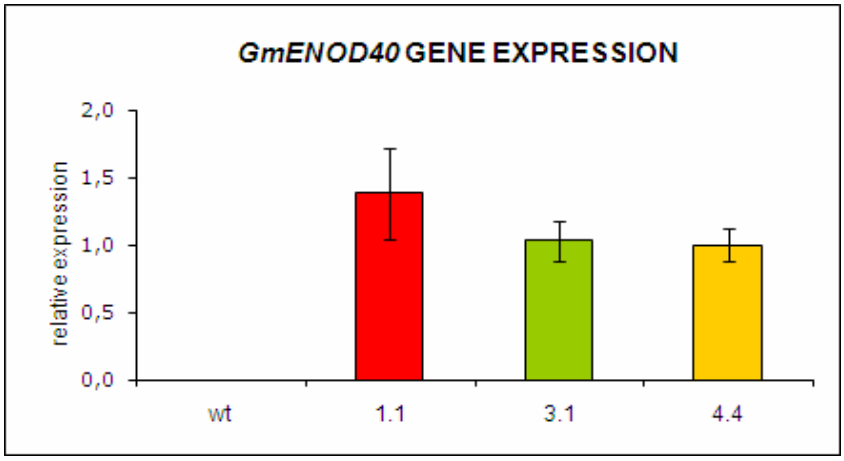


Fig.3.2 Real-Time RT-PCR for the expression o f *GmENOD40* in *Arabidopsis thaliana*. Bars represents mean ± SE for RNA samples extracted from a pool of rosette leaves from eight plants. 4.4 line was used as the calibrator.

LC-MS metabolite profiling of *A. thaliana* methanol extract

Untargeted metabolomics analysis of the methanolic extracts was performed for each line by LC-MS to verify whether the overexpression of *GmENOD40* gene has effects on plant metabolism. Three biological and two technical replicates for each line were analyzed by LC-MS in negative and positive ionization mode and the obtained data were processed by MZmine software, as described in “Material and Methods” section. LC-MS analysis revealed 1360 signals, 1071 detected in negative mode and 289 in

.

positive ionization mode. Among all the signals detected, 98 molecules were putatively identified (tables 3.1 and 3.2). These included sugars, glucosinolates, flavonoids, sinapic acid derivatives, other hydroxycinnamic acid derivatives, hydroxybenzoic acid derivatives, non aromatic organic acids, aminoacids. Between the glucosinolates, only indolic and aliphatic glucosinolates were putatively identified, while aromatic glucosinolates were not evident.

The more abundant flavonoid was the tetrahydroxyflavonol kaempferol, with various glycosylation patterns. The sinapoyl malate, that is an ester of sinapic and malic acid, was the most represented molecule in the methanol extractable metabolome under analysis. Also many sinapoyl glucosides and the caffeoyl quinic acid (chlorogenic acid) were evident. Between the non aromatic acids, citric acid and malic acid were the more represented.

Most glucosinolates, flavonoids and sinapoyl derivatives were identified in negative mode, while in positive ionization mode mainly fragments of glucosinolates and some amino acids were detected.

The identification of glucosinolates, typical molecules of Arabidopsis, was facilitated in negative ionization mode by the reproducible fragmentation of the molecules to a sulphated glucose anion (m/z 259) (Rochfort *et al*., 2008).

All the identified molecules are reported in tables 3.1 and 3.2.

Analysis of metabolomic data by chemometric approach

In a metabolomics context, bioinformatical and statistical tools are required to complete the global interpretation of the results; for this reason the use of chemometrics tools, e.g., principal component analysis (PCA), and Bidirectional Orthogonal Projections to Latent Structures Discriminant Analysis (O2PLS-DA), are of great importance as they include efficient, validated, and robust methods for modeling and interpretation of complex chemical and biological data. In order to compare wild type and transformed plants, LC-MS dataset was processed by MZmine software and the obtained data matrix (data obtained with positive and negative ionization modes, together) was imported in SIMCA-P software v.12.0 for the analysis.

Unsupervised PCA analysis (UV scaling), performed on the whole dataset shows that samples are clustered into four groups, and the three transformed lines are clearly

.

separated from the control, as reported in score plot (Fig.3.3).

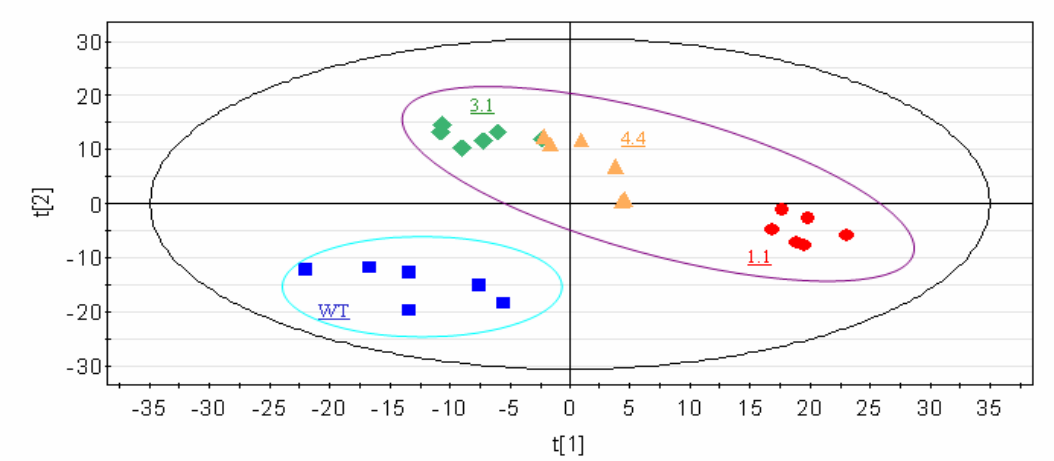


Fig.3.3 PCA score plot (UV scaling) of metabolite profiling of wild type (blue square) and the three *GmENOD40* overexpressing lines (1.1: red circle, 3.1: green diamond, 4.4: yellow triangle) using the LC-MS quantified molecules as x variables. In this plot each point represents one observation; all the three biological and the two technical replicates are reported. The intensities of the 1360 signals quantified both in negative and in positive ionization mode by LC-MS after MZmine analysis are used as variables.

The separation is even more strong when the supervised O2PLS-DA analysis is used (Fig.3.4).

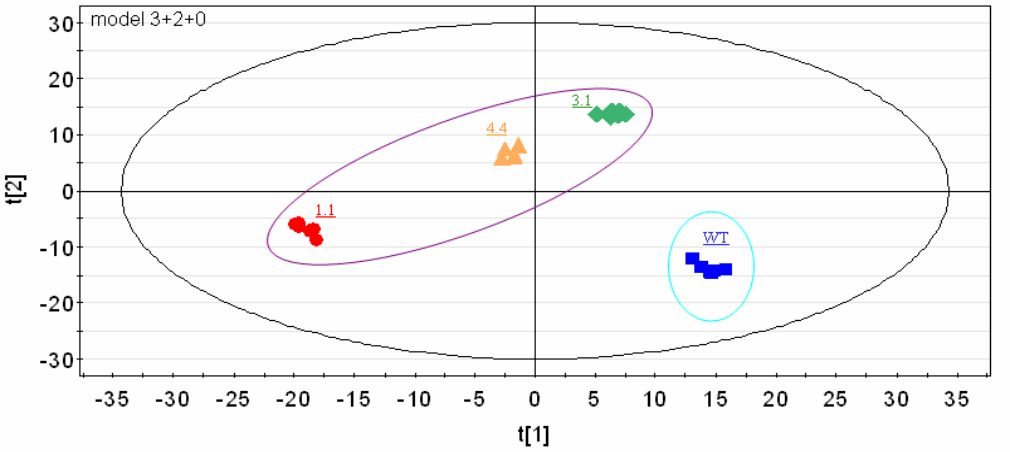


Fig.3.4 O2PLS-DA score plot (UV scaling) of wild type sample (blue square) and each of the three *GmENOD40* overexpressing lines (1.1: red circle, 3.1: green diamond, 4.4: yellow triangle) using as x variables LC-MS quantified signals.

Furthermore, the PCA and O2PLS-DA score plots show similarity in the metabolite profiles of the transformed lines 3.1 and 4.4.

In order to show which molecules cause the separation between the various cell lines, a pq(corr)1/pq(corr)2 loading plot was used, in which the relationships between

.

molecules and groups of samples are shown (fig. 3.5), and metabolites are coloured according to their chemical class.

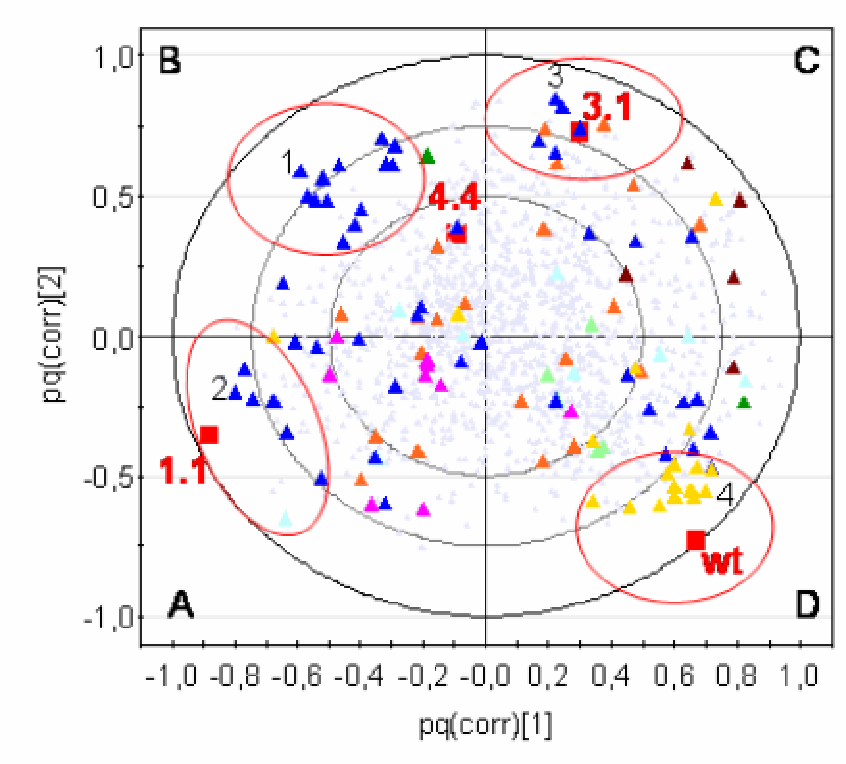


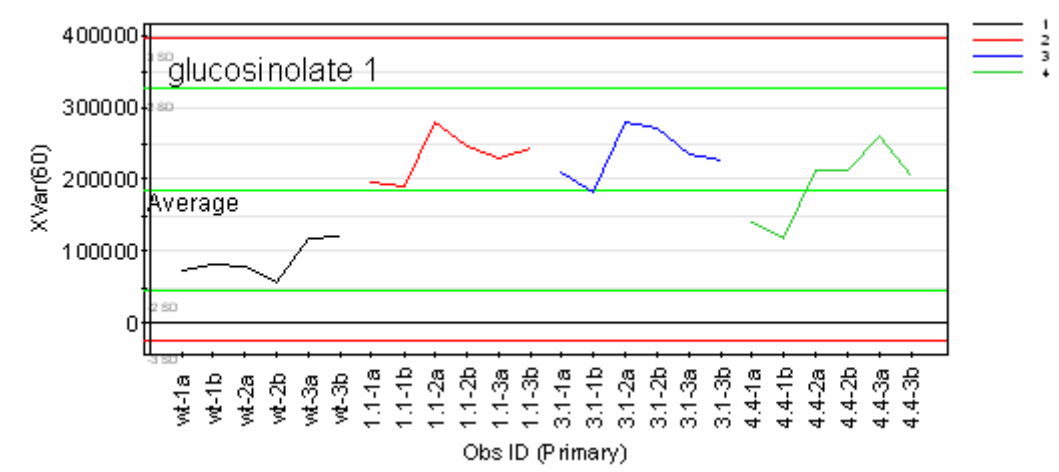
Fig. 3.5 pq(corr)1/pq(corr)2 loading plot. Relationships between classes of samples (red squares) and metabolites (coloured triangles) are shown. Blue: glucosinolates; pale blue: sugars; green: hydroxybenzoic acid derivatives; orange: synapic acid derivatives; pale green: other organic acids; fucsia: aminoacid; yellow: flavonoids; brown: other hydroxycinnamic acid derivatives; grey: unknown.

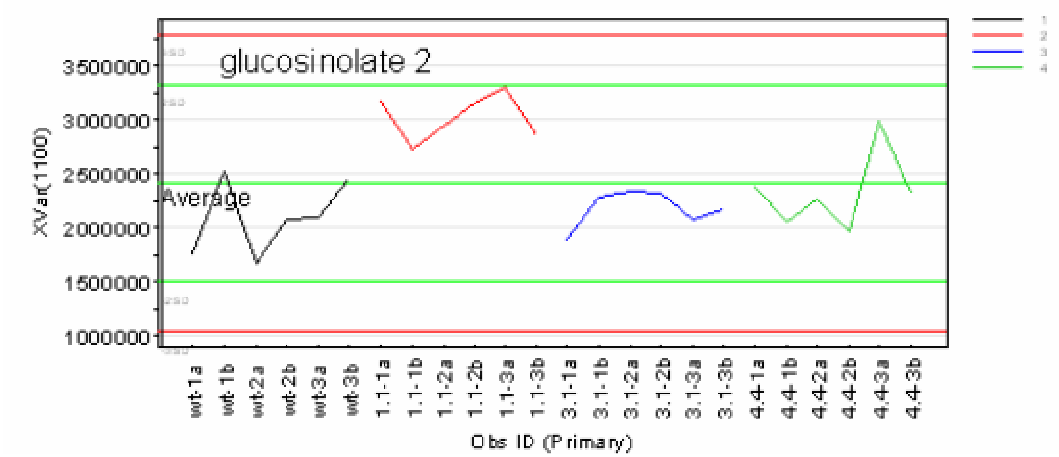
According to the model shown in Fig.3.5, the first component P1, which explains 12% of the total variability, separates 1.1 and 4.4 transformed lines from wild type and 3.1, while the second component P2, that explains 9% of the total variability, separates 4.4 and 3.1 lines from wt and 1.1 samples. Metabolites in quadrants A,B,C with high absolute pq(corr) value (for instance those inside the red circles) significantly characterized one or more transformed lines whereas metabolites in the quadrant D and with high absolute pq(corr) value significantly characterized the wild type samples. The inner metabolites, with low absolute pq(corr) values do not vary between samples.

A consistent number of glucosinolates (blue triangles) characterize the transformed lines (A,B,C quadrants), while the flavonoids (yellow triangle) mainly characterized the wild type samples (D quadrant). The amino acids (fucsia triangles) characterize the 1.1 line (quadrant A), but with low significance (low pq(corr) value), whereas the very abundant sinapic acid derivatives do not characterize any of the groups of samples.

.

To better clarify the model, in Fig. 3.6 four metabolites with different position in the loading plot (1, 2, 3, 4 in Fig. 3.5) are analysed for their signal intensities in the different classes. Glucosinolate 1 is correlated with all transformed lines, glucosinolate 2 correlates only with line 1.1, glucosinolate 3 with lines 3.3 and 4.1, and flavonoid 4 with wild type.





.

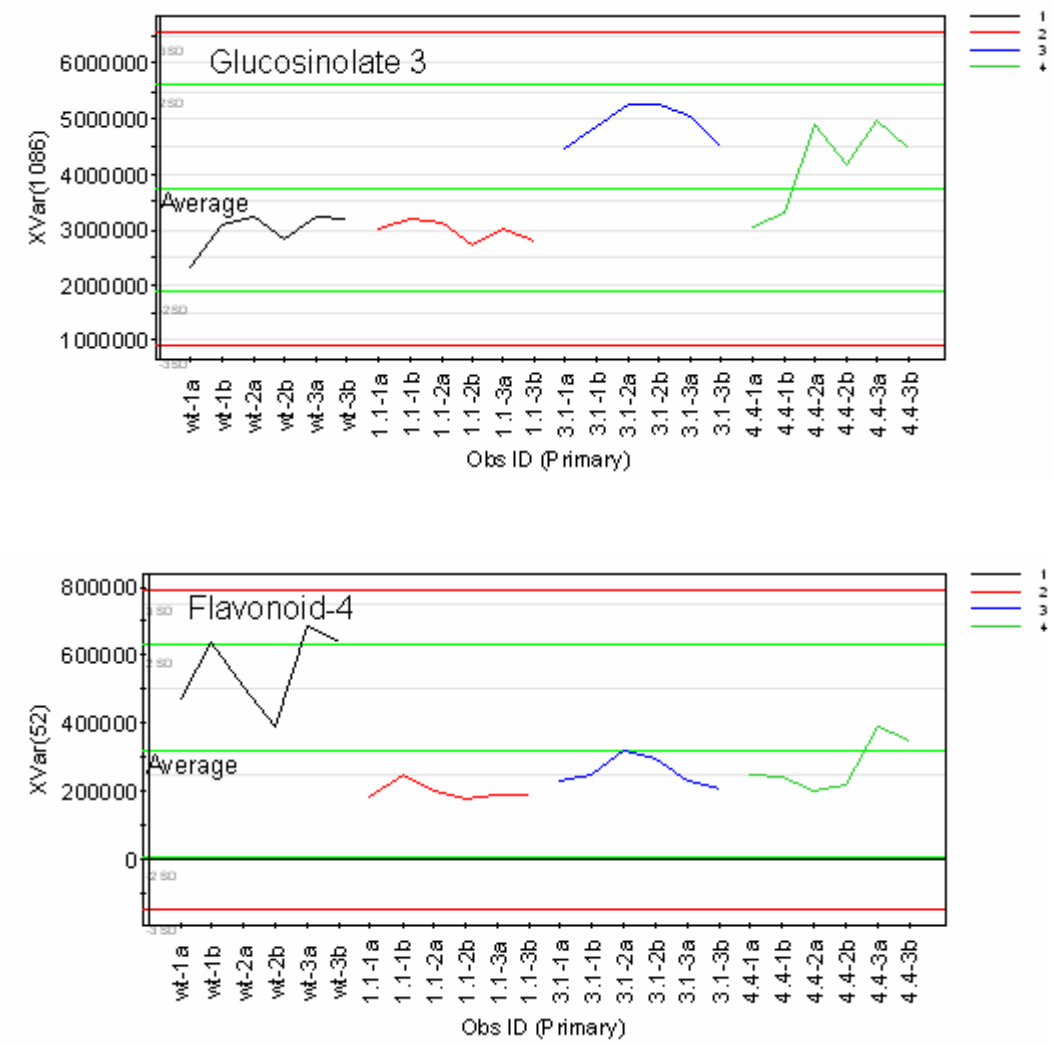


Fig. 3.6 Quantification of the metabolites 1,2,3, which are glucosinolates, and 4 (a flavonoid) in the various samples, according to the dataset used for the analysis, is reported.

