Analysis of EST Libraries from *Leptosphaeria maculans* and Blackleg Infected *Brassica napus*

J.A. Christianson[*,1], S.R. Rimmer[2], A.G. Good[1], and D.J. Lydiate[2]

[1]*Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada*

[2]*Saskatoon Research Center, Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, Saskatchewan, S7N 0X2, Canada*

Abstract: The ascomycete *Leptosphaeria maculans* is the causal agent of blackleg of crucifers. cDNA libraries were constructed and partially sequenced to increase the amount of *L. maculans* sequence data available and to survey genes expressed during pathogenesis. Two libraries were built, one from cultured mycelia (LM) and one from *Brassica napus* leaf tissue infected with *L. maculans* (ILS). A total of 741 expressed sequence tags (ESTs) were obtained from the LM library, and 1284 were obtained from the ILS library. The vast majority of the ESTs obtained from the ILS library appear to originate from *B. napus* (70-81% depending on the identification method). Various methods for determining sequence origin in mixed libraries were compared. cDNAs with a putative role in pathogenesis were identified from the ILS library. These and other cDNAs of interest from the LM library are discussed.

INTRODUCTION

Leptosphaeria maculans (anamorph. *Phoma lingam*) is the causal agent of blackleg (Phoma stem canker) of *Brassica* species and is responsible for considerable seed yield loss in oilseed rape crops worldwide. *Leptosphaeria maculans* is an ascomycete in the order Dothidiomycetes. *L. maculans* grows septate, branched, hyaline mycelia when young, but these turn darker with the buildup of pigment over time. This fungus produces wind borne ascospores after mating, which are thought to be the primary source of infection [1]. Asexual pycnidiospores may also be produced; these emerge in pinkish exudates and are likely spread to neighboring plants through rain splash [1, 2]. *L. maculans* infects hosts through wounds or natural openings such as stomata [3, 4]. Hyphae grow in the intercellular spaces of host mesophyll tissue, establishing small colonies within four days post infection [4]. Biotrophic growth continues in the intercellular space until approximately eight to ten days post infection and no symptoms are visible during this period; afterward, chlorotic patches and necrotic lesions begin to develop in infected tissues [3]. These symptoms develop behind the growing hyphal front, which colonizes the vascular system and, subsequently causes the characteristic basal stem canker from which the disease name is derived [3, 5]. *L. maculans* is also able to grow saprophytically on residue in the field and can persist for up to five seasons. The molecular mechanisms that enable the fungus to grow and develop within the host and evade the host defense response are largely unknown. Although, in general, the number of genes from plant pathogens cloned and studied is increasing rapidly, only three *L. maculans* pathogenicity genes have been reported as yet [6-8].

*Address correspondence to this author at the Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada; Tel: +61 02 6246 4914; Fax: +61 02 6246 5000;
E-mail: Jed.Christianson@csiro.au

Collections of expressed sequence tags (EST) provide a relatively inexpensive way to obtain valuable sequence data from organisms that are not yet well represented in sequence databases [9] and have been used successfully to study plant pathogen interactions. cDNA libraries have been sequenced from soybean tissues infected with *Phytophthora sojae* [10, 11]; rice tissue infected with *Magnaporthe grisea* [12]; potato tissue infected with *Phytophthora infestans* [13], *B. napus* infected with *Sclerotinia sclerotiorum* [14], *Brassica oleracea* infected with *Alternaria brassicicola* [15] and one library of ESTs derived from *B. napus* infected with *L. maculans* [16]. One of the difficulties in working with libraries which contain a mix of sequences from different organisms is distinguishing which organism any particular sequence is derived from. Various methods for distinguishing sequence origin have been utilized in the past. These include: hybridization of arrayed ESTs to genomic DNA from each contributor organism [16]; using sequence alignment tools to compare each EST to public databases and inferring sequence origin from closely matched sequences [10, 12, 13, 15]; Use of %G+C content to separate individual ESTs [10, 13]; calculating codon use patterns for individual ESTs and comparing codon frequencies to averages observed in sequences known to be from each contributor organism [14]; or computational approaches that take into account both %G+C and codon use bias [11]. To increase the availability of *L. maculans* sequences and to gain understanding of the molecular biology underlying pathogenicity of this organism, ESTs were generated from a cDNA library made from leaf tissue of a susceptible *B. napus* host eight days after infection with *L. maculans* as well as from *L. maculans* cultured *in vitro*. To date, 741 ESTs from *in vitro* cultured *L. maculans* and 1284 ESTs from infected leaf tissue have been collected. The infected leaf library was subjected to various computational methods to establish the source organism for each sequence. Genes with putative roles in pathogenicity were identified and are discussed.

