Analysis of a set of 14,549 *Brassica napus* unique ESTs and signature genes related to *Sclerotinia* resistance

MAO Han, DONG Caihua, SU Zhen, JI Ruiqin, GUO Xuelan, LIU Shengyi

Institute of Oil Crops Research, Chinese Academy of Agricultural Sciences, Wuhan, 430062, China.
Email: liusy@oilcrops.cn

Abstract

Sclerotinia sclerotiorum (Ss) is one of the most devastating diseases in oilseed rape (*Brassica napus*). To study the resistance mechanism against Ss in oilseed rape, a normalized cDNA library of seedling leaves treated by BTH (bennzothiadiazole), MJ (methyl-jsamonate), OA (oxalic acid, a toxin produced by Ss) or inoculated with Ss was constructed to produce single-pass expressed sequence tags (ESTs) in the present study. After vector sequences, mitochondrial sequences, rRNA and viral sequences was masked and sequences less than 200bp in length were excluded, in total 35,325 high-quality sequences were produced. Of them, 3,489 ESTs were classified as singlets while 31,804 were assembled into 11,060 contigs, resulting in a non-redundant set of 14,549 unique ESTs. About 90.87% of the unique ESTs have identifiable homology in the *Arabidopsis* genome. Genes corresponding to metabolism have remained most conserved between these two plant genomes. The putative products of these unique ESTs were annotated according to their homology with the categorized proteins of *Arabidopsis*. The abundance of transcripts related to the disease was evaluated by RNA blots. The further analysis is still in process. Most of the disease defense response genes reported in *Arabidopsis* have their homologies in *Brassica napus*, including those in very up-stream and very down-stream of the signaling pathways. Interestingly, much more genes were involved in the JA-dependent signaling pathway than in the SA-dependent signaling pathway.

Key words: Oilseed Rape; ESTs; *Sclerotinia* Resistance

Introduction

Brassica. napus (oilseed rape) belongs to the *Brassicaceae* family, an economically important plant family that includes vegetables such as cabbage, broccoli, mustard, radish. Oilseed rape is the third largest oilseed crop in the world, providing approximately 13% of the world's supply of vegetable oil, and it is widely cultivated as an economically important crop in China. *Sclerotinia sclerotiorum* is a ubiquitous necrotrophic fungal pathogen capable of infecting at least 408 plant species of 75 families (Boland and Hall, 1994). In *B. napus*, the pathogen causes rot of stems, leaves and podsunder favorable weather conditions, resulting in severe yield losses. Oilseed rape is cultivated on close to 8 million hectares in China and yield losses caused by *S. sclerotiorum* were normally 10 %～20%. To date, complete resistance to this pathogen has not been identified, although partial resistance was recently reported in some *B. napus* cultivars like Zhongyou 821 (Buchwaldt et al., 2003). Although some hormones and metabolism pathways had been found to play an important role in the process of disease resistance, interactions between *S. sclerotiorum* and plants were still indistinct, especially on *B. napus*. Identification and selection for genes against *S. sclerotiorum* is an effective way to reduce the yield loss.Genome sequencing is an efficient way toward investigation of functional genome of plant species. Sequencing of *Arabidopsis thaliana* genome has been completed, which provides us with a detailed view of the gene content and genome organization of the plant species. Yet the degree of conserved gene content, gene number, and genome organization among plant species remains unresolved. And we know that oilseed rape is an amphidiploid (allotetraploids), arising from the hybridization of an A genome progenitor *B. rapa* (AA,2n=20) and a C genome progenitor *B. oleracea* (CC, 2n=18)(U, 1935). Its chromosome number is n=19 and the size of its genome is ca. 1200Mb (Arumuganthan and Earle, 1991). Therefore it would be difficult to undertake sequencing of the whole genome at present. However we can still gain some information about the *B. napus* genome by single pass sequencing of cDNA clones to generate ESTs and analysis of these ESTs.