In order to determine which variables can be considered as biomarkers for transformed lines, O2PLS-DA was performed on the whole dataset.

Biomarkers are defined as molecules indicating or correlating with a physiological change; this could include both positive and negative biomarkers; in this investigation the positive biomarkers characterize all the three transformed lines, while the negative biomarkers characterized the wild type class.

To identify the biomarkers, three distinct two-classes O2PLS-DA models were built for each transformed line compared with the wild-type plants (model M4:WT vs1.1, model M5: WT vs 3.1 and model M6: WT vs 4.4). These three models (M4, M5 and M6) were used to investigate the difference between the control (wild-type plants) and each

.

transformed line; subsequently, the transformed lines were compared pairwise by Shared and Unique Structure plot (SUS-plot), using the results of the previous models.

Therefore, SUS-plot was used to investigate the role that a measured variable has to discriminate all the transformed lines compared with wild type plants; moreover, it is able to show shared and unique structures among the samples. In figure 3.7 a) an example of SUS-plot between the model M4 (WT vs 1.1) and the model M5 (WT vs 3.1) is reported. Each metabolite has one coordinate from each model and is symbolized by a triangle. The green and the yellow zones show the regions in which the metabolites characterized only one transformed line compared with the control; in this case the variables close to the green region characterized only the line 3.1 and the metabolites close to yellow region differentiate the line 1.1 from wild type but not from the line 3.1. Instead, the metabolites that discriminate both the transformed lines,1.1 and 3.1, from wild type plants are close to the corners (red regions).

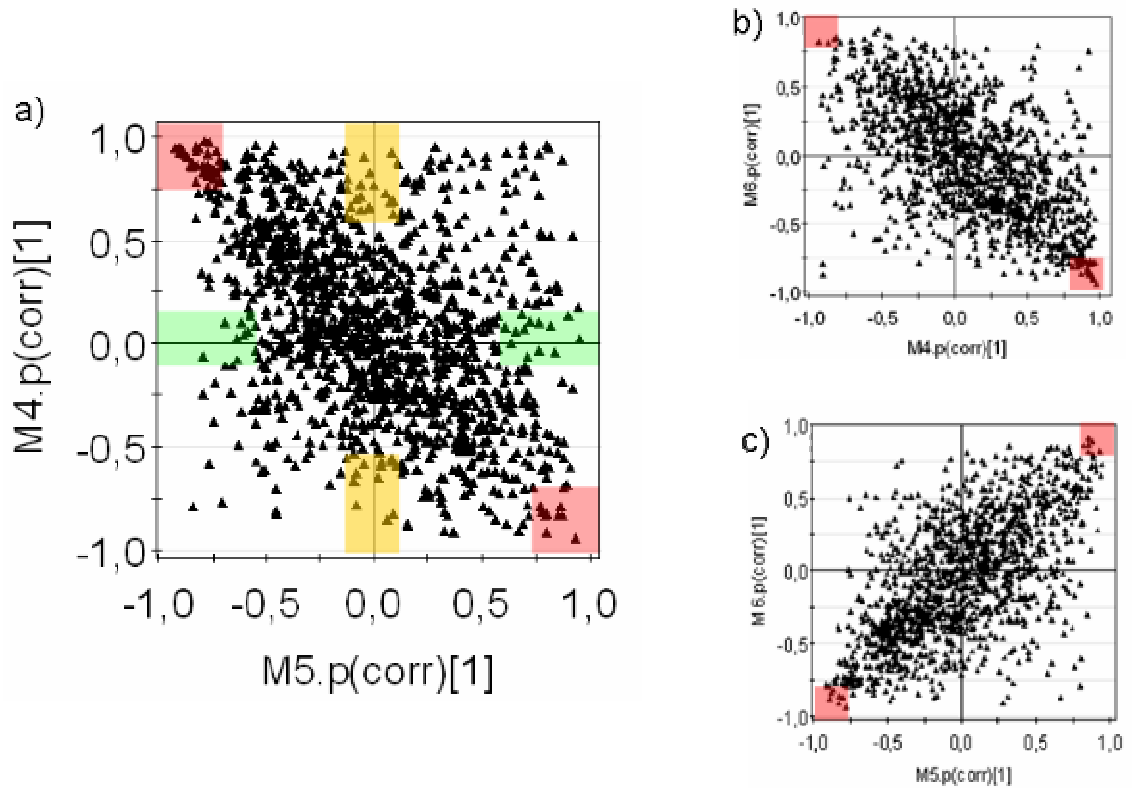


Fig.3.7 a) SUS-plot between the model M4 (WT vs 1.1) and M5 (WT vs 3.1).
b) SUS-plot between the model M4 (WT vs 1.1) and M6 (WT vs 4.4).
c) SUS-plot between the model M5 (WT vs 3.1) and M6 (WT vs 4.4).
The metabolites with coordinates close to (1,-1) and (-1,1) characterize two transformed lines from wild type plants (red regions).

.

By comparing the possible combinations of all the three models (Fig. 3.7), biomarkers specific for all the three transformed plants compared with the wild type line were identified (Tab 3.3). Six glucosinolates were identified as positive biomarkers, whereas five flavonoids and one hydroxybenzoyl hexoside were identified as negative biomarkers. Therefore, only glucosinolate metabolites characterize as positive biomarkers the transformed lines compared with wild type, whereas flavonoids, negative biomarkers of transformed lines, characterized wild type plants.

rt	m/z	positive biomarkers
4.939	406	3-methylthiopropyl (glucoiberverin)
4.954	166	3-methylthiopropyl fragment (glucoiberverin)
8.521	420	4-methylthiobutyl (glucoerucin)
26.804	462	putative glucosinolate
26.813	530	glucosinolate derivative
34.646	476	4-methoxy-3-indolylmethyl (4-methoxyglucobrassicin)

rt	m/z	negative biomarkers
20.585	755	kaempferol hexose hexose deoxyhexose
18.445	740	kaempferol hexose-deoxyhexose isotope
4.187	299	hydroxybenzoyl hexoside
25.217	577	kaempferol deoxyhexose deoxyhexose
25.212	640	kaempferol deoxyhexose deoxyhexose derivative
18.436	739	kaempferol hexose-deoxyhexose-deoxyhexose

Tab.3.3 Positive and negative biomarkers of *ENOD40* overexpressing lines. Six glucosinolate molecules (positive biomarkers) characterize all the transformed lines compared to wild type, whereas negative biomarkers, absent in the transformed lines, characterize wild type plant. Glucosinolates trivial name between bracket.

Moreover, the molecules that characterize each transformed line compared with the control were analyzed; in the line 1.1 and 3.1 the increase of other specific glucosinolates was observed.

4-hydroxy glucobrassicin increased in both lines, whereas glucohirsutin and neoglucobrassicin increased in line 1.1 and 3.1 respectively.

.

Glucosinolates in *A. thaliana*

Since in the literature a correlation between glucosinolate accumulation and developmental stages is reported (Petersen *et al.*, 2002; Brown *et al.*, 2003), the determination of glucosinolate in leaves with different ages was performed by LC-MS.
To test the glucosinolates concentration of leaves, a metabolic profiling was performed on young and old wild type leaves by LC-MS. The data were analysed by Mzmine and the total content of glucosinolates (sum of the centroid areas) was quantified (Fig 3.6).

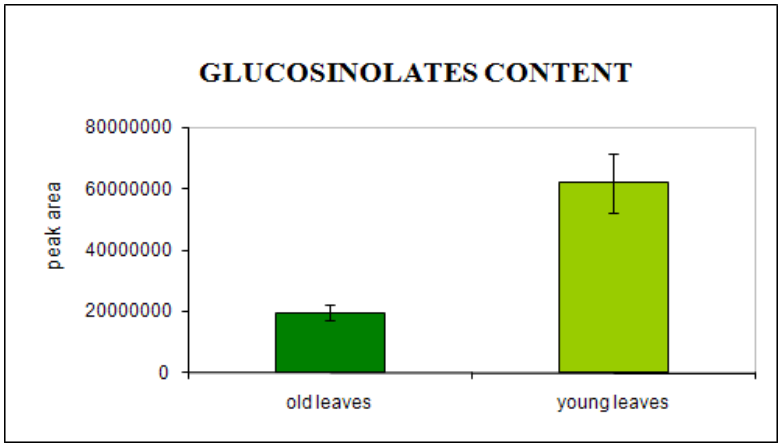


Fig 3.6 Quantitative analysis of glucosinolates in young and old leaves of *A.thaliana* wild type. Metabolic profiling was performed by LC-MS after MZmine analysis.

The glucosinolates concentration was much higher in young leaves than in old ones.

DISCUSSION

Up to now little is known about the precise function of *ENOD40* gene. *ENOD40* homologues have been found among legume and non-legume plants, suggesting an involvement of the gene in processes common to all plants. Here *A. thaliana* plants transformed with soybean *ENOD40* gene driven by 35S promoter were used to investigate the possible role of *ENOD40* in the non-legume plant *Arabidopsis thaliana*. Metabolic profiling of three independent transformed lines (1.1, 3.1 and 4.4) and wild type plants were compared to investigate *ENOD40* gene role.
The LC-ESI-MS analysis of methanolic extracts of wild type and transformed plants

.

allowed to detect 1360 signals from which 98 molecules (parent molecular ions) have been identified.

The metabolite profiling obtained by LC-MS analysis includes small primary metabolites such as sucrose, some amino acids, some organic acids and many secondary metabolites. Among these, glucosinolates, flavonoids and sinapic acid derivatives were mainly detected.

To look for possible correlation among the samples and between metabolites and samples, chemometric approach was used to analyse the data.

Unsupervised PCA, performed to investigate the variability in the dataset, and supervised O2PLS-DA, used to look for possible correlations between metabolites and cluster of samples, show in the score plot a clear separation between wild type and transformed plants, suggesting an effect of *GmENOD40* on plant metabolism.

The loading plot showed that this separation was mainly due to the glucosinolates, that characterized the transformed lines, and the flavonoids, that characterized the wild type line.

Furthermore, biomarkers were identified for wild type and transformed plants by two-class O2PLS-DA analysis. This analysis revealed some flavonoids as negative biomarkers and some glucosinolates as positive biomarkers for transformed plants.

Flavonoids are secondary metabolites widely distributed in the plant kingdom; they constitute a relatively diverse family of aromatic molecules that are derived from the amino acids phenylalanine and malonyl-coenzyme A. The first biosynthetic step is catalyzed by chalcone synthase (CHS), which uses the malonyl CoA and 4-coumaroyl CoA as substrates, and produces the first subgroup of compounds called chalcones. Indeed, flavonoids include six major subgroups that are found in most higher plants: the chalcones, flavones, flavonols, flavandiols, anthocyanins, and condensed tannins (or proanthocyanidins); a seventh group, the aurones, is widespread, but not ubiquitous (Winkel-Shirley, 2001). These compounds are thought to play a vast array of biological functions.

Kaempferol derivatives are the major flavonoids isolated in Arabidopsis plants (Veit and Pauli, 1999), as revealed also by our metabolomics analysis, which reveals many tetrahydroxyflavones with various glycosylation patterns (Tab. 3.1 and 3.2). Flavonoids were found as negative biomarker in transformed plants. Since their vast array of

.

biological functions, the meaning of the low abundance of flavonoids in the transformed lines is not easy to understand.

Glucosinolates, positive biomarkers for transformed lines, are a well-studied class of plant secondary metabolites. Glucosinolates are of interest for their role in plant defence against microorganisms and herbivores (Schreiner *et al.*, 2009; Beekwilder *et al*., 2008) and for their importance for anticarcinogenic properties (Keum *et al.*, 2005). Depending on the amino acid precursor, glucosinolates are grouped into different classes. Aliphatic glucosinolates originate from Alanine, Methionine, Leucine, Isoleucine or Valine, aromatic glucosinolates from Phenylalanine or Tyrosine, and indole glucosinolates are derived from Tryptophan. Regulators of indolic glucosinolates are MYB51, MYB122 and MYB34, whereas MYB28, MYB29 and MYB76 were shown to act as regulators of aliphatic glucosinolate biosynthesis (Gigolashvili *et al.*, 2008). In our work, only aliphatic and indolic glucosinolates were identified.

Composition and content of glucosinolates have been extensively characterised in different organs and in developmental stages of *Arabidopsis thaliana* (Petersen *et al*., 2002; Brown *et al.*, 2003), in which 40 different glucosinolates have been identified, derived from Methionine, Phenylalanine, or Tryptophan (Kliebenstein *et al*., 2001; Reichelt *et al.*, 2002). With regard to the concentration, dormant and germinating seeds have the highest content of glucosinolates, followed by inflorescences, younger leaves and roots; moreover, younger is the organ analysed, higher is the content of glucosinolates (Brown *et al.*, 2003).

Our analysis performed on leaves with different ages shows similar results: glucosinolates amount is higher in young leaves than in old ones.

REFERENCES

Asad (1994) Isolation and characterization of cDNA and genomic clones of *MsENOD40*; Transcripts are detected in meristematic cells of alfalfa. Protoplasma 183: 10-23

Beecher (2002) Metabolomics: a new 'Omics' technology. Am Genomic/proteomic Technol 2: 40-43

.

Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, Verbocht H, de Vos RC, Morandini P, Aarts MG, Bovy A (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in Arabidopsis. PLoS One 3: e2068

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of *Arabidopsis thaliana*. Phytochemistry 62: 471-481

Bylesjo M, Eriksson D, Kusano M, Moritz T, Trygg J (2007) Data integration in plant biology: the O2PLS method for combined modeling of transcript and metabolite data. Plant J 52: 1181-1191

Ceoldo S, Toffali K, Mantovani S, Baldan G, Levi M, Guzzo F (2009) Metabolomics of *Daucus carota* cultured cell lines under stressing conditions reveals interactions between phenolic compounds. Plant Science 176: 553-568

Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M (1998) Patterns of *ENOD40* gene expression in stem-borne nodules of *Sesbania rostrata*. Plant Mol Biol 37: 67-76

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. Embo J 13: 5099-5112

Gigolashvili T, Engqvist M, Yatusevich R, Muller C, Flugge UI (2008) HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in Arabidopsis thaliana. New Phytol 177: 627-642

Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by enod40 in Arabidopsis thaliana. J Exp Bot 56: 507-513

Keum YS, Jeong WS, Kong AN (2005) Chemopreventive functions of isothiocyanates. Drug News Perspect 18: 445-451

Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T (2001) Genetic control of natural variation in Arabidopsis glucosinolate accumulation. Plant Physiol 126: 811-825

Kouchi H, Takane K, So RB, Ladha JK, Reddy PM (1999) *Rice ENOD40*: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J 18: 121-129

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408

.

Papadopoulou K, Roussis A, Katinakis P (1996) *Phaseolus ENOD40* is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. Plant Mol Biol 30: 403-417

Petersen BL, Chen S, Hansen CH, Olsen CE, Halkier BA (2002) Composition and content of glucosinolates in developing *Arabidopsis thaliana*. Planta 214: 562-571

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002) Benzoic acid glucosinolate esters and other glucosinolates from *Arabidopsis thaliana*. Phytochemistry 59: 663-671

Rochfort SJ, Trenerry VC, Imsic M, Panozzo J, Jones R (2008) Class targeted metabolomics: ESI ion trap screening methods for glucosinolates based on MSn fragmentation. Phytochemistry 69: 1671-1679

Schreiner M, Krumbein A, Ruppel S (2009) Interaction between plants and bacteria: glucosinolates and phyllospheric colonization of cruciferous vegetables by Enterobacter radicincitans DSM 16656. J Mol Microbiol Biotechnol 17: 124-135

Trygg J, Holmes E, Lundstedt T (2007) Chemometrics in metabonomics. J Proteome Res 6: 469-479

Veit M, Pauli GF (1999) Major flavonoids from Arabidopsis thaliana leaves. J Nat Prod 62: 1301-1303

Verpoorte CY, Mustafa NR, Kim HK (2008) Metabolomics: back to basic. Phytochem reviews: 525-537

Verpoorte R, Choi, Y.H., Kim, H.K (2007) NMR-based metabolomics at work in phytochemistry. Phytochemistry Review 1: 3-14

Wan X, Hontelez J, Lillo A, Guarnerio C, van de Peut D, Fedorova E, Bisseling T, Franssen H (2007) *Medicago truncatula ENOD40-1* and *ENOD40-2* are both involved in nodule initiation and bacteroid development. J Exp Bot 58: 2033-2041

Winkel-Shirley B (2001) Flavonoid Biosynthesis. A Colorful Model for Genetics, Biochemistry, Cell Biology, and Biotechnology. Plant physiology 126: 485-493

.