1874-4370/08

MATERIALS AND METHODS

Culture Conditions

Leptosphaeria maculans strain PL86-12 from the collection at the AAFC Saskatoon Research Centre was grown on V8 agar (10% V8 juice; 0.02% $CaCO_3$; 1.5% agar). A $1cm^2$ plug of agar containing actively growing mycelium was transferred to 200 ml liquid V8 medium. This culture was grown for 5 days at room temperature on an orbital shaker at ~50rpm before RNA extraction.

V8 agar plates of PL86-12 were incubated at room temperature under continuous fluorescent light. Spores were collected in sterile water after 2 weeks and filtered through sterilized Whatman filter paper (Grade 2) and the concentration adjusted to 1×10^7 spores ml^{-1} in sterile distilled water. Spore suspensions were stored at -20°C.

Inoculation of *Brassica napus* with *Leptosphaeria maculans*

Plants of *B. napus* accession N-o-1, a doubled haploid line derived from Westar [17], were grown in a growth cabinet for four weeks and expanded leaves were inoculated with 1ml of a 10^7 spores ml^{-1} suspension of PL86-12 pycnidiospores. Spores were infiltrated through the bottom leaf surface using a 10 ml syringe without needle until approximately 75% of the leaf area showed signs of water soaking. Eight days post infection, a single inoculated leaf showing signs of widespread infection and a single leaf of equivalent age from a non-inoculated plant were used for RNA extraction.

RNA Extraction and Subtractive cDNA Library Construction

Five grams of 5-day-old fungal mycelia and 10 g each of eight day infected and non-infected leaf tissues were flash frozen in liquid N_2 and ground to a fine powder. Total RNA was extracted from plant and fungal tissues and mRNA was subsequently purified using oligo dT columns. Three cDNA libraries were constructed using the Superscript Plasmid System for cDNA Synthesis and Plasmid cloning kit (Gibco BRL Life Technologies Cat. No. 18248-013); one library from *in vitro* cultured *L. maculans* (LM), one from infected *B. napus* leaf tissue (IL) and one from uninfected *B. napus* leaf tissue (BL). The cDNAs were cloned into pSPORT1 or pSPORT2 digested with *Sal I* and *Not I*. These plasmid vectors differ only in the orientation of their multiple cloning sites. The IL library in pSPORT1 was used to produce target single stranded DNA (ssDNA). Biotinylated driver RNA was produced from the LM and BL libraries in pSPORT2 and target ssDNA was allowed to hybridize with biotinylated driver RNA. Avidin was used to bind and remove all driver sequences and any target sequences that had hybridized to them, resulting in the ILS library.

Plasmid Preparation

Bacteria transformed with a portion of the cultured *L. maculans* cDNA (designated with the prefix LM) and bacteria transformed with subtracted infected leaf library cDNA (designated with the prefix ILS) were plated on LB agar containing 50 □g ml^{-1} carbenicillin and X-gal (5-bromo-4-chloro-3-indoyl-□-D-galactosidase). White colonies were selected and transferred individually to 96-well plates containing 1.4 ml liquid TB media (1.2% w/v tryptone, 2.4% w/v yeast extract, 0.4%v/v glycerol, 17mM KH_2PO_4, 72mM K_2HPO_4) and 50 mg ml^{-1} carbenicillin and incubated overnight at 37°C. Plasmid DNA was extracted using the alkaline lysis procedure and purified with QIAprep 96 Turbo miniprep kits according to the manufacturer's protocol (Qiagen Cat. No. 27191). Purified plasmids were diluted to a final concentration of 100 ng ml^{-1}.