In this study, we used large-scale EST sequencing for gene expression profiling at early infection stages in *B. napus*. We constructed normalized cDNA libraries using mRNA isolated from the fungus-infected leaf tissues, and mRNA from non-infected leaf tissues as control. In order to obtain more disease defense genes involved in different signaling pathways, we also treated leaves by salicylic acid homolog bennzothiadiazole (BTH), methyl jasmonate (MJ) and oxalic acid (OA). We performed extensive analysis of the ESTs using a variety of computational methods.

Materials and methods

Plant Materials and cDNA Library Construction

An Oilseed Rape cv. Zhongshuang 9 was grown in a growth room for RNA isolation from leaves of 6-week-old seedlings. Leaves were treated by different chemicals, including 0.1mM BTH, 0.1mM MJ (methyl-jsamonate), 5mM OA (oxalic acid, a toxin produced by *S. sclerotiorum*), or inoculated by *S. sclerotiorum*. For the control, *B. napus* was inoculated with water. The inoculated plants were placed in a plastic box (covered tightly) and incubated in the dark for 24h at 25°C, and

leaf tissues were collected 24h after inoculation.

Total RNA was isolated separately from each treatment using Trizol (GIBCO BRL) according to the manufacturer's instructions. Poly(A+) mRNA isolated from each treatment by oligo(dT) chromatography, and were pooled pro rata, and cDNA were synthesized with oligo(dT) primer carrying a *Sfi* I site. cDNA were selected to enrich for length>1.0kb and subjected to a normalization process, and then were cloned directly into the *Sfi* I A /*Sfi* I B sites of a re-constructed *pBluescript* II SK vector (the *EcoR* I/ *Not* I site was replaced with the *Sfi* I A/*Sfi* I B site).

EST Sequencing and Assembly of EST into Contigs

Plasmids were isolated according to a standard alkaline lysis protocol and used for capillary sequencing (MegaBACE 1000 and ABI3700). All the clones were randomly selected from the normalized library and subjected to single-pass sequencing reactions from the 5′-end using a T7 primer (5′-TAATACGACTCACTATAGGG-3′). Raw data as chromatogram files were processed for base-calling and quality assessment using the phred software program (Phred-Phrap-Consed package; Ewing and Green, 1998; Ewing et al., 1998). Vector sequences, mitochondrial sequences, rRNA and viral sequences were masked with CROSS_MATCH (version 0.990319, Phil Green). Finally, sequences whose length was less than 200bp were excluded from the following analysis. These EST sequences were then assembled into contigs using phrap (http://www.phrap.com) with the default parameters.

Individual ESTs and the unique sequences set were annotated by BLASTX search against the proteome of *Arabidopsis* whose functional categories were assigned by TAIR (http://www.arabidopsis.org/) and TIGR (http://www.tigr.org/) with the terms from the Gene Ontology Consortium controlled vocabularies (http://www.geneontology.org). In the annotation, the following criteria were used. In general, E value scores lower than 1.0 e_4 were considered to be significant. When multiple hits were found, the one with the longest extended homology was selected.

Results and Discussion

Normalization can enrich cDNAs with relatively low copies and should increase the number of unique ESTs in random sequencing (Soares et al., 1994; Smith et al., 2001). Our normalized *B. napus* cDNA library was synthesized using mRNA from *B. napus* leaves treated by *S. sclerotiorum*, BTH, MJ, and OA. The library consisted of at least 3.4 x10^7 primary recombinant plasmids with an average insert size of 1.2 kb. About 600 ESTs were randomly selected from the library to investigate cDNA length. Most of their length ranged from 1.0kb to 2.0kb (Fig. 1). After vectors were removed and contaminations were masked, 35,325 high-quality ESTs with at least 200 bp in length were obtained. These sequences assembled into 11,060 tentative consensus (TC) sequences and 3,489 singletons, thus representing 14,549 unique ESTs(Tab1.)