Table 3.1 LC-MS compounds identification in positive ionization mode

n°	rt	m/z	ms/ms (ms3)	putative identification	ab b
51	2.114	110	100	glucoiberverin fragment	GL
23	2.497	382	**318**, 220 (189)	1-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
38	2.681	366	237	1-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
13	2.956	182	118	tyrosine	AA
20	2.987	118		valine	AA
24	3.030	308		glutathione (reduced form)	
174	3.906	132	86	leucin/isoleucin	AA
9	3.920	87		leucin/isoleucin, fragment	AA
170	4.954	166	118	glucoiberverin fragment	GL
14	4.967	310	**148** (100)	glucoiberverin fragment	GL
63	4.971	328	166, 118	glucoiberverin fragment	GL
88	5.001	312	148,100	glucoiberverin	GL
46	5.283	369	351-**207** (190)	4-hydroxyglucobrassicin, fragment	GL
10	5.718	120	110, 93	phenylalanine, fragment	AA
281	5.723	166	120	phenylalanine	AA
28	11.871	369	207, 130	3-indolylmethylglucosinolate (fragment)	GL
21	11.884	207	130	3-indolylmethylglucosinolate (fragment)	GL
65	11.986	351		4-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
55	12.184	205	188	tryptophan	AA
48	12.188	188		tryptophan, fragment(-NH3)	AA
5	12.474	277	260-114	coumaroylagmatine (plant J 2009, 57:555) trans	
1	14.163	395	364, 233	sinapoymalate (isomer,fragment)	HC
8	14.509	252	184	8-Methylsulfinyl-n-octylglucosinolate (fragment)	GL
29	14.509	381	252-184	8-Methylsulfinyl-n-octylglucosinolate (fragment)	GL
3	14.514	414	**252** (184)	8-Methylsulfinyl-n-octylglucosinolate (fragment)(glucohirsutin)	GL
34	14.514	415	381, **252** (184,136)	glucohirsutin fragment, isotope	GL
11	15.844	277	260-114	coumaroylagmatine (plant J 2009, 57:555) cis	
60	16.390	207	175	sinapoymalate (isomer,fragment)	HC
15	16.710	399	**237** (160)	1-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
50	16.710	237	160	1-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
35	16.748	383	237-160	1-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
147	18.441	433	287	kaempferol deoxyhexose hexose deoxyhexose, fragment	FL

.

72	18.980	435	273	trihydroxy hexoside flavanone	FL
158	19.555	306	166, 118	glucoiberverin fragment	GL
77	20.765	399	364-237	4-Methoxyindol-3-ylmethylglucosinolate (fragment)	GL
108	22.376	595	433, 287	kaempferol deoxyhexose hexose	FL
22	24.528	207	**175** (147)	sinapoymalate (fragment)	HC
2	24.657	363	229, 157	glucoiberverin fragment	GL
64	25.219	579	433, **287** (241, 185)	kaempferol deoxyhexose deoxyhexose	FL
36	25.227	433	287	kaempferol deoxyhexose deoxyhexose, fragment	FL
16	28.034	202	**138** (121, 93)	3-indolylmethylglucosinolate (fragment)	GL
136	28.034	224	160	glucosinolate fragment	GL
17	34.647	382		Methoxy indolyl 3 methylglucosinolate, fragment	GL
141	43.019	220	143	pantothenate	

Table 3.2 LC-MS compounds identification in negative ionization mode

n°	rt	m/z	ms/ms (ms3)	putative identification	abb
219	2.183	389	341(179,143)	sucrose, formic adduct isotope	
205	2.279	381	341, 179	sucrose, chlorine adduct, isotope	
184	2.288	343	341,179	sucrose, isotope	
133	2.329	379	341, 179	sucrose, chlorine adduct, isotope	
58	2.333	387	**341**(179,143)	sucrose, formic adduct	
65	2.346	341	179, 143	sucrose	
116	2.520	377	341, 179	sucrose, chlorine adduct	
159	2.541	133	115	malic acid	AAO
99	2.574	115	71	fumaric acid	AAO
55	2.647	425	407,**358**, 259 (259,162)	glucoiberin isotope	GL
1	2.965	422	407,**358**, 259 (259,162)	3-methylsulphinylpropyl (glucoiberin)	GL
152	2.997	191	173, 111	citric acid	AAO
100	3.019	436	372, 259	glucoraphanin	GL
404	3.065	111	67	citric acid, fragment	AAO
18	3.301	115	71	maleic acid	AAO
250	4.187	299	252, 173, 136	hydroxybenzoyl hexoside	HB
7	4.939	406	326, 275, **259**, 229 (241,198)	glucoiberverin	GL
17	5.269	463	**285**, 267, 259 (204, 97)	4-hydroxyglucobrassicin	GL

.

16	6.144	315	**152** (109)	dihydroxybenzoyl glucoside	HB
12	8.237	478	**414** (275, 259)	glucosinolate	GL
60	8.521	420	340, 275, 259, 242, 227	glucoerucin	GL
13	11.860	449	275, **259**, 195 (139, 97)	3-indolylmethylglucosinolate, isotope	GL
4	11.865	447	275, **259**, 195 (139, 97)	3-indolylmethylglucosinolate	GL
66	11.877	450	275, **259**, 195 (139, 97)	3-indolylmethylglucosinolate isotope	GL
681	14.155	353	191, 176	caffeoylquinic acid	HC
194	14.159	191	176	caffeoylquinic acid, fragment	HC
35	14.511	495	259	glucohirsutin isotope	GL
2	14.516	492	428, 186	8-Methylsulfinyl-n-octylglucosinolate (glucohirsutin)	GL
76	14.525	496	259	8-Methylsulfinyl-n-octylglucosinolate (glucohirsutin) isotope	GL
10	14.534	494	428, 186	glucohirsutin, isotope	GL
122	15.776	431	385 (179)	sinapoyl hexoside 1 formic adduct	HC
545	16.055	432	179	sinapoyl hexoside 1 formic adduct isotope	HC
15	16.350	385	265, 247, **223**, 205 (208, 164)	sinapoyl hexoside 1	HC
45	16.433	431	**385** (179)	sinapoylhexoside 1 formic adduct	HC
6	16.713	477	259	Methoxy indolyl 3 methylglucosinolate (neoglucobrassicin)	GL
125	16.713	480	259	Methoxy indolyl 3 methylglucosinolate (neoglucobrassicin), isotope	GL
29	16.717	479	259	Methoxy indolyl 3 methylglucosinolate (neoglucobrassicin), isotope	GL
256	16.778	386	179	sinapoyl hexoside 1 isotope	HC
48	17.032	385	179	sinapoyl hexoside 2	HC
329	17.329	386	179	sinapoyl hexoside 2 isotope	HC
49	17.366	385	179	sinapoyl hexoside 2	HC
807	18.423	775	609, 285	kaempferol hexose deoxyhexose deoxyhexose chlorine adduct	FL
52	18.437	739	593, 285	kaempferol hexose deoxyhexose deoxyhexose	FL
693	18.441	741	593, 285	kaempferol hexose deoxyhexose deoxyhexose isotope	FL
234	18.446	740	593, 285	kaempferol hexose deoxyhexose deoxyhexose isotope	FL
499	18.606	501	**223** (208,164, 149, 135)	sinapoyl malate hexoside	HC
124	19.219	681	**519** (358)	monohydroxytetramethoxyflavone hexosyl hexoside	FL
64	19.387	431	**385** (179)	sinapoyl hexoside 3 formic adduct	HC
20	19.957	431	385 (179)	sinapoyl hexoside 4 formic adduct	HC
377	20.273	385	179	sinapoyl hexoside 4	HC

.

182	20.585	755	**609** (285)	kaempferol hexose hexose deoxyhexose	FL
9	20.759	477	446, 259	4-Methoxyindol-3-methylglucosinolate	GL
51	20.759	479	**431** (285)	kaempferol deoxyhexose deoxyhexose isomer	FL
397	22.385	595	431-285 (kaempferol) 447(quercetin)	co-eluting quercetin-deoxyhexose deoxyhexose and kaempferol hexoside-deoxyhexose, isomer	FLs
131	22.394	594	431-285 (kaempferol) 447(quercetin)	co-eluting quercetin-deoxyhexose deoxyhexose and kaempferol hexoside-deoxyhexoside	FLs
3	24.501	223	208, 179, **164**, 149 (149)	sinapic acid, fragment of sinapoyl malate 1	HC
5	24.547	339	**223** (208,164, 149, 135)	sinapoyl malate 1	HC
655	25.185	580	**431**, 285	kaempferol deoxyhexose deoxyhexose isotope	FL
63	25.213	640	**577** (285, 257)	kaempferol di-deoxyhexose derivative	FL
11	25.217	577	205, 265, 344	kaempferol deoxyhexose deoxyhexose	FL
135	25.217	579	**431** (285)	kaempferol deoxyhexose deoxyhexose isotope	FL
46	25.243	578	**431**, 285	kaempferol deoxyhexose deoxyhexose isotope	FL
27	26.804	462	259	glucosinolate	GL
145	26.804	464	259	glucosinolate, isotope	GL
694	26.813	530	259	glucosinolate derivative	GL
242	29.448	535	419, 223	sinapic acid derivative	HC
425	34.632	480	259	Methoxy indolyl 3 methylglucosinolate, isotope	GL
31	34.641	478	259	Methoxy indolyl 3 methylglucosinolate, isotope	GL
154	34.641	479	259	Methoxy indolyl 3 methylglucosinolate, isotope	GL
8	34.646	476	298,275, **259**, 195 (138)	Methoxy indolyl 3 methylglucosinolate	GL
208	34.650	544	259	Methoxy indolyl 3 methylglucosinolate derivative (= + 68)	GL
108	38.718	339	223 (208,164, 149, 135)	sinapoyl malate 2	HC
598	38.746	340	**223** (208,164, 149, 135)	sinapoyl malate 2 isotope	HC
255	38.843	223	208, 179, 164, 149 (149)	sinapic acid, fragment of sinapoyl malate 2	HC

Tab. 3.1 and 3.2 LC-MS compounds identification in positive and negative ionization modality. 1360 signals were detected and 42 and 58 molecules, respectively, in positive and negative ionization mode were identified (the isotopes are reported on the tables but are not counted as new identified molecules). Several identification of the metabolites were performed comparing our data with those obtained by Rochfort *et.al.,* (2008).Legend: n°: molecule identification number (abundance of molecule), rt:retention time (minutes), abb: metabolites abbreviation; GL: glucosinolate, FL: flavonoid, HC: hydroxycinnamic acid; HB: hydroxybenzoic acid, AA: amino acid, AAO: organic acid. The base ione, with its fragmentation is reported between bracket, in bold.

.

.

Chapter 4

Transcriptional profiling of *Arabidopsis thaliana* plants overexpressing *ENOD40*

.

.

ABSTRACT

Homologues of *ENOD40* gene have been identified in several plant species across the plant kingdom; but, to date, the precise function of the gene is still unclear.

In this chapter, transcriptional profiling of three different transformed *GmENOD40 Arabidopsis thaliana* lines and wild type plants were compared to study the effect of *ENOD40* in non-legume plants. Biomarkers analysis revealed that, in transformed plants, several upregulated genes are closely related to cell wall processes, such as wall strengthening and delivery of new components.

INTRODUCTION

ENOD40 is an early noduline gene induced during nodule formation in symbiotic interaction between legume plants and Rhizobia. The highest expression levels of *ENOD40* is during nodule formation and, therefore, several researches have been done during this process to unravel *ENOD40* function. In the last years it was demonstrated that *ENOD40* gene is essential for nodule initiation and nodule organogenesis (Wan *et al.*, 2007; Kumagai *et al*., 2006).

However, homologues of the gene were also identified in non-legume plants like rice, maize, citrus, tomato and tobacco, indicating that the role of *ENOD40* is not only restricted to the nodulation process. *ENOD40* expression patterns are similar in non-symbiotic tissue of legumes and in non legume plants. In rice, *ENOD40* expression was detected at the early stages of the development of lateral vascular bundles that conjoin the emerging leaf (Kouchi *et al*., 1999); in tomato, it is found in flower and lateral root development, in germinated seeds and in vascular tissue of stems, young leaves and roots (Vleghels *et al*., 2003). In legume plants, gene expression was detected in root and stem vascular tissues (Varkonyi-Gasic and White, 2002) and in non symbiotic meristematic tissues, such as developing lateral roots (Asad, 1994; Papadopoulou *et al.*, 1996; Fang and Hirsch, 1998), young leaves, stipule primordia (Asad, 1994; Corich *et al.*, 1998) and embryonic tissues (Flemetakis *et al.*, 2000).

Although different expression patterns were observed and some putative functions, as transport or organogenesis, were proposed, the precise function of *ENOD40* in non-legume plants is still unclear.

In this chapter, the role of *ENOD40* was investigated in *Arabidopsis thaliana* plants.

.

Previous studies in Arabidopsis plants have demonstrated that ectopic expression of *GmENOD40* did not lead to severe changes in overall plant architecture but led to a reduction in cell size (Guzzo *et al*., 2005).

In recent years, the use of microarray analysis to study genome-wide gene expression has been a very powerful tool for the identification of differentially expressed genes specific to a given biological process or treatment. Here microarray techniques were used to study the effect of *ENOD40* overexpression, comparing expression profiling of three Arabidopsis lines transformed with 35S::*GmENOD40* and wild-type plants. Thereafter, bioinformatical and statistical tools were used to extract the information from the results.

MATERIALS AND METHOD

Plant Material

Wild type and transformed (1.1, 3.1 and 4.4) plants were collected, ground in liquid nitrogen lyophilized and then the obtained power were split in two part, one used for transcriptomic analysis and the other used for metabolomic analysis. Plants details are reported in chapter 3.

RNA extraction, quantification and integrity check

RNA was extracted from lyophilized plant material using the Spectrum Plant Total RNA kit (Sigma-Aldrich). After RNA isolation, samples were treated with RQ1 RNase-Free DNase (Promega) to remove any contaminating DNA. The RNA was quantified by NanoDrop spectrophotometer (ND-1000, Thermo Fischer Scientific Inc. Waltham). The RNA samples used for CombiMatrix array gene chip experiments (CombiMatrix, Mukilteo) were quantified and quality checked by Agilent 2100 BYOANALYZER (Santa Clara).

.

Array conception

A total of 27763 oligonucleotide probes (35- to 40-mer) were designed on the last 1500 bp of the transcripts. CustomArray 90K was designed with three replicates of each probe randomly distributed across the array to allow measurement of the variability within the array. 23 bacterial and viral oligonucleotides sequences and 48 Ambion spike-in specific oligos (Cat.1780) were used as negative controls. Three biological replicates, each one composed by eight plants as described in "Material and Methods" section chapter 3, were used for each hybridization experiment.

Amplification, labeling and hybridization procedures

Amino allyl aRNA synthesis and dye Cy5-ULS labeling were performed using the RNA ampULSe: Amplification and Labeling Kit for Combimatrix arrays (KREATECH biotechnology). 2 μg of extracted RNA in the amplification procedure and 6 μg of aRNA for the labeling procedure were used.

The purified labeled aRNA was quantified by spectrophotometer (RNA at λ_{260nm} and incorporated Cy5 at λ_{650nm}) and 5 μg of labeled aRNA were used for array hybridization. Pre-hybridization, RNA fragmentation, hybridization, washings and imaging were performed as indicated in CustomArray™ 90K Microarray Hybridization and Imaging Protocol (PTL020).

Microarray data analysis

The slides were scanned at 5 μm resolution, using a GenePix 4300A microarray scanner with Gene Pix Pro 7 software. PMT gain and laser power values were adjusted in order to get a signal around 70% of maximum measurable signal. Scanning is followed by densiometric quantification of the spots using the Microarray Imager 5.8 software (http://bioapps.combimatrix.com).

Data were normalized using quantile normalization and background signal was calculated on the basis of negative control signal intensities, as the average of the signal of negative controls plus 2 times their standard deviation. The data were corrected by filtering out the signals below the calculated background. Data were statistically analyzed by tMev software (http://www.tm4.org/index.html) using SAM (Statistical

.

Analysis for Microarrays) multiclass analysis (Tusher *et al.*, 2001). The SAM analysis assigns a score to each gene based on the change in gene expression relative to the standard deviation of repeated measurement. The percentage of genes identified as differentially expressed for pure randomness is defined as false discovery rate (FDR) and is estimated through the use of permutations. In order to select the set of gene differentially expressed , an FDR threshold of 15% was selected.

The obtained matrix with differentially expressed genes was imported in SIMCA-P software (v. 12.0, Umetrics). Principal Component Analysis (PCA) and Bidirectional Orthogonal Projections to Latent Structures Discriminant Analysis (OPLS-DA) were performed on the data matrix using unit variance scaling, as described in Material and Methods Data Analysis, Chapter 3. In this work, variables with p(corr) $\geq \pm$ 0.9 have been considered as biomarkers.

Real-Time PCR

The RNA extracted from lyophilized plant material with the Spectrum Plant Total RNA kit (Sigma-Aldrich) and treated with RQ1 RNase-Free DNase (Promega) used for microarray analysis was also used to perform qRT-PCR. The cDNA was synthesized using the Improm-II^{TM} Reverse Transcription System kit (Promega); 1μg of RNA treated with DNase was reverse-transcribed, whereas the RNA not reverse-transcribed was used as negative control in subsequent Real-Time PCR experiments.

PCR reaction mix contained: 5 μl of cDNA diluted 1:10, 12.5 μl of Platinum ® SYBR ® Green qPCR SuperMix UDG with ROX (Invitrogen), 0.5 μl of each primer (stock 20 μM), and sterile double-distilled water to a final volume of 25 μl. Each amplification was carried out in triplicate.

Real-time PCR amplification was performed on an ABI PRISM 7000 sequence detection system (Applied Biosystems) programmed for initial denaturing at 95°C for 10 min, and 40 cycles of denaturing at 95°C for 15 sec, annealing at 55°C for 30 sec and extension at 72°C for 30 sec Analysis of the dissociation curve of PCR products at the end of the PCR run revealed a single narrow peak for each amplification product. The relative expression was calculated according to $2^{-\Delta\Delta Ct}$ method as described by Livak and Schmittgen (2001).

.

Primer Design

All the primers were designed using Primer 3 (http://frodo.wi.mit.edu/).