cDNA Sequencing

Sequencing reactions were carried out using an ABI 877 integrated thermocycler with Big Dye v3.0 sequencing chemistry(Applied Biosystems) at half scale using 300 ng of template. Forward and reverse sequencing reaction were carried out separately on each template using 3.2 pmol of M13 Forward or M13 reverse universal primers. Cycle sequencing was conducted using the following conditions; 15 cycles of 96° C for 10s, 55° C for 5s, 70° C for 1 min followed by 15 cycles of 96° C for 10s, 70° C for 1 min with 1° C s^{-1} ramp between cycle steps. Sequencing products were purified to remove unincorporated nucleotides using either the Bio-Rad SEQueaky clean dye terminator removal kit (Bio-Rad laboratories) or with ethanol precipitation. Purified sequencing products were separated using an ABI 377 DNA sequencer. All *L. maculans* ESTs have been submitted to the genbank dbEST (accession numbers EB762676-EB763482 and EC091584-EC091606).

Sequence Analysis

All cDNA sequences were trimmed for vector sequences and for sequence quality using Sequencher software (Gene codes corporation, Ann Arbor, MI). Sequences were trimmed from the 5' end until the first 25 bases of sequence contained fewer than three ambiguous base calls or up to 25% of the total sequence length. Sequences were trimmed from the 3' end until the last 25 bases of sequence contained fewer than three ambiguous base calls. After sequence trimming, forward and reverse sequences from each clone were aligned where possible to form a single contig. In cases where EST sequences did not overlap the individual ESTs were provided with the same clone name but denoted with the suffix "F" or "R". Where only one end of the cDNA was successfully sequenced, this was again denoted by an ''F'' or ''R'' suffix. All ESTs were compared to public domain databases (non-redundant GenBank CDS; http://www.ncbi.nlm.nih.gov/) using the BLAST and gapped BLASTX algorithms [18]. All ESTs were also compared using the TBLASTX algorithm [18] to both a customized plant sequence database composed of the *Arabidopsis thaliana* (WWW.ARABIDOPSIS.ORG) genome sequences and a selection of *B. napus* ESTs available in the non-redundant GenBank CDS database and a customized fungal sequence database comprised of the *Neurospora crassa*, *Aspergillus nidulans* and *Stagonospora nodorum* genome sequences available through the Fungal Genome Initiative (WWW.BROAD.MIT.EDU) and all ascomycete ESTs available through Cogeme's Phytopathogenic Fungi and Oomycete EST database (http://www.cogeme.ex.ac.uk).

Hexamer frequency analysis was performed using a program originally described by White *et al.* [19]. This program was written in the C programming language and can be downloaded from the TIGR website: (ftp://ftp.tigr.org/pub/

software/qc/). The program employs a sliding hexamer window to tabulate the frequency of each of the 4,096 possible 6 bp sequences in two training sets from different organisms. The frequency of hexamers is tabulated for an individual test sequence and this result is compared with the frequency distribution derived from the two training sets using a log-likelihood ratio function to generate a dissimilarity value (t-value) for the test sequence. In this case, negative t-values result when a sequence is more plant-like and positive t-values result when a sequence is more fungus-like, with t=0 being a distinct threshold. Error rates for analysis can be determined by calculating dissimilarity values for each coding sequence in each of the training sequence sets used. In this manner the number of known plant sequences which would be identified as fungal-like (and vice-versa), or the misidentification rate, can be determined.

Following the procedures outlined in Hraber and Weller [11], *B. napus* coding sequences excluding mitochondrial, plastid and non-coding tRNA and rRNA sequences were retrieved from NCBI's Genbank using Entrez. The Entrez search was also limited to exclude EST, STS (sequence tagged sites), GSS (genomic survey sequences), working draft and patented sequences. *Leptosphaeria maculans* sequences were obtained in the same manner. However, only 111 sequences fitting this description were found in Genbank. These were supplemented by the EST sequences from the LM library in order to increase the available sequence data for the *L. maculans* control sequence set. All sequences were screened for polynucleotide termini and sequences that were less than 300 nt were not used in the analysis.