Table 1. Statistics of a *Brassica napus* Unique EST Set

	No. of Sequences
Total number of ESTs	35293
ESTs in contigs	31804
Singlet ESTs	3489
Total number of contigs	11060
Total number of unique sequences(Unigenes)	14549

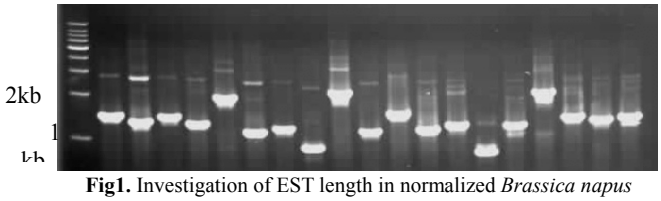

Fig1. Investigation of EST length in normalized *Brassica napus* libray

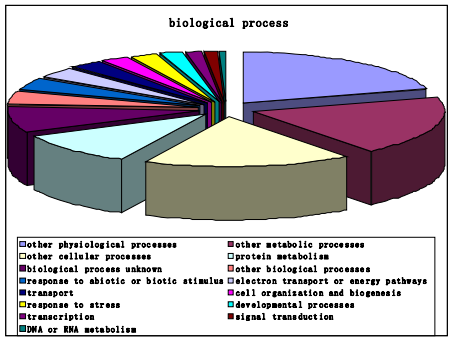

Fig.2 Functional classification of *Brassica napus* unique ESTs

About 90.87% of the unique ESTs have identifiable homology in the *Arabidopsis* genome. Genes corresponding to metabolism have remained most conserved between these two plant genomes. All the unique ESTs were categorized with respect to functionally annotated genes in *Arabidopsis* and grouped into 15 categories of biological functions (Fig.2). The largest proportion of functionally assigned unique ESTs fell into four categories: (1) other physiological processes, (2) other metabolic processes, (3) other cellular processe, (4) biological process unknown. Disease defense-related unique ESTs were further classified into 10 catalogs according to their functional annotation.

These disease defense-related genes involve in SA, JA, ethylene (ET) or oxidative burst (OB) signal pathways, respectively. Of them, more than 300 genes were involved in oxidative stress and ROS, and 4 genes involved in virus inducing gene silencing. Surprisingly Many R-genes and up-stream genes in signalling pathways were expressed and sequenced. For examples, CC-NBS-LRR,TIR-NBS-LRR and RPPS genes; EDS1, EDS5, PAD4, SGT and EDR1; MPK, MAPK, MAPKK, CTR1 and other kinases; MAD-box, AP2, WRKY, bZip, TGA; SSI2, CET1; NPR1, PR1, PDF1.2 and other PR proteins; Genes involved in metabolisms of JA, SA, phenolics, flavoniod and ubiquitins. According to the comparative analysis with *Arabidopsis* information, about 70% of *Arabidopsis* disease defense-related genes were found in this unique EST library. Interestingly, much more genes sequenced were involved in the JA-dependent signaling pathway than that in the SA-dependent signaling pathway in this library, implying that the JA-dependent signaling pathway may play the more important role in the response to *S. sclerotiorum* in *B. napus* in comparison with other plant-pathogen systems.

Referances

Arumuganthan K., Earie, E.D.(1991) Nuclear DNA content of some important plant species. Plant Mol Biol Report 9, 208-218.

Boland G.J., Hall R. (1994). Index of plant hosts of *Sclerotinia*. Can J Plant Pathol 16: 93-108.

Buchwaldt L., Yu F. Q., Rimmer S.R. Hegedus D.D. (2003). Resistance to *Sclerotinia sclerotiorum* in a Chinese *Brassica napus* cultivar. In: International Congress of Plant Pathology, Christchurch, New Zealand, 2–7 February.

Ewing B., Green P. (1998). Basecalling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8:186-194.

Ewing B., Hillier L., Wendl M., Green P. (1998). Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Research 8:175-185.

Smith, T.P.L., et al. (2001). Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. Genome Res 11: 626–630.

Soares M.B., Bonaldo M.D.F., Jelene P., Su L., Lawton L., Efstratiadis A. (1994). Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91: 9228–9232.

U,N.(1935). Genome analysis in Brassica with special reference to the experimental formation of *Brassica napus* and peculiar mode of fertilization. Jpn J Bot 7: 389-452.