Primer Name	Primer Sequence (5'-3')
p-glycoprot21 forward	CTTACGCCGCCAGTTTCTAC
p-glycoprot21 reverse	ATAGCTACAGCCGCCATTGT
xylogluc forward	CCCTTCTTTTCCCGGTACTC
xylogluc reverse	GCTTTTCCGTCTTCCATTTG
MYB29 forward	CATGTTGTGTGGGAGAAGGA
MYB29 reverse	AGTCCAGCTTTTTGGGGAAT
COBL2 forward	GGTGGTCAAGCAACAGAACA
COBL2 reverse	TCACACCGCCTCTACAACAG
AKN2 forward	ATGCAGATGCGACAGACAAC
AKN2 reverse	CGCCGTCGAGTGTGTAAGTA
EXL3 forward	GCTCCGATCAAATCTTCGAC
EXL3 reverse	TGAACGGTGATGTTGTTGGT
XTR7 forward	TGGCAGGAGGAGACTCAGAT
XTR7 reverse	TCTTGCATTCTGGAGGGAAT
GASA1 forward	ACGACAAGTGCCAGTGCTAC
GASA1 reverse	ACGACATCGTCCCATTATCC
SQD2 forward	TACCTGAAGCTCGGATTGCT
SQD2 reverse	TGTGAGAGTTCATCGCCTTG
EXLA1 forward	ACATTGCTCAAGTCGGTTCA
EXLA1 reverse	CGTCGTATCCACCGGTTACT
XTH33 forward	TACAAGTACGCCCCTTTCGT
XTH33 reverse	CCATCAACAGGGTCCAGACT
PAT2 forward	GTCCTTTCTGGTCCATCCAA
PAT2 reverse	TAACTACGGCCTCGGTGACT
MIPS1 forward	AGAAGACCGTCAATGGCACT
MIPS1 reverse	CCCACAAGCATAACCCCTAA
HKL1 forward	GGTCCGATTCACAAGGTGTT
HKL1 reverse	ATCTTGCCCAACCATTTCAC
EXPA5 forward	CGCATGTTTTGAGCTGATGT
EXPA5 reverse	AAGATCGAAATGGTGGTTGG
PCS1 forward	CATTTCCACCACAACGTCAC

.

PCS1 reverse	CAACGAAGCCAGGAGAGTTC
GoIS1 forward	ACTACGTGAAAGGGGTGGTG
GoIS1 reverse	CAACTGTTTCCGGTGATCCT
AGL87 forward	CTCGCCGAGTTACATTCAGA
AGL87 reverse	AGCTCAGGGCCATCTTTGTA
GUN4 forward	TTTTAAGCCATCCTGCGTTT
GUN4 reverse	CTGCTCCTACTCCTGCCTGT
PIP 1;5 forward	TCCCAATATGTGTGCCTCTG
PIP 1;5 reverse	CTTCCTCGCCAAGAACAAAC
Actin1 forward	CCGAGCGTGGTTACTCTTTC
Actin1 reverse	GAGCTGGTTTTGGCTGTCTC

RESULTS

Transcriptomic analysis was carried out on the same transformed and wild type *A. thaliana* plants used for metabolomic analysis (chapter 3). The three lines 1.1, 3.1 and 4.4 were transformed with *GmENOD40* and the expression of the transgene was confirmed by Real-Time PCR using *GmENOD40* specific primers, as reported in "Results", chapter 3.

Transcriptional profiling and identification of differentially modulated genes

In order to investigate transcriptomic changes due to *GmENOD40* gene, transcriptional profiling of transformed and wild type plants were performed by microarray analysis technique using a 90K Combimatrix chip.

The chip contains 27763 oligonucleotides probes that match 28675 transcripts and 71 negative controls. Three biological replicates were analysed for each line and the obtained data were normalized by quantile method. A good Pearson correlation, ranging from 0.92 to 0.96, was determined for biological replicates of each line. The datasets were analyzed by a significance analysis of microarray, SAM multi-class analysis (Tusher *et al*., 2001) using TMeV software v. 3.0 with a selected False Discovery Rate (FDR) ≤15%. In total, 382 transcripts were identified as differentially modulated (Table 4.1); of these, 142 were upregulated and 165 downregulated in all the three transformed lines compared to wild-type plants, while 75 genes were up or downregulated in only one or two transformed lines compared with the control.

.

Analysis of transcriptomic data by chemometric approach

To identify biomarker genes within the group of 382 differentially modulated transcripts identified with the SAM multi-class analysis, the data matrix containing the differentially expressed genes were imported in SIMCA-P software v.12.0. Biomarkers, as already reported in chapter 3, are defined as molecules indicating or correlating with a physiological change; this could include both positive and negative biomarkers, the former characterizing all the three transformed lines and the latter characterizing the wild type plants.

First of all, unsupervised PCA and O2PLS discriminant analysis were performed on the dataset (the 382 differentially modulated transcripts identified with the SAM multi-class analysis) to look for possible correlation, based on these transcripts, among the four classes. Unsupervised PCA and supervised O2PLS discriminant analysis show a clear separation between controls and the three transformed lines along the first component, as reported in the O2PLS-DA score plot (Fig. 4.1).

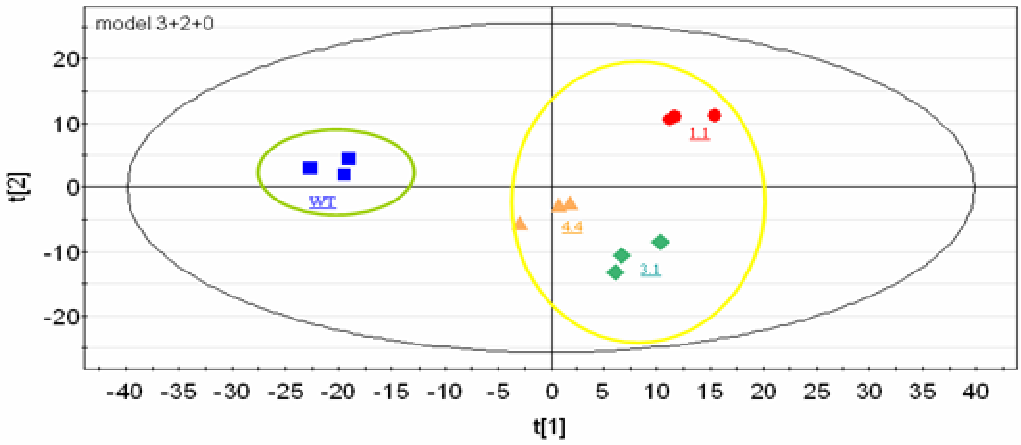


Fig. 4.1 O2PLS-DA score plot (UV scaling) using as classes the wild type sample (blue square) and each of three *GmENOD40* overexpressing lines (1.1: red circle, 3.1: green diamond, 4.4: yellow triangle); using as x variables the 382 differentially expressed genes. In this plot each point represents a biological replicate.

This result means that, as expected, in the transcript modulation induced by the expression of the *ENOD40* gene, some common features clearly distinguish all the three transformed lines compared with the control; furthermore, the separation of the 1.1 transformed line along the second component, indicates that this line has some

.

characteristic features compared with the two 3.1 and 4.4 transformed line. Moreover, the position close to the axes origin of the 4.4 transformed line indicates that the latter has intermediate features between control line and 1.1 and 3.1 lines, being probably the one where the effects of transformation are lower.

If PCA and O2PLS-DA are performed on the transformed lines only, the difference between 1.1 line compared with 3.1 and 4.4 lines is confirmed, as shown by the score plot of figure 4.2.

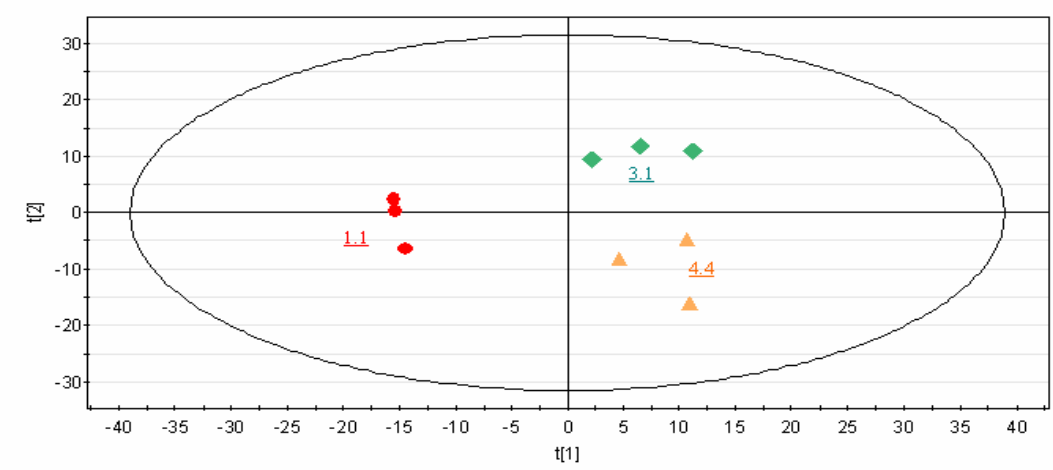


Fig. 4.2 PCA score plot (UV scaling) of transcriptional profiling of the three *GmENOD40* overexpressing lines (1.1: red circle, 3.1: green diamond, 4.4: yellow triangle). In this plot each point represents a biological replicates and were used as x variables the 383 differentially expressed genes.
The same outcome was obtain by O2PLS-DA analysis (data not shown).

Part of these differences among the transcriptomes of the transformed lines could be ENOD40 transcript dose-dependent, since, as shown in chapter 3, the 1.1 transformed line has the highest *GmENOD40* expression, evaluated by RT-PCR (see fig. 3.2).

Afterward, three distinct two-classes O2PLS-DA models were developed for each transformed line compared with wild type plants (model M6:WT vs1.1, model M7: WT vs 3.1 and model M8: WT vs 4.4) to investigate the difference between the wild-type plants and each transformed line. Subsequently the transformed lines were compared pairwise by Shared and Unique Structure plot (SUS-plot) using the results of the previous models. SUS-plot is a useful tool to compare the outcomes of multiple classification models compared to a common reference, as reported in chapter 3.

Biomarkers specific for all the three transformed lines were identified comparing the

.

possible combinations of the three models (Fig. 4.3).

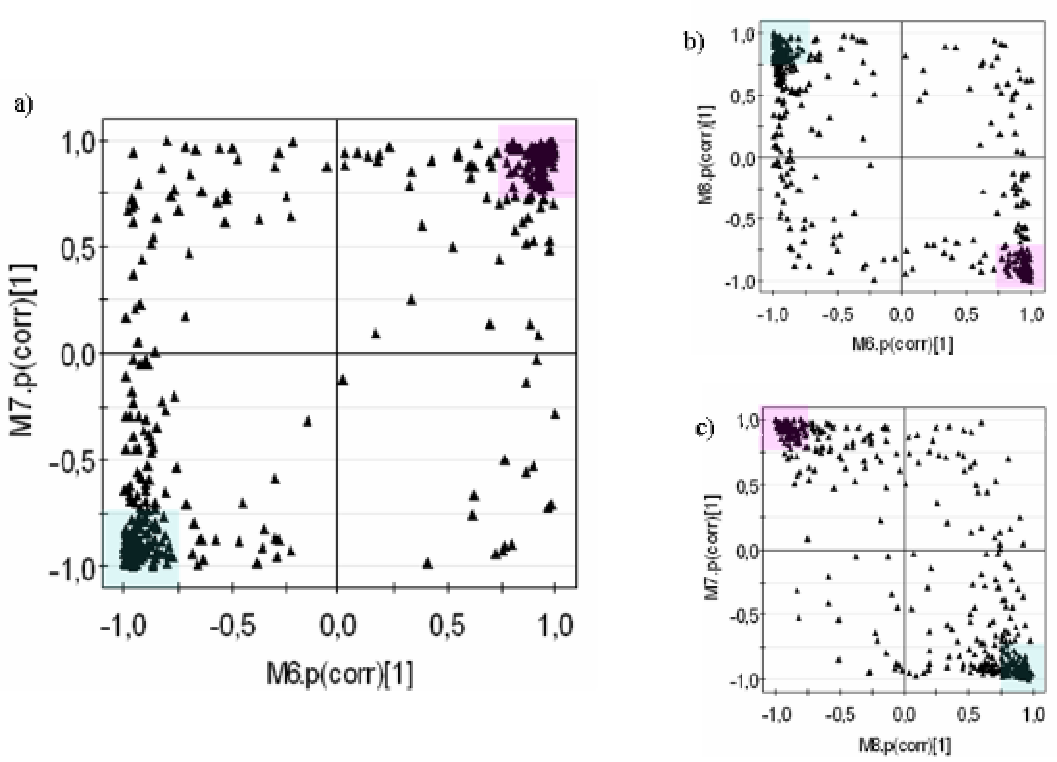


Fig. 4.3 a) SUS-plot between the model M6 (WT vs 1.1) and M7 (WT vs 3.1) b) SUS-plot between the model M6 (WT vs 1.1) and M8 (WT vs 4.4). c) SUS-plot between the model M8 (WT vs 4.4) and M7 (WT vs 3.1). 382 differentially modulated genes were analysed; each triangle represents one gene. Blu region: genes upper regulated in wild type plants, pink region: genes downregulated in wild type plants.

Twenty-three genes were identified as positive biomarkers of the transformed lines, whereas eighteen genes were identified as negative biomarkers. (Tab 4.2). Among positive biomarkers, most of the transcripts identified (12 out of 23) were correlated with processes occurring in the cell wall; some of these were involved in cell wall remodelling, as xyloglucan endotransglycosylase/hydrolase (XTH33, At1g22740; XTR7, At4g14130; xyloglucan:xyloglucosyltransferase putative, At5g65730) or EXPANSIN-LIKE A1 (ATEXLA1, At3g45970), and other were involved in the delivery (RAB GTPASE HOMOLOG A5E, At1g05810) or in the biosynthesis (MYO-INOSITOL OXYGENASE 2, At2g19800) of new cell wall components. Among negative biomarkers, no cluster of genes involved in the same process was identified; two modulated genes were involved in the glucosinolate biosynthetic pathway (APS

.

kinase 1, At2g14750 and APS kinase 2, At4g39940); two genes were involved in the biosynthesis and transport of lipids (STEROL 4-ALPHA METHYL OXIDASE 1-3, At4g22753, and a lipid transfer protein, At5g64080), and other two genes were involved in catalytic processes (phosphodiesterase / nucleotide pyrophosphatase related, At4g29710, and the phenazine biosynthesis PhzC/PhzF family protein, At4g02850).

Among all the 382 differentially expressed genes (Table 4.1), further genes, induced in the three transformed lines, although not identified as biomarkers, were correlated with processes occurring in the cell wall. Between these, genes involved in the cell wall remodelling (xyloglucan endotransglycosylase/hydrolase18, At4g30280; endo-xyloglucan transferase putative, At3g23730; Touch4, At5g57560; meristem-5, At4g30270), in lignin biosynthetic process (METHIONINE OVER-ACCUMULATOR 3, At3g17390) or encoded structural constituent of cell wall (leucine-rich repeat family protein / extensin family protein, At4g13340) were identified.

With regard to the genes downregulated in the transformed plants, various other genes, not included in the biomarker analysis and involved in the biosynthetic glucosinolate pathway were found (CYP79B3, At2g22330; CYP79B2, At4g39950; SULFOTRANSFERASE17, At1g18590; FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 3, At1g62560); among these also two MYB factors (MYB29, At5g07690 and MYB34, At5g60890), considered as regulators of glucosinolate pathway, were identified. Moreover, four expansin genes (EXPANSIN A5, At3g29030; EXPANSIN A16, At3g55500; EXPANSIN B3, At4g28250 and EXPANSIN A10, At1g26770) were also downregulated (Table 4.1).

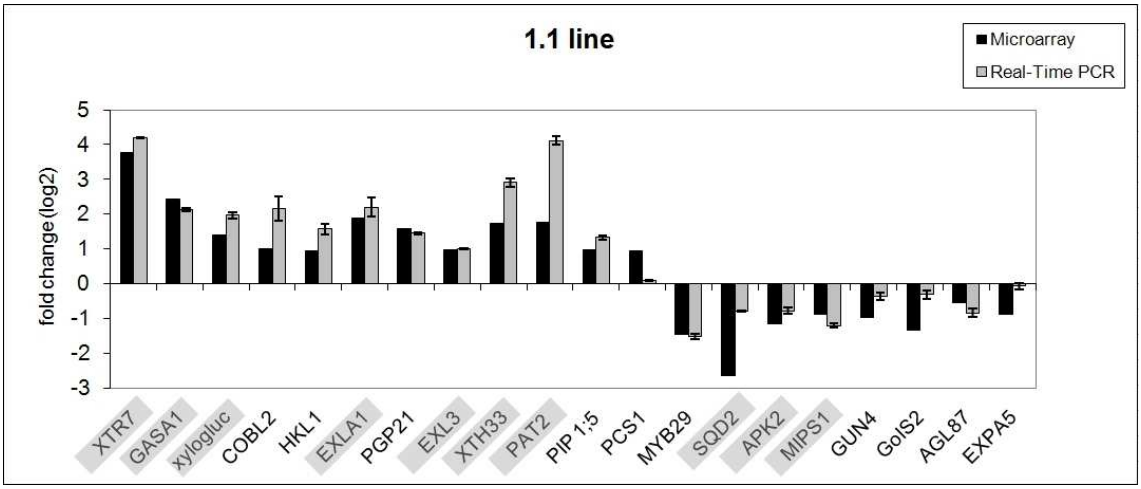
.