RESULTS

General Characteristics of the EST Libraries

Both cDNA libraries were of high quality. The average length of the ESTs generated from the cultured *L. maculans* (LM) and the infected leaf (ILS) libraries was 563nt. In the LM libraries, 741 ESTs were obtained from 551 cDNAs (those cDNAs for which the forward and reverse sequences did not overlap are present twice in the EST collection), representing 461 unique sequences (16% redundant sequences). From the ILS library, 1284 ESTs were obtained from 1203 cDNAs, representing 976 unique sequences (25% redundant sequences). Sequence assembly using both the LM and ILS ESTs identified 23 LM ESTs and 16 ILS ESTs which formed 13 contigs, indicating a low level of overlap in these sequence sets.

Determining Sequence Origin in the Mixed ILS Library

All ESTs obtained from *L. maculans* and infected *B. napus* leaf tissue were compared to the Genbank non-redundant nucleotide and protein databases using the BLASTN and BLASTX algorithms [18].

Although complete genome sequence for various filamentous fungi is available, the BLASTN search suggested that only 20% of the LM ESTs (from the cultured fungus library) have any significant (e-value < $1e^{-5}$) similarity to fungal nucleotide sequences available in the Genbank nr database (Fig. **1A**). For the majority of LM ESTs (69%) no significant matches to known sequences could be found. Although another report, using 100 publicly available *L. maculans* ESTs, was able to identify 43% of these sequences as fungal, they used a much higher e-value (<$1e^{-1}$) cutoff for their assignments. As expected, the same search indicated that most (70%) of the ILS library ESTs (from the infected leaf library) had a high degree of nucleotide sequence similarity to *A. thaliana* or other *Brassica* species (e-value < $1e^{-5}$) and only a small percentage (3%) fungal sequences. Surprisingly, a low level (5%) of the LM library ESTs also have nucleotide sequences similar to plant sequences (e-value < $1e^{-5}$) (Fig. **1A**), however the degree of similarity for these matches, while exhibiting e-values less than $1e^{-5}$ is still weak and presumably due to sequence conservation between these organisms.

When a comparison of the six reading frame translations of each LM library EST to the Genbank protein database was performed using BLASTX, the proportion of LM library ESTs exhibiting similarity to filamentous fungi protein sequences rose to 57% (e-value of < $1e^{-5}$) (Fig. **1B**). In contrast, when the ILS library ESTs were compared to the Genbank protein database using the same method, the proportion of ESTs for which the strongest match is a fungal protein sequence only rose from 3% to 4% (Fig. **1B**).

A more rigourous BLAST-based approach was attempted in order to determine the derivation of each of the ILS ESTs. TBLASTX comparisons were made using the ILS EST set, the LM EST set and a collection of 426 *B. napus* ESTs obtained from Genbank against collections of plant and fungal sequences (Fig. **1C**). These included the entire *A. thaliana* genome and a set of fungal sequences containing the entire the entire *Stagonospora nodorum* genome obtained from The *Arabidopsis* Information Resource (HTTP://WWW.ARABIDOPSIS.ORG), Genbank (HTTP://WWW.NCBI.NLM.NIH.GOV/ENTREZ/), the Broad Institute (HTTP://WWW.BROAD.MIT.EDU/) and Cogeme's phytopathogenic EST database (http://www.cogeme.ex.ac.uk). Particular ESTs were classed as being more plant-like or more fungus-like based on a comparison of the e-values of their best matches in each of the plant and fungal sequence collections. When this approach was applied to the collection of *B. napus* ESTs, all ESTs were determined to be plant-like. When applied to the LM ESTs 87% of the ESTs were classified as fungal, 2% as plant-like and 11% remained unclassified (neither plant nor fungal sequence collections provided a match with an e-value less than $1e^{-5}$). The high level of fungal classification achieved by this method is primarily due to similarity with *Stagonospora nodorum*. Both *S. nodorum* and *L. maculans* are dothideomycetes, and 68% of those LM sequences identified as fungal were most similar to *S. nodorum* sequences. In comparison the ILS EST set was composed of 7% fungal, 81% plant-like and 12% unclassified sequences (Fig. **1C**).