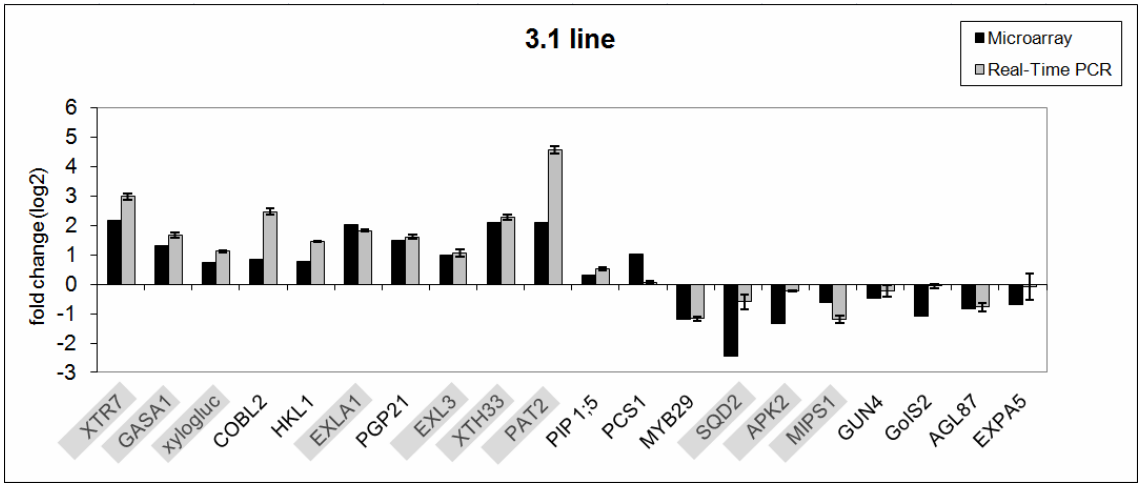
Validation of Microarray data

The validation of microarray results was performed by quantitative Real-Time PCR.

A subset of twenty differentially expressed genes was randomly selected from the dataset containing the 382 significantly modulated genes. The RNA of wild type and transformed lines used for qRT-PCR was the same used for microarray analysis. Actin gene was used as a internal control to normalize gene expression. Fold changes were compared between microarray analysis and qRT-PCR analysis (Fig. 4.4).

Although there was difference in fold-changes between the two techniques, for all the genes tested the modulation observed by Real-Time PCR was in agreement with the microarray analysis.





.

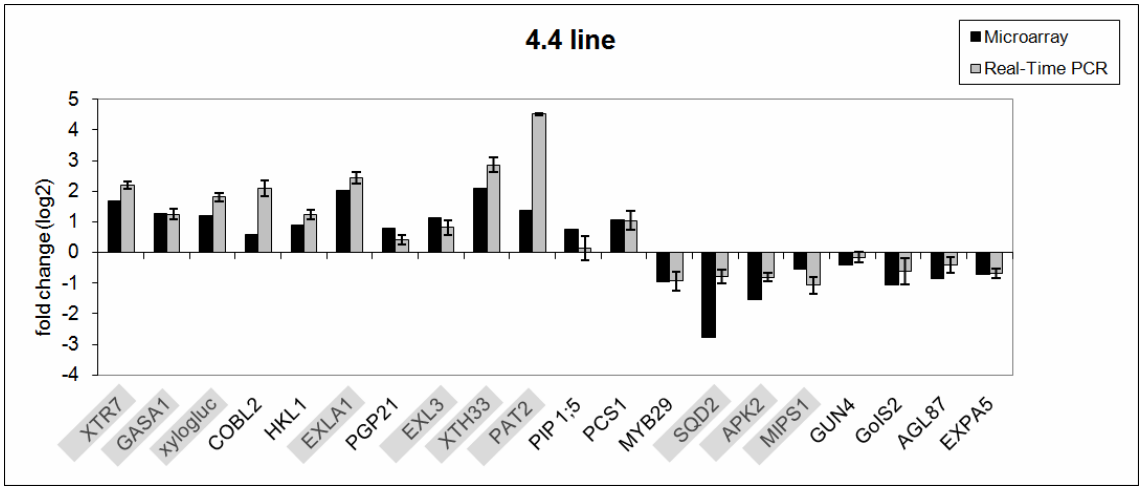


Fig.4.4 Validation of Microarray result using Real-Time qRT-PCR. Values are expressed as fold change (log_2; mean ± SE). Black and grey bars represent, respectively, the fold change obtained with microarray analysis and qRT-PCR.
XTR7, XYLOGLUCAN ENDOTRANSGLYCOSYLASE 7; GASA1, GAST1 PROTEIN HOMOLOG 1; xylogluc, xyloglucan:xyloglucosyl transferase, putative; COBL2, COBRA-LIKE PROTEIN 2 PRECURSOR; HKL1, HEXOKINASE-LIKE 1; EXLA1, EXPANSIN-LIKE A1; PGP21, P-glycoprotein 21; EXL3, EXORDIUM LIKE 3; XTH33, xyloglucan endotransglucosylase/hydrolase; PAT2, PATELLIN 2; PIP 1;5, PLASMA MEMBRANE INTRINSIC PROTEIN 1;5; PCS1, PROMOTION OF CELL SURVIVAL 1; MYB29, MYB domain protein 29; SQD2, sulfoquinovosyldiacylglycerol 2; APK2, APS KINASE 2; MIPS1, MYO-INOSITOL-1-PHOSTPATE SYNTHASE 1; GUN4;GoIS1, galactinol synthase 1; AGL87, AGAMOUS-LIKE 87; EXPA5, EXPANSIN A5. The gene identified as biomarkers are highlighted.

DISCUSSION

Homologues of *ENOD40* gene have been identified in legume and non-legume plants; in a recent work, an *ENOD40* gene was also identified in Arabidopsis (Gultyaev and Roussis, 2007), as reported in "General introduction". The gene expression, observed in legume and non-legume plants, both in symbiotic and in non symbiotic tissues, suggests a not-yet elucidated general role of *ENOD40* in plants, not limited to the nodule initiation and maintenance, where the role of this gene has been more extensively investigated. To date, although extensive researches have been done in the past, the function of the gene remains unclear.

In this research, microarray analysis was used to identify the genes modulated by *ENOD40* overexpression and to get clues on *ENOD40* role in non-legume plants. Therefore, the transcriptional profiles of three transformed *Arabidopsis thaliana* lines named 1.1, 3.1 and 4.4, compared to wild type plants were investigated. A previous work, in which the same transformed Arabidopsis lines were used, showed that

.

ENOD40 overexpression induces a reduction of cell size in various tissues, although the organs maintain the normal size (Guzzo *et al.*, 2005).

Using the microarray analysis, 382 genes were identified as differentially modulated.

Statistical Multivariate analysis (PCA and O2PLS-DA) performed on the 382 differentially expressed genes clearly discriminated between the transformed lines and wild type plants, indicating differences in transcriptional profiling induced by *ENOD40* overexpression. Moreover, as observed in the "Results", some differences exists also between the transformed lines, and these differences correlate with the level of *GmENOD40* gene expression, suggesting that there are some effects due to ENOD40 transcript doses.

A multivariate-based analysis method, called biomarker analysis, was used to further restrict the group of transcripts differentially expressed in all the transformed lines compared to the control one. The proposed marker analysis method was aimed to mark, within the group of 382 differentially modulated transcripts identified with the SAM multi-class analysis, those transcripts whose modulation in all the three transformed lines compared to the control showed the highest significance.

Twenty-three genes were identified as positive biomarkers in all the three transformed lines. Based on database searches, the *ENOD40* induced genes were annotated and twelve out of twenty-three genes were involved in cell wall related processes. Further genes, induced in the three transformed lines, although with lower significance than those identified as biomarkers, were correlated with processes occurring in the cell wall.

The plant cell wall is a dynamic structure that exhibits considerable spatial and temporal variability in terms of composition and organization, reflecting a balance between wall synthesis, deposition, reorganization and selective disassembly.

This is achieved through the coordinated action of a battery of wall synthesizing and modifying enzymes that collectively provide a mechanism for regulating cell size and shape (Carpita, 2000). Indeed, the cell wall is the mechanical determinant of shape and size of each plant cell (Zhong and Ye, 2007) and a coordinate expansion and differentiation of the individual cell is due to a dynamic remodelling of the wall structure (Carpita and Gibeaut, 1993).

The cell wall is composed of cellulose microfibrils embedded in a hydrated matrix that

.

is made of complex polysaccharides, (hemicelluloses and pectins), and a small amount of structural proteins; all the components are associated in a mature wall (Carpita and Gibeaut, 1993). This polysaccharide-rich extracellular matrix determines the cell shape and cell wall tensile properties that provide mechanical strength to the plant (Rose and Bennett, 1999).

Among the genes identified as positive biomarkers, characterising the three transformed lines, one EXPANSIN-LIKE A1 (*ATEXLA1*) and three xyloglucan endotransglycosylase/hydrolase, XTHs (*XTR7, XTH33* and a xyloglucan: xyloglucosyltransferase putative) were identified. Expansins are primary wall loosening agents and their action alone is sufficient to restore extensibility of cell walls (McQueen-Mason *et al*., 1992; Cosgrove, 1998). Expansins appear to have a key function in the regulation of cell expansion by coordinating not only hydrogen bond breakdown and so cell wall loosening, but also cellulose synthesis, deposition and spatial organization in relation to other polymers and protein constituents of the cell wall (Zenoni *et al.,* 2004). Whereas EXP A and B are reported to induce cell wall loosening, EXPansin-like A and B are known only for their gene sequences and their role is unclear (Cosgrove, 2005).

The XTHs genes carry out various functions: wall loosening, wall strengthening, integrating new xyloglucans into the wall and hydrolysing xyloglucans (Rose *et al.*, 2002). The role of the XTH family in the regulation of cell wall loosening versus wall strengthening is not clear but a strengthening function seems more probable (Cosgrove, 2005; Maris *et al*., 2009). Another gene involved in cell wall remodelling is pectin acetylesterase (Vercauteren *et al*., 2002). Furthermore, some genes as RAB GTPASE (*RABA5E* and *RABG3B*) are involved in cell vesicle trafficking; in particular, *RABA5E* plays a role in delivery of new cell wall components needed for cell wall synthesis or cell wall maintenance (Vernoud *et al.,* 2003); PATELLIN 2 (*PATL2)* is involved in membrane trafficking events associated with cell plate formation and expansion. MYO-INOSITOL OXYGENASE 2 (*MIOX2*) is involved in the biosynthesis of new cell components (Kanter *et al.,* 2005; Endres and Tenhaken, 2009), whereas gibberellic acid-stimulated (*GASA1*) belongs to a gene family encoding putative small cell wall proteins.

The upregulation of many genes involved in the cell wall component delivery and

.

strengthening in the transformed lines, and the reduced size of cells in the transformed lines suggest that in this context the EXPANSINS-LIKE-A1 gene could mainly play a role in the cell wall remodelling rather then in the wall loosening process merely finalised to cell expansion. However, four expansin genes (EXPANSIN A5, EXPANSIN A16, EXPANSIN B3 and EXPANSIN A10) were found among the genes downregulated in the transformed lines, although not identified as negative biomarkers. Since the wild type cells showed higher cell enlargement compared to the transformed ones (Guzzo *et al.,* 2005), these four expansin genes downregulated in the transformed lines and more active in the control plants could be involved in the cell wall loosening process necessary for cell expansion.

Among the eighteen genes identified as negative biomarkers, and so downregulated in the transformed lines, both APS kinase 1 and 2 (APK1 and APK2) were found. These genes are involved in the synthesis of activated sulfate and are part of the glucosinolate biosynthesis network (Yatusevich *et al.*, 2009; Mugford *et al*. 2009). Other genes, involved in the glucosinolate biosynthesis pathway, were downregulated in the transformed lines but not identified as negative biomarkers (CYP79B3, CYP79B2, SULFOTRANSFERASE 17, FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 3). Among these genes, two R2R3 MYB transcription factors, MYB29 and MYB34, that control aliphatic and indolic glucosinolate biosynthesis, were also identified. These MYB transcription factors regulate the APS kinase 1 and 2 (Mugford *et al*., 2009; Yatusevich *et al*., 2009).

In the wild type plants, which have lower glucosinolate level, the glucosinolate biosynthetic pathway is switched on, as shown by the higher level of APS kinase genes and MYB transcription factors; this probably indicates that wild type plants were collected during a moment of increased glucosinolate demand, whereas transformed plants had sufficient glucosinolates level (Yatusevich *et al.*, 2009).

.

REFERENCES

Asad (1994) Isolation and characterization of cDNA and genomic clones of *MsENOD40*; Transcripts are detected in meristematic cells of alfalfa. Protoplasma 183: 10-23

Carpita NaM, M. (2000) The cell wall. *In* BB Buchanan, Gruissem, W. and Jones, R.L., ed, Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville, MD, pp 52-108

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J 3: 1-30

Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M (1998) Patterns of ENOD40 gene expression in stem-borne nodules of Sesbania rostrata. Plant Mol Biol 37: 67-76

Cosgrove DJ (1998) Cell wall loosening by expansins. Plant Physiol 118: 333-339

Cosgrove DJ (2005) Growth of the plant cell wall. Nat Rev Mol Cell Biol 6: 850-861

Endres S, Tenhaken R (2009) Myoinositol oxygenase controls the level of myoinositol in Arabidopsis, but does not increase ascorbic acid. Plant Physiol 149: 1042-1049

Fang Y, Hirsch AM (1998) Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol 116: 53-68

Flemetakis E, Kavroulakis N, Quaedvlieg NE, Spaink HP, Dimou M, Roussis A, Katinakis P (2000) Lotus japonicus contains two distinct ENOD40 genes that are expressed in symbiotic, nonsymbiotic, and embryonic tissues. Mol Plant Microbe Interact 13: 987-994

Gultyaev AP, Roussis A (2007) Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res 35: 3144-3152

Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by enod40 in Arabidopsis thaliana. J Exp Bot 56: 507-513

Kanter U, Usadel B, Guerineau F, Li Y, Pauly M, Tenhaken R (2005) The inositol oxygenase gene family of Arabidopsis is involved in the biosynthesis of nucleotide sugar precursors for cell-wall matrix polysaccharides. Planta 221: 243-254

Kouchi H, Takane K, So RB, Ladha JK, Reddy PM (1999) Rice ENOD40: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J 18: 121-

.

129

Kumagai H, Kinoshita E, Ridge RW, Kouchi H (2006) RNAi knock-down of ENOD40s leads to significant suppression of nodule formation in Lotus japonicus. Plant Cell Physiol 47: 1102-1111

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408

Maris A, Suslov D, Fry SC, Verbelen JP, Vissenberg K (2009) Enzymic characterization of two recombinant xyloglucan endotransglucosylase/hydrolase (XTH) proteins of Arabidopsis and their effect on root growth and cell wall extension. J Exp Bot 60: 3959-3972

McQueen-Mason S, Durachko DM, Cosgrove DJ (1992) Two endogenous proteins that induce cell wall extension in plants. Plant Cell 4: 1425-1433

Mugford SG, Yoshimoto N, Reichelt M, Wirtz M, Hill L, Mugford ST, Nakazato Y, Noji M, Takahashi H, Kramell R, Gigolashvili T, Flugge UI, Wasternack C, Gershenzon J, Hell R, Saito K, Kopriva S (2009) Disruption of adenosine-5'-phosphosulfate kinase in Arabidopsis reduces levels of sulfated secondary metabolites. Plant Cell 21: 910-927

Papadopoulou K, Roussis A, Katinakis P (1996) Phaseolus ENOD40 is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. Plant Mol Biol 30: 403-417

Rose JK, Bennett AB (1999) Cooperative disassembly of the cellulose-xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening. Trends Plant Sci 4: 176-183

Rose JK, Braam J, Fry SC, Nishitani K (2002) The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. Plant Cell Physiol 43: 1421-1435

Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116-5121

Varkonyi-Gasic E, White DW (2002) The white clover enod40 gene family. Expression patterns of two types of genes indicate a role in vascular function. Plant Physiol 129: 1107-1118

Vercauteren I, de Almeida Engler J, De Groodt R, Gheysen G (2002) An Arabidopsis thaliana pectin acetylesterase gene is upregulated in nematode feeding sites induced by root-knot and cyst nematodes. Mol Plant Microbe Interact 15: 404-407

Vernoud V, Horton AC, Yang Z, Nielsen E (2003) Analysis of the small GTPase gene superfamily of Arabidopsis. Plant Physiol 131: 1191-1208

.

Vleghels I, Hontelez J, Ribeiro A, Fransz P, Bisseling T, Franssen H (2003) Expression of ENOD40 during tomato plant development. Planta 218: 42-49

Wan X, Hontelez J, Lillo A, Guarnerio C, van de Peut D, Fedorova E, Bisseling T, Franssen H (2007) Medicago truncatula ENOD40-1 and ENOD40-2 are both involved in nodule initiation and bacteroid development. J Exp Bot 58: 2033-2041

Yatusevich R, Mugford SG, Matthewman C, Gigolashvili T, Frerigmann H, Delaney S, Koprivova A, Flugge UI, Kopriva S (2009) Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in Arabidopsis thaliana. Plant J

Zenoni S, Reale L, Tornielli GB, Lanfaloni L, Porceddu A, Ferrarini A, Moretti C, Zamboni A, Speghini A, Ferranti F, Pezzotti M (2004) Downregulation of the Petunia hybrida alpha-expansin gene PhEXP1 reduces the amount of crystalline cellulose in cell walls and leads to phenotypic changes in petal limbs. Plant Cell 16: 295-308

Zhong R, Ye ZH (2007) Regulation of cell wall biosynthesis. Curr Opin Plant Biol 10: 564-572

.

Tab 4.1 The 382 transcripts differentially modulated. FC: fold change (log_2)

ID locus	Definition	FC 1.1	FC 3.1	FC 4.4
At1g01720	ATAF1	-0.910	-0.806	-0.837
At1g02205	ECERIFERUM 1 (CER1)	-1.112	-0.890	-0.518
At1g02640	BETA-XYLOSIDASE 2 (BXL2)	1.950	0.712	1.066
At1g02850	BETA GLUCOSIDASE 11 (BGLU11)	-0.625	0.201	-0.019
At1g03870	FASCICLIN-LIKE ARABINOOGALACTAN 9 (FLA9)	1.914	1.485	1.311
At1g04040	acid phosphatase class B family protein	0.822	0.969	-0.080
At1g04800	glycine-rich protein	-0.963	-1.478	-0.798
At1g05810	RAB GTPASE HOMOLOG A5E (RABA5E)	0.857	0.685	0.654
At1g06080	DELTA 9 DESATURASE 1 (ADS1)	0.147	-0.688	-0.361
At1g06160	OCTADECANOID-RESPONSIVE AP2/ERF 59 (ORA59)	0.357	1.454	0.816
At1g06830	glutaredoxin family protein	-1.112	-0.729	-0.818
At1g07430	protein phosphatase 2C, putative / PP2C, putative	-1.171	-0.539	-0.258
At1g07570	APK1A	-0.114	-0.930	-0.472
At1g08930	EARLY RESPONSE TO DEHYDRATION 6 (ERD6)	0.949	0.754	-0.114
At1g09350	Arabidopsis thaliana galactinol synthase 3 (AtGolS3)	-0.129	-0.733	-1.076
At1g09500	cinnamyl-alcohol dehydrogenase family / CAD family	-0.274	0.991	-0.067
At1g10020	unknown protein	0.920	0.839	0.933
At1g10550	XTH33	1.751	2.100	2.082
At1g10560	PLANT U-BOX 18 (PUB18)	-0.116	-0.798	-0.743
At1g11260	SUGAR TRANSPORTER 1 (STP1)	1.032	0.665	0.967
At1g12080	unknown protein	0.445	-0.956	-0.912
At1g12080	unknown protein	0.408	-0.789	-0.657
At1g12710	Phloem protein 2-A12 (AtPP2-A12)	0.400	0.950	0.644
At1g13080	CYTOCHROME P450 71B2 (CYP71B2)	-1.000	-0.796	-0.432
At1g13110	CYP71B7	0.358	0.889	0.277
At1g13470	unknown protein	0.522	0.419	-0.542
At1g14250	nucleoside phosphatase family protein / GDA1/CD39 family protein	-1.581	-0.450	-0.541
At1g15125	S-adenosylmethionine-dependent methyltransferase/ methyltransferase	1.313	1.903	1.827
At1g18570	MYB DOMAIN PROTEIN 51 (MYB51)	1.039	0.785	0.150
At1g18590	SULFOTRANSFERASE 17 (SOT17)	-1.102	-0.951	-0.314
At1g19320	pathogenesis-related thaumatin family protein	-0.483	-1.214	-1.105
At1g19530	unknown protein	0.906	0.151	0.120
At1g19670	CORONATINE-INDUCED PROTEIN 1 (ATCLH1)	-1.748	-0.450	-1.077

.

At1g20070	unknown protein	-0.530	-1.378	-1.237
At1g20840	TONOPLAST MONOSACCHARIDE TRANSPORTER1 (TMT1)	0.500	0.709	0.036
At1g21050	unknown protein	0.706	0.537	0.026
At1g22530	PATELLIN 2 (PATL2)	1.777	2.111	1.382
At1g22570	proton-dependent oligopeptide transport (POT) family protein	0.689	1.424	0.392
At1g22590	AGAMOUS-LIKE 87 (AGL87)	-0.553	-0.835	-0.847
At1g22690	gibberellin-responsive protein, putative	0.807	0.037	-0.248
At1g22740	RABG3B	1.187	0.846	0.962
At1g23480	CELLULOSE SYNTHASE-LIKE A3 (ATCSLA03)	0.890	1.016	0.914
At1g24020	MLP-LIKE PROTEIN 423 (MLP423)	-1.321	-1.239	-0.828
At1g24100	UDP-glucosyl transferase 74B1 (UGT74B1)	-0.942	-1.042	-0.661
At1g24280	GLUCOSE-6-PHOSPHATE DEHYDROGENASE 3 (G6PD3)	-1.145	-0.908	-0.872
At1g26770	ARABIDOPSIS THALIANA EXPANSIN A 10 (ATEXPA10)	-1.192	-1.123	-1.402
At1g28330	DORMANCY-ASSOCIATED PROTEIN-LIKE 1 (DYL1)	1.035	0.730	0.788
At1g28330	DORMANCY-ASSOCIATED PROTEIN-LIKE 1 (DYL1)	1.039	0.887	0.638
At1g30380	photosystem I subunit K (PSAK)	-1.027	-0.744	-0.490
At1g30720	FAD-binding domain-containing protein	-0.103	-1.086	-0.058
At1g32460	unknown protein	0.929	0.344	-0.622
At1g35140	PHOSPHATE-INDUCED 1 (PHI-1)	1.535	1.873	0.996
At1g35350	LOCATED IN: integral to membrane	0.772	1.167	0.791
At1g50460	HEXOKINASE-LIKE 1 (HKL1)	0.943	0.775	0.886
At1g51140	basic helix-loop-helix (bHLH) family protein	-1.089	-0.663	-0.330
At1g52000	jacalin lectin family protein	-1.678	-0.576	-0.618
At1g52030	MYROSINASE-BINDING PROTEIN 2 (MBP2)	-1.031	0.233	0.104
At1g52200	unknown protein	0.709	0.472	-0.789
At1g52400	BETA GLUCOSIDASE 18 (BGLU18)	-0.941	0.053	-0.642
At1g52690	late embryogenesis abundant protein, putative / LEA protein, putative	0.025	1.406	1.028
At1g54740	FUNCTIONS IN: molecular_function unknown	1.038	0.217	0.822
At1g54820	protein kinase family protein	-1.318	-0.404	-0.426
At1g55330	AGP21	1.133	0.589	1.132
At1g55850	ATCSLE1	0.179	1.201	0.866
At1g56430	NICOTIANAMINE SYNTHASE 4 (NAS4)	-1.000	-0.609	-0.766
At1g56580	unknown protein	-0.684	-1.019	-0.481
At1g56600	Arabidopsis thaliana galactinol synthase 2 (AtGolS2)	-1.325	-1.083	-1.048
At1g62480	vacuolar calcium-binding protein-related	0.711	-0.402	-0.209

.

At1g62560	FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 3	-1.122	-1.229	-0.899
At1g62570	FLAVIN-MONOOXYGENASE GLUCOSINOLATE S-OXYGENASE 4	-1.007	-0.491	-0.461
At1g64200	VACUOLAR H+-ATPASE SUBUNIT E ISOFORM 3 (VHA-E3)	-0.803	-0.711	-0.426
At1g65390	ARABIDOPSIS THALIANA PHLOEM PROTEIN 2 A5 (ATPP2A5)	0.488	-0.682	-0.422
At1g65840	ARABIDOPSIS THALIANA POLYAMINE OXIDASE 4 (ATPAO4)	0.041	-0.107	-1.116
At1g65900	unknown protein	-0.931	-0.568	-0.307
At1g66350	RGA-LIKE 1 (RGL1)	-0.241	-0.639	-0.673
At1g66700	PXMT1	-0.484	1.314	0.501
At1g66940	protein kinase-related	0.773	0.715	-0.023
At1g67810	SULFUR E 2 (SUFE2)	-0.235	1.212	-0.055
At1g68600	unknown protein	-1.404	-0.548	-0.735
At1g70260	nodulin MtN21 family protein	-1.176	-1.725	-0.578
At1g70810	C2 domain-containing protein	-0.788	0.211	0.085
At1g72430	auxin-responsive protein-related	1.225	0.978	1.022
At1g73260	trypsin and protease inhibitor family protein / Kunitz family protein	-0.664	1.735	-0.297
At1g73325	trypsin and protease inhibitor /Kunitz family protein	-0.856	-0.837	-0.682
At1g74090	DESULFO-GLUCOSINOLATE SULFOTRANSFERASE 18 (SOT18)	-1.457	-1.148	-0.605
At1g75040	PATHOGENESIS-RELATED GENE 5 (PR5)	0.232	1.108	-0.385
At1g75750	GAST1 PROTEIN HOMOLOG 1 (GASA1)	2.428	1.316	1.266
At1g76090	STEROL METHYLTRANSFERASE 3 (SMT3)	1.095	1.140	1.253
At1g76160	SKU5 Similar 5 (sks5)	1.130	0.842	0.406
At1g76790	O-methyltransferase family 2 protein	-1.536	-0.389	-0.770
At1g76990	ACR3	0.814	0.961	0.203
At1g78270	UDP-glucosyl transferase 85A4 (AtUGT85A4)	0.693	1.286	0.720
At1g78820	curculin-like (mannose-binding) lectin family protein /PAN domain-containing protein	-0.977	-0.387	-0.226
At1g80310	sulfate transmembrane transporter	-0.892	-0.038	-0.048
At1g80760	NOD26-LIKE INTRINSIC PROTEIN 6	-0.775	-1.069	-0.545
At2g03310	unknown protein	-0.244	0.568	0.255
At2g03980	GDSL-motif lipase/hydrolase family protein	-0.905	-0.258	-0.804
At2g05380	GLYCINE-RICH PROTEIN 3 SHORT ISOFORM (GRP3S)	1.141	0.445	0.207

.

At2g14247	Expressed protein	-0.462	1.167	0.533
At2g14560	LATE UPREGULATED IN RESPONSE TO HYALOPERONOSPORAPARASITICA (LURP1)	0.909	0.841	0.093
At2g14610	PATHOGENESIS-RELATED GENE 1 (PR1)	-0.447	0.993	-1.062
At2g14750	APS KINASE 1 (APK 1)	-0.949	-1.000	-0.968
At2g15050	LTP	-1.447	-0.527	-0.012
At2g15620	NITRITE REDUCTASE 1 (NIR1)	-1.340	-1.118	-0.607
At2g15780	glycine-rich protein	-1.009	-0.893	-0.416
At2g16660	nodulin family protein	0.845	1.330	1.440
At2g16890	UDP-glucoronosyl/UDP-glucosyl transferase family protein	-0.958	-0.685	-0.764
At2g19800	MYO-INOSITOL OXYGENASE 2 (MIOX2)	2.240	1.235	0.454
At2g20142	transmembrane receptor	1.086	0.620	0.148
At2g20340	tyrosine decarboxylase, putative	-1.348	-0.750	-0.748
At2g20880	AP2 domain-containing transcription factor, putative	-0.103	-1.237	-0.611
At2g22330	CYP79B3	-1.130	-0.435	-0.897
At2g23000	serine carboxypeptidase-like 10 (scpl10)	-1.548	-0.910	-0.491
At2g23130	ARABINOGALACTAN PROTEIN 17 (AGP17)	0.619	1.045	0.996
At2g23600	ACETONE-CYANOHYDRIN LYASE (ACL)	0.886	0.992	0.159
At2g26010	plant defensin 1.3 (PDF1.3)	-0.783	3.553	1.838
At2g26020	plant defensin 1.2b (PDF1.2b)	-0.245	2.214	1.749
At2g26560	PHOSPHOLIPASE A 2A (PLA2A)	0.810	1.547	0.727
At2g27402	FUNCTIONS IN: molecular_function unknown	-0.553	-1.616	-1.405
At2g28190	COPPER/ZINC SUPEROXIDE DISMUTASE 2 (CSD2)	0.665	-0.261	0.418
At2g29340	short-chain dehydrogenase/reductase (SDR) family protein	-0.689	-0.178	0.310
At2g29450	GLUTATHIONE S-TRANSFERASE TAU 5 (ATGSTU5)	-1.488	-1.121	-1.273
At2g29460	GLUTATHIONE S-TRANSFERASE TAU 4 (ATGSTU4)	-1.017	0.089	-0.645
At2g30010	INVOLVED IN: biological_process unknown	1.063	0.856	0.151
At2g30600	BTB/POZ domain-containing protein	1.127	0.959	0.550
At2g30770	CYP71A13	-0.230	1.085	0.015
At2g30930	unknown protein	1.650	1.923	1.177
At2g31120	At2g31120.1//	0.623	0.923	0.611
At2g31160	LIGHT SENSITIVE HYPOCOTYLS 3 (LSH3)	-0.962	-1.104	-0.692
At2g31790	UDP-glucoronosyl/UDP-glucosyl transferase family protein	-1.039	-0.804	-0.597
At2g31810	acetolactate synthase small subunit, putative	0.967	0.193	0.133
At2g32100	ARABIDOPSIS THALIANA OVATE FAMILY PROTEIN 16 (OFP16)	-1.227	-0.002	-0.391

.

At2g33380	RESPONSIVE TO DESSICATION 20 (RD20)	-1.241	-0.709	-1.060
At2g33830	dormancy/auxin associated family protein	0.738	0.744	0.732
At2g34510	FUNCTIONS IN: molecular_function unknown	0.871	1.427	0.840
At2g34620	mitochondrial transcription termination factor-related	-1.064	-0.941	-0.609
At2g34770	FATTY ACID HYDROXYLASE 1 (FAH1)	0.812	0.838	0.843
At2g36130	peptidyl-prolyl cis-trans isomerase, putative / cyclophilin, putative / rotamase, putative	1.741	1.786	2.139
At2g37240	INVOLVED IN: biological_process unknown	-0.988	-0.270	-0.280
At2g37540	short-chain dehydrogenase/reductase (SDR) family protein	0.636	0.921	0.805
At2g37870	lipid transfer protein (LTP) family protein/seed storage	-0.492	0.895	0.553
At2g38180	GDSL-motif lipase/hydrolase family protein	-0.318	0.595	-0.081
At2g38870	protease inhibitor, putative	-0.813	-0.949	0.099
At2g39030	GCN5-related N-acetyltransferase (GNAT) family protein	-2.196	0.862	0.201
At2g39330	JACALIN-RELATED LECTIN 23 (JAL23)	-1.406	-0.699	-0.827
At2g39800	DELTA1-PYRROLINE-5-CARBOXYLATE SYNTHASE 1 (P5CS1)	-1.200	-0.418	0.132
At2g40080	EARLY FLOWERING 4 (ELF4)	0.491	0.987	0.226
At2g40330	Bet v I allergen family protein	2.078	1.770	0.739
At2g41100	TOUCH 3 (TCH3)	1.091	0.109	0.309
At2g41990	unknown protein	1.152	0.507	0.372
At2g42040	FUNCTIONS IN: molecular_function unknown	0.795	0.706	0.791
At2g43100	aconitase C-terminal domain-containing protein	-1.291	-1.027	-0.666
At2g43570	chitinase, putative	-0.232	1.045	0.673
At2g46370	JASMONATE RESISTANT 1 (JAR1)	-1.012	-0.941	-1.493
At2g46680	ARABIDOPSIS THALIANA HOMEOBOX 7 (ATHB-7)	-1.084	-0.021	-0.010
At2g46830	CIRCADIAN CLOCK ASSOCIATED 1 (CCA1)	-0.008	0.068	0.581
At2g47180	Arabidopsis thaliana galactinol synthase 1 (AtGolS1)	-0.831	-0.729	-0.298
At2g48030	endonuclease/exonuclease/phosphatase family protein	0.763	0.779	0.172
At3g01060	unknown protein	-1.068	-0.508	-0.507
At3g01290	band 7 family protein	1.809	1.361	0.272
At3g01500	CARBONIC ANHYDRASE 1 (CA1)	-1.573	-1.585	-0.333
At3g02410	INVOLVED IN: biological_process unknown	2.273	1.801	1.626
At3g02480	ABA-responsive protein-related	0.461	2.097	1.050
At3g02570	MATERNAL EFFECT EMBRYO ARREST 31 (MEE31)	-0.216	-0.750	-1.208
At3g03190	GLUTATHIONE S-TRANSFERASE F11 (ATGSTF11)	-1.237	-1.250	-0.485
At3g03780	AtMS2	-0.916	-0.752	-0.581

.

At3g04210	disease resistance protein (TIR-NBS class), putative	0.589	0.988	0.225
At3g04630	WDL1	-0.247	-1.158	-0.716
At3g04720	PATHOGENESIS-RELATED 4 (PR4)	-0.241	1.922	1.065
At3g05180	GDSL-motif lipase/hydrolase family protein	-0.921	-0.376	-0.307
At3g05730	LOCATED IN: endomembrane system	-0.401	-1.104	-0.551
At3g08770	LTP6	-0.810	-1.091	-0.537
At3g08860	alanine-glyoxylate aminotransferase, putative /beta-alanine-pyruvate aminotransferase, putative / AGT, putative	-0.993	-0.041	0.153
At3g12500	ARABIDOPSIS THALIANA BASIC CHITINASE (ATHCHIB)	-0.980	2.464	1.185
At3g13470	chaperonin, putative	-1.045	-1.160	-1.334
At3g14230	RAP2.2	0.922	0.632	0.369
At3g15450	unknown protein	1.291	0.687	0.992
At3g15540	INDOLE-3-ACETIC ACID INDUCIBLE 19 (IAA19)	0.707	-0.364	-0.374
At3g15630	unknown protein	1.596	0.922	0.985
At3g16420	PYK10-BINDING PROTEIN 1 (PBP1)	-0.343	0.605	0.301
At3g17070	peroxidase, putative	-0.811	-0.161	0.030
At3g17390	METHIONINE OVER-ACCUMULATOR 3 (MTO3)	0.672	0.892	0.562
At3g20030	F-box family protein	0.166	0.784	1.932
At3g21320	FUNCTIONS IN: molecular_function unknown	0.871	-0.093	-0.213
At3g21670	nitrate transporter (NTP3)	-1.124	-0.414	-0.772
At3g22060	receptor protein kinase-related	-1.238	-1.157	-0.977
At3g22600	lipid transfer protein (LTP) family protein/seed storage	-0.468	1.058	0.389
At3g22740	HMT3	-0.745	-1.251	-0.361
At3g23730	xyloglucosyl transferase, putative /xyloglucan endotransglycosylase, putative / endo-xyloglucan transferase, putative	0.862	0.364	1.095
At3g25760	ALLENE OXIDE CYCLASE 1 (AOC1)	-1.843	-0.662	-1.747
At3g25770	ALLENE OXIDE CYCLASE 2 (AOC2)	-1.216	-0.225	-0.973
At3g26740	CCR-LIKE (CCL)	-0.123	0.809	0.442
At3g28220	meprin and TRAF homology domain-containing protein / MATH domain-containing protein	-2.875	0.508	-1.000
At3g29030	EXPANSIN A5 (EXPA5)	-0.870	-0.683	-0.705
At3g29810	COBRA-LIKE PROTEIN 2 PRECURSOR (COBL2)	0.998	0.864	0.603
At3g30775	EARLY RESPONSIVE TO DEHYDRATION 5 (ERD5)	2.072	1.362	0.541
At3g42628	phosphoenolpyruvate carboxylase-related / PEP carboxylase-related	-1.033	-0.261	-0.031
At3g44430	unknown protein	-0.853	-0.674	-0.700
At3g44860	farnesoic acid carboxyl-O-methyltransferase (FAMT)	-1.247	-1.091	-1.360

.