The average percentage G+C content was calculated for sequences in the LM library and found to be 51%, with a standard deviation of 8%. The average %G+C content for all of the EST sequences for *L. maculans* currently available in the Genbank EST and Cogeme databases (613 in total) was 50% with a standard deviation of 6%. For comparison, a sample of *B. napus* coding sequences was retrieved from Genbank using the Entrez batch retrieval tool. The *B. napus* sequences have a mean G+C content of 43%, with a standard deviation of 5%. The means for the LM library sequences and the retrieved *B. napus* sequences are significantly different when compared using Student's two tailed t-test assum-

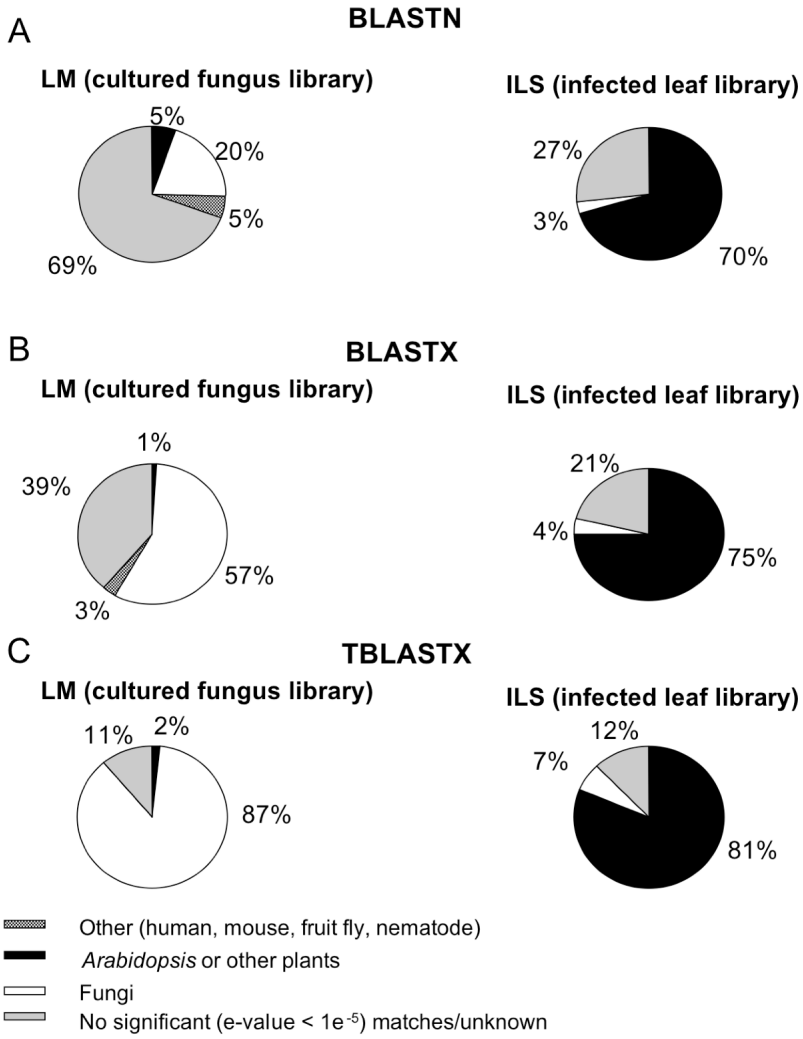

Fig. (1). Similarity of library ESTs to sequences derived from other organisms. (**A**) Results based on EST nucleotide sequences compared to the Genbank nr nucleotide database using the BLASTN algorithm. (**B**) Results based on amino acid translations of EST sequences compared to the Genbank nr protein database using the BLASTX algorithm. (**C**) Results based on amino acid translations of EST sequences compared to custom sequence databases using the TBLASTX algorithm.

ing unequal variances (t=19, P<0.0005). However, the G+C content distributions for each set overlap substantially, making it difficult to predict the origin of any particular EST with a high degree of confidence (Fig. **2**).

A second computational approach to determining sequence origin, based on an analysis of hexamer frequencies as described by Hraber and Weller [11] was attempted. The frequency of occurrence of each possible nucleotide hexamer in any coding sequence will be influenced by the overall %G+C content for that organism as well as the codon use preference in coding sequences of that organism.

The hexamer analysis was performed on ILS sequences using the *B. napus* and *L. maculans* sequence collections described for calculation of %G+C content as training sets. Of the ILS sequences, 158 (12.3%) were removed from the ILS EST set before the analysis due to inadequate length (<300nt); these were considered unclassified. An analysis of the hexamer frequencies of each of the remaining ILS ESTs indicated that 838 (65.3%) of the ILS ESTs appeared to originate from *B. napus* and 288 (22.4%) appeared to originate from *L. maculans*. Analysis of dissimilarity values for each sequence in the *B. napus* training set indicated a misidentification rate of 17% (*B. napus* training set sequences appear to be more similar to fungal sequences), and analysis of each of the *L. maculans* training set sequences indicated a misidentification rate of 7% (*L. maculans* sequences appear more plant-like). If only 83% of the plant sequences in the ILS library will be correctly identified as such by the hexamer analysis, then it can be expected that the 838 sequences identified as plant like are only 83% of the total number of plant sequences and the other 17% (~172 sequences) are classified as fungal. Given that of the 288 ESTs classified as fungal-like by this method, over half (172/288, or 60%) are likely misidentified plant sequences, assignment of ESTs were ultimately based on the TBLASTX comparisons to plant and fungal sequence sets.

Analysis of Putative Functions

For each EST, attempts were made to infer function based on comparisons of amino acid translations of that EST in each of the possible six reading frames to the entire non-

redundant GenBank CDS database using the BLASTX algorithm.

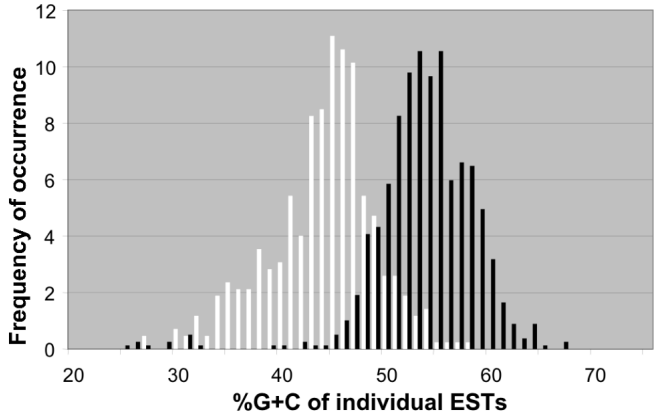

Fig. (2). Histogram displaying %G+C content for *B. napus* and *L. maculans* ESTs. %G+C content was calculated for *B. napus* sequences obtained from Genbank and from the LM EST set. Each bar represents the percentage fraction of the respective training set ESTs corresponding to a %G+C content. Black bars *L. maculans* sequences; White bars *B. napus* sequences

For 191 of the 741 LM ESTs a putative function could be assigned based on that ESTs top match in the Genbank nr protein database (Table **2**). Genes involved in metabolism comprised the largest group of those with a functional annotation, although genes involved in cellular transport and cell rescue, defense and virulence were also highly represented. There was an unexpected abundance of ESTs with a high degree of similarity to alcohol dehydrogenase and pyruvate decarboxylase genes which may indicate that the fungal culture was not thoroughly aerated and may have been suffering anaerobic stress despite being grown in with constant shaking. Although the LM EST set was derived from RNA expressed during normal growth in culture, a number of ESTs from this library could potentially have a role in infection. These include transcripts with a likely role in the detoxification of reactive oxygen species, production of toxins or defense against toxic compounds, e.g. LM9-194R and LM6-147 (Table **2**, see Discussion).

The 86 ILS ESTs that were identified as being derived from *L. maculans* were compared to the Genbank nr nucleotide database using the BLASTN algorithm, the Genbank nr protein database using the BLASTX algorithm and the set of *S. nodorum* annotated genes (HTTP://WWW.BROAD.MIT.EDU) using the BLASTN algorithm. The ESTs were partitioned into functional categories in accordance with the MIPS functional catalogue (Table **1**). Just under half (49%) of the fungal ILS ESTs to which a function could be assigned appear to be involved in ribosome biogenesis. The fungal ILS ESTs also contain genes which may be involved in pathogenesis, e.g. ILS-409R, ILS-1220, ILS-312F, ILS-598 and ILS-718R (Table **1**, see Discussion).