At3g44990	XYLOGLUCAN ENDO-TRANSGLYCOSYLASE-RELATED 8 (XTR8)	-1.576	-0.528	-0.693
At3g45140	LIPOXYGENASE 2 (LOX2)	-1.718	-0.678	-0.628
At3g45970	ARABIDOPSIS THALIANA EXPANSIN-LIKE A1 (ATEXLA1)	1.902	2.042	2.021
At3g47340	GLUTAMINE-DEPENDENT ASPARAGINE SYNTHASE 1 (ASN1)	2.180	1.657	1.238
At3g47420	glycerol-3-phosphate transporter, putative/glycerol 3-phosphate permease, putative	-0.788	0.041	-0.886
At3g48460	GDSL-motif lipase/hydrolase family protein	-0.955	-0.794	-0.402
At3g49620	DARK INDUCIBLE 11 (DIN11)	-0.876	1.490	0.883
At3g50440	METHYL ESTERASE 10 (MES10)	-0.818	-0.023	0.186
At3g50770	calmodulin-related protein, putative	0.636	1.342	0.353
At3g51450	strictosidine synthase family protein	-1.144	-0.907	-1.464
At3g52720	ALPHA CARBONIC ANHYDRASE 1 (ACA1)	-1.429	-1.317	-0.921
At3g53960	proton-dependent oligopeptide transport (POT) family protein	-0.949	-0.541	-0.778
At3g54390	transcription factor	-0.869	-0.836	-0.235
At3g54600	DJ-1 family protein	-1.226	-0.964	-0.405
At3g55500	ARABIDOPSIS THALIANA EXPANSIN A16 (ATEXPA16)	-1.350	-0.955	0.083
At3g57520	Arabidopsis thaliana seed imbibition 2 (AtSIP2)	1.011	0.966	0.668
At3g57520	Arabidopsis thaliana seed imbibition 2 (AtSIP2)	1.353	1.473	0.964
At3g59350	serine/threonine protein kinase, putative	1.374	1.272	0.979
At3g59350	serine/threonine protein kinase, putative	1.387	0.997	0.803
At3g59400	GUN4	-0.966	-0.463	-0.420
At3g59900	AUXIN-REGULATED GENE INVOLVED IN ORGAN SIZE (ARGOS)	0.777	-0.074	-0.101
At3g60140	DARK INDUCIBLE 2 (DIN2)	-0.592	1.183	-0.297
At3g60320	DNA binding	0.620	0.800	0.245
At3g61990	O-methyltransferase family 3 protein	0.255	1.223	0.438
At3g62110	glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein	0.422	0.047	-0.615
At3g62150	P-GLYCOPROTEIN 21 (PGP21)	1.577	1.481	0.779
At4g00300	fringe-related protein	0.668	0.243	-0.302
At4g01750	rhamnogalacturonan xylosyltransferase 2 (RGXT2)	0.897	1.115	0.059
At4g02330	ATPMEPCRB	-0.379	-0.983	-0.135
At4g02850	phenazine biosynthesis PhzC/PhzF family protein	-1.310	-1.342	-0.987
At4g04610	APS REDUCTASE 1 (APR1)	-1.328	-1.002	-1.100

.

At4g08870	arginase, putative	-2.079	-0.279	-0.890
At4g09350	DNAJ heat shock N-terminal domain-containing protein	-0.833	-0.377	-0.535
At4g10120	ATSPS4F	-0.786	0.180	-0.214
At4g10770	OLIGOPEPTIDE TRANSPORTER 7 (OPT7)	-0.896	-0.362	0.097
At4g11320	cysteine proteinase, putative	-1.579	-0.440	-0.861
At4g11650	osmotin 34 (ATOSM34)	-1.498	1.858	0.346
At4g12030	bile acid:sodium symporter family protein	-0.959	-1.060	-0.511
At4g12030	bile acid:sodium symporter family protein	-1.610	-1.334	-0.905
At4g12480	pEARLI 1	0.163	0.760	0.929
At4g13180	short-chain dehydrogenase/reductase (SDR) family protein	-0.918	-0.669	-0.987
At4g13340	leucine-rich repeat family protein / extensin family protein	1.220	1.294	0.968
At4g14020	rapid alkalinization factor (RALF) family protein	-0.812	0.070	-0.491
At4g14130	XYLOGLUCAN ENDOTRANSGLYCOSYLASE 7 (XTR7)	3.766	2.161	1.685
At4g15210	ARABIDOPSIS THALIANA BETA-AMYLASE (ATBETA-AMY)	-1.525	-0.822	-1.250
At4g15210	BETA-AMYLASE 5 (BAM5)	-1.872	-1.005	-1.155
At4g15440	HYDROPEROXIDE LYASE 1 (HPL1)	-2.146	-1.057	-0.783
At4g16260	catalytic/ cation binding/ hydrolyzing O-glycosyl compounds	-0.041	1.937	0.837
At4g16563	aspartyl protease family protein	1.030	1.112	0.868
At4g16590	ATCSLA01	-1.140	-0.829	-0.907
At4g17870	FUNCTIONS IN: molecular_function unknown	1.269	0.871	0.381
At4g18440	adenylosuccinate lyase, putative / adenylosuccinase, putative	-1.356	-1.066	-1.272
At4g19420	pectinacetylesterase family protein	0.715	1.060	0.788
At4g19420	pectinacetylesterase family protein	0.464	0.886	0.757
At4g19430	unknown protein	2.080	1.426	0.808
At4g20820	FAD-binding domain-containing protein	-0.713	-0.475	0.368
At4g20830	FAD-binding domain-containing protein	-0.426	-1.040	-1.080
At4g21280	PHOTOSYSTEM II SUBUNIT QA (PSBQA)	-1.010	-0.995	-0.425
At4g21870	26.5 kDa class P-related heat shock protein (HSP26.5-P)	2.292	2.672	0.634
At4g21910	MATE efflux family protein	-1.332	-0.235	-0.681
At4g22200	ARABIDOPSIS POTASSIUM TRANSPORT 2/3 (AKT2/3)	-1.093	-0.802	-0.963
At4g22753	STEROL 4-ALPHA METHYL OXIDASE 1-3 (SMO1-3)	-1.007	-0.844	-1.074
At4g23220	kinase	-0.353	-0.997	-0.641
At4g23400	PLASMA MEMBRANE INTRINSIC PROTEIN 1	0.973	0.306	0.771
At4g23600	CORONATINE INDUCED 1 (CORI3)	-1.703	-0.332	-0.954
At4g23820	glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein	1.107	0.678	0.599

.

At4g24780	pectate lyase family protein	-1.259	-1.090	-0.749
At4g25100	FE SUPEROXIDE DISMUTASE 1 (FSD1)	-1.238	-0.563	-0.704
At4g26530	fructose-bisphosphate aldolase, putative	-1.194	-0.414	-0.100
At4g28250	ARABIDOPSIS THALIANA EXPANSIN B3 (ATEXPB3)	-1.084	-1.245	-0.340
At4g29030	glycine-rich protein	-0.976	-1.067	-0.643
At4g29150	IQ-domain 25 (IQD25)	-0.567	-1.423	-3.461
At4g29700	type I phosphodiesterase/nucleotide pyrophosphatase family protein	-1.188	-1.069	-0.923
At4g29710	phosphodiesterase/nucleotide pyrophosphatase-related	-0.909	-0.712	-0.547
At4g29905	unknown protein	-1.009	-0.809	0.175
At4g30270	meristem-5 (MERI5B)	1.105	1.056	1.124
At4g30280	XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE 18 (XTH18)	1.272	0.688	0.926
At4g30460	glycine-rich protein	-1.004	-0.524	-0.166
At4g32770	VITAMIN E DEFICIENT 1 (VTE1)	-0.999	-0.627	-0.506
At4g33550	lipid binding	-0.743	0.508	0.503
At4g33660	unknown protein	-1.075	-0.695	-0.138
At4g33666	unknown protein	-0.593	-0.514	0.183
At4g33720	pathogenesis-related protein, putative	0.313	1.709	0.926
At4g33980	unknown protein	1.040	1.470	0.795
At4g35720	unknown protein	0.365	-0.658	-0.060
At4g36850	INVOLVED IN: biological_process unknown	0.717	-0.307	-0.303
At4g37770	ACS8	0.987	0.068	-0.319
At4g38410	dehydrin, putative	-0.183	-1.124	-0.836
At4g39260	cold, circadian rhythm, and RNA binding 1 (CCR1)	0.373	0.820	-0.139
At4g39260	cold, circadian rhythm, and RNA binding 1 (CCR1)	0.503	0.353	-1.211
At4g39510	CYP96A12	-1.123	-0.675	-0.516
At4g39800	MYO-INOSITOL-1-PHOSTPATE SYNTHASE 1 (MIPS1)	-0.888	-0.618	-0.553
At4g39830	L-ascorbate oxidase, putative	0.312	-0.317	-0.640
At4g39940	APS-kinase 2 (APK2)	-1.159	-1.328	-1.549
At4g39950	CYP79B2	-0.979	-0.527	-1.069
At5g01220	sulfoquinovosyldiacylglycerol 2 (SQD2)	-2.621	-2.463	-2.780
At5g01750	unknown protein	-1.792	-1.634	-1.587
At5g02150	binding	3.473	3.555	3.644
At5g02190	PROMOTION OF CELL SURVIVAL 1 (PCS1)	0.952	1.021	1.053
At5g02450	60S ribosomal protein L36 (RPL36C)	-1.224	-1.256	-1.052
At5g02500	HEAT SHOCK COGNATE PROTEIN 70-1 (HSC70-1)	-1.058	-1.090	-0.886

.

At5g02540	short-chain dehydrogenase/reductase (SDR) family protein	0.246	-0.873	-0.789
At5g02580	unknown protein	0.770	0.508	-0.042
At5g02940	LOCATED IN: chloroplast, chloroplast envelope	-1.324	-0.593	-1.010
At5g03120	unknown protein	1.444	2.016	1.417
At5g04950	NICOTIANAMINE SYNTHASE 1 (NAS1)	-1.092	-0.613	-0.147
At5g05270	chalcone-flavanone isomerase family protein	-1.192	-0.847	-0.440
At5g05440	unknown protein	1.105	0.993	0.778
At5g05600	oxidoreductase, 2OG-Fe(II) oxygenase family protein	-1.246	-0.244	-1.156
At5g06510	NUCLEAR FACTOR Y, SUBUNIT A10 (NF-YA10)	0.933	1.340	1.421
At5g07010	SULFOTRANSFERASE 2A (ST2A)	-0.093	-0.094	-1.427
At5g07580	DNA binding / transcription factor	0.745	1.101	0.847
At5g07690	ARABIDOPSIS THALIANA MYB DOMAIN PROTEIN 29 (ATMYB29)	-1.448	-1.212	-0.940
At5g08050	FUNCTIONS IN: molecular_function unknown	-0.939	-1.091	-0.621
At5g12860	dicarboxylate transporter 1 (DiT1)	-1.134	-0.821	-0.930
At5g13170	SENESCENCE-ASSOCIATED PROTEIN 29 (SAG29)	-0.283	0.900	0.226
At5g13710	STEROL METHYLTRANSFERASE 1 (SMT1)	1.039	0.545	0.358
At5g13930	TRANSPARENT TESTA 4 (TT4)	-1.317	-0.170	-0.470
At5g14060	CARAB-AK-LYS	-0.948	-0.472	-0.489
At5g14090	unknown protein	0.812	0.563	0.448
At5g14550	unknown protein	-1.445	0.233	-0.225
At5g14640	SHAGGY-LIKE KINASE 13 (SK13)	-0.859	-0.597	-0.375
At5g14740	CARBONIC ANHYDRASE 2 (CA2)	-0.617	-0.786	0.058
At5g14920	gibberellin-regulated family protein	1.713	1.458	0.022
At5g15500	ankyrin repeat family protein	-0.796	0.479	0.335
At5g15950	adenosylmethionine decarboxylase family protein	-0.706	-0.786	-0.995
At5g17220	ATH GLUTATHIONE S-TRANSFERASE PHI 12 (ATGSTF12)	-2.261	-1.031	-1.208
At5g18840	sugar transporter, putative	1.666	2.205	0.722
At5g19470	Arabidopsis thaliana Nudix hydrolase homolog 24 (atnudt24)	-1.401	-0.057	-1.157
At5g20250	DARK INDUCIBLE 10 (DIN10)	1.927	1.393	62.000
At5g22650	HISTONE DEACETYLASE 2B (HD2B)	-0.155	-0.347	-1.001
At5g22920	zinc finger (C3HC4-type RING finger) family protein	1.213	0.994	0.488
At5g23060	Calcium sensing receptor (CaS)	-0.913	-0.218	-0.307
At5g23240	DNAJ heat shock N-terminal domain-containing protein	0.514	1.293	0.248
At5g23410	FUNCTIONS IN: molecular_function unknown	0.655	0.881	0.288
At5g23820	MD-2-related lipid recognition domain-containing protein	-1.207	-0.577	-0.651

.

At5g23940	embryo defective 3009 (EMB3009)	-0.594	-0.802	0.247
At5g24580	copper-binding family protein	0.533	-0.475	-0.294
At5g24770	VEGETATIVE STORAGE PROTEIN 2 (VSP2)	-2.293	-0.074	-1.590
At5g25610	RD22	-0.653	0.394	0.460
At5g27350	SFP1	-0.981	-0.026	-0.260
At5g35480	unknown protein	-0.870	0.103	-0.140
At5g35735	auxin-responsive family protein	0.670	0.795	-0.314
At5g38120	4-coumarate-CoA ligase family protein / 4-coumaroyl-CoA synthase family protein	-1.225	-0.794	-0.969
At5g40240	nodulin MtN21 family protein	0.206	0.915	-0.400
At5g40850	UROPHORPHYRIN METHYLASE 1 (UPM1)	-1.131	-1.304	-0.595
At5g41900	hydrolase, alpha/beta fold family protein	-0.123	-1.049	-0.320
At5g42650	ALLENE OXIDE SYNTHASE (AOS)	-1.333	-0.742	-1.623
At5g42900	unknown protein	1.158	1.208	0.835
At5g43170	ARABIDOPSIS ZINC-FINGER PROTEIN 3 (AZF3)	0.614	0.857	0.199
At5g43580	serine-type endopeptidase inhibitor	-0.497	0.681	-0.284
At5g44420	PDF1.2	-0.578	2.365	1.875
At5g44430	plant defensin 1.2c (PDF1.2c)	0.171	3.643	3.400
At5g44720	molybdenum cofactor sulfurase family protein	-0.789	-0.843	-0.965
At5g45640	subtilase family protein	-0.866	-0.139	-0.866
At5g45650	subtilase family protein	-0.832	-0.067	-0.355
At5g48490	lipid transfer protein (LTP) family/ proteinseed storage	-0.138	-0.344	1.052
At5g48850	SULPHUR DEFICIENCY-INDUCED 1 (ATSDI1)	-0.734	-0.177	0.529
At5g51550	EXORDIUM LIKE 3 (EXL3)	0.977	0.982	1.120
At5g51720	unknown protein	-0.589	-1.055	0.000
At5g54280	ARABIDOPSIS THALIANA MYOSIN 2 (ATM2)	-1.165	-0.801	-0.935
At5g54585	unknown protein	-1.200	-0.714	0.125
At5g55570	unknown protein	-0.915	-0.784	-0.694
At5g57560	Touch 4 (TCH4)	2.069	1.594	1.321
At5g57760	unknown protein	0.826	0.211	-0.336
At5g59310	LIPID TRANSFER PROTEIN 4 (LTP4)	-0.703	0.855	0.377
At5g59320	LIPID TRANSFER PROTEIN 3 (LTP3)	0.038	1.883	1.602
At5g59330	lipid binding	0.025	2.116	1.854
At5g59480	haloacid dehalogenase-like hydrolase family protein	-0.724	-1.050	-1.151
At5g60890	MYB DOMAIN PROTEIN 34 (MYB34)	-1.438	-1.127	-0.974
At5g61160	anthocyanin 5-aromatic acyltransferase 1 (AACT1)	0.324	2.170	0.932

.

At5g61590	AP2 domain-containing transcription factor family protein	0.878	1.477	0.404
At5g61600	ethylene-responsive element-binding family protein	1.304	1.574	0.350
At5g62360	unknown protein	0.387	1.148	-0.101
At5g62430	CYCLING DOF FACTOR 1 (CDF1)	-0.755	-0.794	-0.253
At5g64080	protease inhibitor/seed storage/ lipid transfer protein (LTP)	-1.063	-0.704	-0.584
At5g64120	peroxidase, putative	-1.161	-1.377	-0.639
At5g64770	unknown protein	-1.290	-0.851	-1.022
At5g65010	ASPARAGINE SYNTHETASE 2 (ASN2)	-1.262	-0.715	0.000
At5g65390	AGP7	0.837	0.798	1.217
At5g65730	xyloglucan:xyloglucosyl transferase,putative/xyloglucan endotransglycosylase, putative/endo-xyloglucan transferase, putative	1.405	0.744	1.216
At5g65890	ACT Domain Repeat 1 (ACR1)	-0.871	-0.444	-0.444
At5g66400	RESPONSIVE TO ABA 18 (RAB18)	-0.969	0.401	0.042

.

Table 4.2: Positive and negative biomarkers.

ID locus	positive biomarkers	abb.
At1g05810	RAB GTPASE HOMOLOG A5E (RABA5E)	RABA5E
At1g10550	xyloglucan:xyloglucosyl transferase 33	XTH33
At1g22530	PATELLIN 2	PATL2
At1g22740	RAB GTPASE HOMOLOG G3B	RABG3B
At1g75750	GAST1 PROTEIN HOMOLOG 1	GASA1
At1g76090	STEROL METHYLTRANSFERASE 3	SMT3
At2g19800	MYO-INOSITOL OXYGENASE 2	MIOX2
At2g30600	BTB/POZ domain-containing protein	
At2g33830	dormancy/auxin associated family protein	
At2g36130	peptidyl-prolyl cis-trans isomerase, putative / cyclophilin, putative	
At2g40330	Bet v I allergen family protein	
At3g15630	unknown protein	
At3g45970	EXPANSIN-LIKE A1	ATEXLA1
At3g59350	serine/threonine protein kinase, putative	
At4g14130	xyloglucan endotransglycosylase 7	XTR7
At4g16563	aspartyl protease family protein	
At4g19420	pectinacetylesterase family protein	
At5g02150	binding	
At5g03120	unknown protein	
At5g14090	unknown protein	
At5g18840	sugar transporter, putative	
At5g51550	EXORDIUM LIKE 3	EXL3
At5g65730	xyloglucan:xyloglucosyl transferase, putative	

ID locus	negative biomarkers	abb.
At1g56580	unknown protein	
At2g14750	APS KINASE 1	APK 1
At2g15620	NITRITE REDUCTASE 1	NIR1
At2g16890	UDP-glucoronosyl/UDP-glucosyl transferase family protein	
At3g52720	ALPHA CARBONIC ANHYDRASE 1	ACA1
At4g02850	phenazine biosynthesis PhzC/PhzF family protein	
At4g22753	STEROL 4-ALPHA METHYL OXIDASE 1-3	SMO1-3
At4g29700	type I phosphodiesterase/nucleotide pyrophosphatase family protein	
At4g29710	phosphodiesterase/nucleotide pyrophosphatase-related	
At4g39800	MYO-INOSITOL-1-PHOSTPATE SYNTHASE 1	MIPS1
At4g39940	APS KINASE 2	APK 2
At5g01220	sulfoquinovosyldiacylglycerol 2	SQD2
At5g01750	unknown protein	
At5g02450	60S ribosomal protein L36	RPL36C
At5g02500	HEAT SHOCK COGNATE PROTEIN 70-1	HSC70-1
At5g54280	ARABIDOPSIS THALIANA MYOSIN 2	ATM2
At5g64080	protease inhibitor/seed storage/lipid transfer protein	LTP
At5g64770	unknown protein	

Tab.4.1 Positive biomarkers characterize all the transformed lines whereas negative biomarkers, absent in the transformed lines, characterized wild type plants. The variables that are considered as biomarkers have p(corr) $\geq$ 0.9. The twelve positive biomarkers cell wall related are: XTR7, aspartyl protease, MIOX2, PAT2, GASA1, xyloglucan: xyloglucosyl transferase putative, RABG3B, XTH33, EXL1, RABA5E, EXL3 and pectinacetylesterase family protein.

.

Chapter 5

Concluding Remarks

.

.

ENOD40 gene has been identified in different plant species and its expression is observed during the initiation and development of new organs, such as root nodules (Yang *et al.,* 1993; Fang and Hirsch, 1998; Compaan *et al*., 2001), lateral roots (Asad, 1994; Papadopoulou *et al.,* 1996; Fang and Hirsch, 1998), young leaves and stipule primordia (Asad, 1994; Corich *et al*., 1998).

ENOD40 gene has an unusual structure: it lacks a long open reading frame, but several short ORFs are present. Moreover, at nucleotide level, two regions, named box1 and box2, are highly conserved among all *ENOD40* genes (Vijn *et al.,* 1995; Kouchi *et al*., 1999); in box 1 a highly conserved ORF is present (ORF 1), that encodes a putative peptide of 10-13 amino acids. Furthermore, the gene contains regions corresponding to conserved secondary structures of the transcript (Hofacker *et al*., 2002; Girard *et al.,* 2003; Gultyaev and Roussis, 2007). Indeed, six conserved domains were identified in ENOD40 mRNA and the domains 2 and 3 are strongly conserved among legume and non legume species (Girard *et al.,* 2003). Recent work on ENOD40 secondary structure domains has identified a *ENOD40*-like gene also in Arabidopsis, in which homologous sequences were not previously found (Gultyaev and Roussis, 2007).

Despite several researches, *ENOD40* biological activity has not yet been elucidated. Moreover, if *ENOD40* activity is mediated by the ORF1 peptide, RNA or both is still unclear.

The present work consists of two distinct parts: in the first (Chapter 2), the biochemical research of the tobacco ENOD40 putative peptide has been pursued, unfortunately without producing definitive evidences; in the second part, the possible role of the gene has been investigated through the metabolomics (Chapter 3) and transcriptomics (Chapter 4) characterization of *ENOD40* overexpressing Arabidopsis plants.

That *ENOD40* could act, at least in part, through the production of box 1 peptide is suggested by several observations. In Arabidopsis protoplasts, the transfection with the tobacco full length gene and the administration of the box 1 synthetic peptide confer to the cells the same phenotype (Guzzo *et al*., 2005). *In vitro* assays showed the translation of *Glycine max* and *Medicago truncatula* ORF1 (Rohrig *et al.,* 2002; Sousa *et al.,* 2001). Moreover, translational fusion *NtENOD40* ORF1-GFP (Compaan *et al.,* 2001) and *LlENOD40* ORF1-GUS (Podkowinski *et al*., 2009) resulted in fusion protein production.

.

For all these reasons, extensive efforts were spent in the present work to try to biochemically detect the putative peptide encoded by box 1.
Unfortunately, despite several attempts to set up the purification procedure and the different and sensitive techniques used for the analysis of the putatively peptide-enriched fractions, only MALDI-TOF PSD analysis gave an initial clue of the possible presence of the peptide in *ENOD40* overexpressing BY2 cells.

The other goal of this work was to investigate the role of *ENOD40* gene in non-legume plants.
In a previous work, the cells of different tissues of the Arabidopsis transformed lines 1.1, 3.1 and 4.4, that overexpress soybean *ENOD40* gene, were morphologically characterised. This work demonstrated that the expression of *ENOD40* confers to cells of peripheral tissues and organs, in particular epidermis and mesophyll cells, a reduced cell size compared to the wild type cells (Guzzo *et al.*, 2005). The overall size and shape of the vegetative and reproductive organs of plants of the three transformed lines were entirely comparable to those of wild type plants, suggesting that the smaller cells size of transformed plants is compensated by a higher cell number.
The same cell phenotype was also observed in protoplasts transiently transformed with *ENOD40* gene, in this case the reduced size was due to reduced cell expansion and not to an increased cell division rate, since the reduced size was observed in the protoplasts after the production of the new cell wall but before they re-start to divide (Guzzo *et al.*, 2005).
In the present work, the same transformed Arabidopsis lines were characterized from the molecular point of view, using transcriptomics and metabolomics approaches.
To date, no study on *ENOD40* gene role using these approaches has been reported.
Metabolite and transcriptional profiles of the three Arabidopsis lines overexpressing soybean *ENOD40* gene were acquired and compared to those obtained from wild type plants. Afterward, biomarker analysis of metabolomic and transcriptomic dataset was used in order to identify the metabolites and transcripts that showed the higher correlation with the overexpression of *ENOD40* gene.
In the metabolite profiles, a certain number of glucosinolates positively characterized all the three transformed lines compared with the wild type, while some flavonoids mainly

.

characterized wild type plants. If each of the transformed lines is compared with the wild type, additional glucosinolates positively characterized each transformed line compared with the wild type one.

These results suggest that *ENOD40* expression divert metabolic fluxes from flavonoid pathway to glucosinolate pathway.

Which can be the meaning of this metabolic shift?

Several studies have reported that plants belonging to Brassicaceae family respond to insect damage by systemically accumulating higher levels of glucosinolates, which presumably increase their resistance to subsequent attacks (Agrawal *et al.*, 2002; Mewis *et al.*, 2005). Moreover, a higher level of glucosinolates has been observed in embryos, dormant and germinating seeds, inflorescences, younger leaves and roots (Petersen *et al*., 2002; Brown *et al*., 2003); furthermore, younger is the organs analysed, higher is the content of glucosinolates (Brown *et al.*, 2003).

The high content of glucosinolates in young tissues and organs is probably correlated with their known role as chemical defence against herbivores (Halkier and Gershenzon, 2006); young tissues and organs could be more appetizing than old ones and so could need more effective protection, that could be conferred by higher glucosinolate accumulation. In order to see if also in our system glucosinolates accumulate preferentially in young organs and tissues, metabolites were extracted separately from young and old leaves, and the total content of glucosinolates in the two batches of organs was compared. Glucosinolate levels were much higher in the young leaves, supporting the hypothesis that these molecules play a special role, maybe connected with the chemical defence, in the young organs and tissues.

Surprisingly, transcriptional profiling showed that APS kinase 1 and 2 genes, that are part of the glucosinolate biosynthesis network (Yatusevich *et al.*, 2009; Mugford *et al.*, 2009), are downregulated in the transformed lines. These data are only apparently in conflict, since the biosynthetic pathways of many kinds of secondary metabolites are switched off when the proper metabolite level has already been accumulated in the cells. In the wild type plants, which have lower glucosinolate level, the glucosinolate biosynthetic pathway is probably switched on, as suggested by the APS kinase genes, indicating that probably wild type plants were collected during a moment of increased glucosinolate demand. On the contrary, transformed plants already accumulated higher

.

level of glucosinolates, and transcription of genes related to glucosinolate biosynthesis is downregulated.

With regard to transcriptional profiling, most of the marker genes (twelve out of twenty-three) that characterize all the three transformed lines compared with the wild type line, were correlated with processes occurring in the cell wall. The cell wall is the mechanical determinant of cell shape and size (Zhong and Ye, 2007), and a coordinate expansion and differentiation of the individual cells is due to a dynamic remodelling of the wall structure (Carpita and Gibeaut, 1993). Thus, *ENOD40* gene could be involved in an active process that controls the composition and the dynamics of the cell wall, such as strengthening (XTHs genes) and the delivery of new components to cell wall (RAB GTPASE, MIOX2 genes). Since transformed plants are morphologically characterized by small cell size, genes involved in cell wall dynamics and composition and upregulated in transformed cells, could be involved in the prevention of cell expansion.

Previous morphological studies on *Arabidopsis thaliana ENOD40* transformed lines, characterised by normal organs containing a higher number of smaller cells, and on *ENOD40* transfected Arabidopsis protoplasts characterized by reduced expansion, suggested that the gene could have some role in keeping the cells in a “young” state (Guzzo *et al.*, 2005).

The observation that the *Arabidopsis thaliana ENOD40* transformed plants accumulate high levels of glucosinolates, that are typical of the young tissues, suggests that, also from the metabolic point of view, the transformed cells have features typical of younger cells, whereas wild type cells use their metabolic resources to accumulate flavonoids, another class of secondary metabolites more typical of differentiated state. Of course, it is not possible to exclude that the lower level of flavonoids in transformed cells could have a more specific meaning, not strictly linked to the cell differentiation state.

.

In summary, the transcriptomics and metabolomics analyses showed that *Arabidopsis thaliana ENOD40* transformed lines differed from the wild type ones for three main features: 1) higher accumulation of glucosinolates; 2) lower accumulation of flavonoids; 3) higher expression of some genes correlated with processes occurring in the cell wall.

ENOD40 gene is expressed in young and meristematic tissues during the formation of new organs, such as root nodules (Yang *et al.*, 1993; Fang and Hirsch, 1998; Compaan *et al*., 2001), lateral roots (Asad, 1994; Papadopoulou *et al*., 1996; Fang and Hirsch, 1998), young leaves and stipule primordia (Asad, 1994; Corich *et al.*, 1998). In particular, during nodule formation the gene is expressed before the first cortical cells divisions, and subsequently in nodule primordium, suggesting a very early role in the establishment of this new organ, and in general, a role in organ primordium and meristematic tissue functions.

Our data showed that *ENOD40* overexpression induces genes that control processes occurring in the cell wall, and this ability could explain, at least from the mechanicistic point of view, the small cell size in the organs of Arabidopsis transformed plants. However, the small cell size is also a feature typical of the young less differentiated cells. The observation that *ENOD40* expression induce also from a metabolic point of view features typical of younger cells, suggests that its role is much more complicate than the simple inhibition of cell expansion through the control of cell wall composition. If we consider other features, such as the ability to differentiate normal chloroplasts, the *ENOD40* expressing cells behave as normally differentiated cells, so that the overall appearance of transformed plants is comparable to that of wild type. Thus, the ability of *ENOD40* to induce traits typical of younger cells is limited to some features: the small cell size, the probable consequent increased cell division rate and the glucosinolate accumulation.

Our analyses have been performed on the whole aerial part of 6-weeks-old plants, without discriminating between young and old leaves. However, our observations, together with the pattern of expression of the *ENOD40* genes reported for many species, suggest 1) that this gene has a role on organ primordium and meristematic tissue functions; 2) this role is at least in part connected with its ability to establish and/or keep some features typical of a meristematic cell state.

.

REFERENCES

Agrawal AA, Conner JK, Johnson MT, Wallsgrove R (2002) Ecological genetics of an induced plant defense against herbivores: additive genetic variance and costs of phenotypic plasticity. Evolution 56: 2206-2213

Asad (1994) Isolation and characterization of cDNA and genomic clones of *MsENOD40*; Transcripts are detected in meristematic cells of alfalfa. Protoplasma 183: 10-23

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry 62: 471-481

Carpita NC, Gibeaut DM (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. Plant J 3: 1-30

Compaan B, Yang, W.-C., Bisseling, T., and Franssen, H. (2001) *ENOD40* expression in the pericycle precedes cortical cell division in *Rhizobium*-legume interaction and the highly conserved internal region of the gene does not encode a peptide. Plant Soil 230: 1–8

Corich V, Goormachtig S, Lievens S, Van Montagu M, Holsters M (1998) Patterns of *ENOD40* gene expression in stem-borne nodules of Sesbania rostrata. Plant Mol Biol 37: 67-76

Fang Y, Hirsch AM (1998) Studying early nodulin gene *ENOD40* expression and induction by nodulation factor and cytokinin in transgenic alfalfa. Plant Physiol 116: 53-68

Girard G, Roussis A, Gultyaev AP, Pleij CW, Spaink HP (2003) Structural motifs in the RNA encoded by the early nodulation gene enod40 of soybean. Nucleic Acids Res 31: 5003-5015

Gultyaev AP, Roussis A (2007) Identification of conserved secondary structures and expansion segments in enod40 RNAs reveals new enod40 homologues in plants. Nucleic Acids Res 35: 3144-3152

Guzzo F, Portaluppi P, Grisi R, Barone S, Zampieri S, Franssen H, Levi M (2005) Reduction of cell size induced by enod40 in Arabidopsis thaliana. J Exp Bot 56: 507-513

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol 57: 303-333

Hofacker IL, Fekete M, Stadler PF (2002) Secondary structure prediction for aligned RNA sequences. J Mol Biol 319: 1059-1066

.

Kouchi H, Takane K, So RB, Ladha JK, Reddy PM (1999) Rice *ENOD40*: isolation and expression analysis in rice and transgenic soybean root nodules. Plant J 18: 121-12

Mewis I, Apple HM, Hom A, Raina R, Schultz JC. (2005). Major signaling pathways modulate Arabidopsis glucosinolate accumulation and response to both phloem-feeding and chewing insects. Plant Physiol. 138:1149–62

Mugford SG, Yoshimoto N, Reichelt M, Wirtz M, Hill L, Mugford ST, Nakazato Y, Noji M, Takahashi H, Kramell R, Gigolashvili T, Flugge UI, Wasternack C, Gershenzon J, Hell R, Saito K, Kopriva S (2009) Disruption of adenosine-5'-phosphosulfate kinase in Arabidopsis reduces levels of sulfated secondary metabolites. Plant Cell 21: 910-927

Papadopoulou K, Roussis A, Katinakis P (1996) *Phaseolus ENOD40* is involved in symbiotic and non-symbiotic organogenetic processes: expression during nodule and lateral root development. Plant Mol Biol 30: 403-417

Petersen BL, Chen S, Hansen CH, Olsen CE, Halkier BA (2002) Composition and content of glucosinolates in developing Arabidopsis thaliana. Planta 214: 562-571

Podkowinski J, Zmienko A, Florek B, Wojciechowski P, Rybarczyk A, Wrzesinski J, Ciesiolka J, Blazewicz J, Kondorosi A, Crespi M, Legocki A (2009) Translational and structural analysis of the shortest legume *ENOD40* gene in Lupinus luteus. Acta Biochim Pol 56: 89-102

Rohrig H, Schmidt J, Miklashevichs E, Schell J, John M (2002) *Soybean ENOD40* encodes two peptides that bind to sucrose synthase. Proc Natl Acad Sci U S A 99: 1915-1920

Sousa C, Johansson C, Charon C, Manyani H, Sautter C, Kondorosi A, Crespi M (2001) Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. Mol Cell Biol 21: 354-366

Vijn I, Yang WC, Pallisgard N, Ostergaard Jensen E, van Kammen A, Bisseling T (1995) *VsENOD5, VsENOD12* and *VsENOD40* expression during *Rhizobium*-induced nodule formation on Vicia sativa roots. Plant Mol Biol 28: 1111-1119

Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H (1993) Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean nodule development. Plant J 3: 573-585

Yatusevich R, Mugford SG, Matthewman C, Gigolashvili T, Frerigmann H, Delaney S, Koprivova A, Flugge UI, Kopriva S (2009) Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in *Arabidopsis thaliana*. Plant J

.

Zhong R, Ye ZH (2007) Regulation of cell wall biosynthesis. Curr Opin Plant Biol 10: 564-572