DISCUSSION

Two EST libraries were constructed to gain information about the *L. maculans* - *B. napus* interaction. The first library, constructed from actively growing mycelia of *L. maculans*, has increased the amount of coding sequence information available for this organism. The second library, constructed from infected leaf tissue is a resource for new gene discovery in the area of host-pathogen interactions. The ESTs generated from these libraries represent only a small sample of the genes expressed in either the cultured fungal mycelia or the infected leaf tissue, as evidenced by the low levels of redundancy in the sequence libraries.

A major issue for researchers attempting to analyze collections of sequences derived from a mix of organisms, as is the case in studying sequences from infected tissues, is distinguishing which of the potential contributor organisms any particular sequence originates from. We have used a number of computational approaches which rely on either sequence similarity to previously identified sequences or on characteristics of the sequences themselves (%G+C content, codon use bias) and compared the results from these methods. Initially, BLASTN and BLASTX searches against the Genbank nucleotide and protein databases were performed, however these were only unable to identify at least 21% of the library as either plant or fungal. To identify the origin of a greater number of these unclassified sequences, an analysis of hexamer frequencies in each EST was attempted. This approach compares the frequency of occurrence of every possible hexamer in a test sequence to tables of hexamer frequency distribution calculated from training sets of sequences known to originate from each possible source organism, and has been used previously on libraries from *P. sojae*-infected soybean and different infected *Medicago truncatula* libraries with an error rate of 6% [11]. A comparison of sequences from *B. napus* and *L. maculans* had a high (17%) rate of plant sequences misidentified as fungal. This high false positive rate may reflect a relatively low degree of differentiation in hexamer frequency between *L. maculans* and *B. napus* compared to that between *G. max* and *P. sojae*. In a final effort to elucidate the source of each EST in the ILS library, another BLAST based approach was used. Each of the ILS ESTs was compared using the TBLASTX algorithm to two collections of sequences, one from *A. thaliana* and *B. napus* and the other from various filamentous fungi. ESTs were assigned to a source organism based on their highest BLAST score. This method gave the lowest number of unassigned ESTs (12%) and tended to be in very good agreement with both the other BLAST methods and the hexamer frequency analysis when identifying plant-like sequences. Hsiang and Goodwin [20] also determined that standalone TBLASTX comparisons to custom databases classified more fungal sequences and with greater accuracy than comparisons to the genbank nr database when applied to a broad variety of fungal ESTs from public databases. Further, they also observed that sequences from a variety of plant species were identified at the approximately the same rate when any of the *A. thaliana*, *Oryza sativa* or genbank nr databases were used, likely due to the relative abundance of plant sequences available in the genbank nr database [20]. The ESTs in the unclassified category could not be resolved despite having access to the *A. thaliana* and *S. nodorum* (close relatives of *B. napus* and *L. maculans* respectively) genome sequence data. More of the ESTs in this category may be able to be classified when *L. maculans* genome sequence becomes available, although a recent analysis of ESTs from *Alternaria brassicicola* infected *Brassica oleracea* could not classify the origin of 18% of the ESTs when *A. thaliana* and draft *A. brassicicola*

Table 1. Functional Classification of *Leptosphaeria maculans* ESTs Recovered from Infected *B rassica napus* Leaf Tissue

EST	GenBank Accession Number	Match Description	E-Value	Match Accession Number	Match Database
METABOLISM					
	amino acid metabolism				
ILS-847	EC091603	Pfam identified Cupin	5.6E-45	SNOG_03083	*Stagonospora nodorum genes*
PROTEIN FATE (folding, modification, destination)					
	protein folding and stabilization				
ILS-1148	EB762682	hypothetical protein similar to Tcp1	6.00E-36	SNOG_14059	*Stagonospora nodorum genes*
	protein modification				
ILS-235R	EB762741	UBQ11 (UBIQUITIN 11)	3.00E-71	gi|30679957	genbank nr nucleotide
ILS-1096	EB762679	Pfam identified Protein phosphatase 2C	1.00E-07	SNOG_05256	*Stagonospora nodorum genes*
PROTEIN SYNTHESIS					
	Ribosome biogenesis				
ILS-1208	EB762684	18S ribosomal RNA gene	5.00E-116	gi|58339197	genbank nr nucleotide
ILS-1227	EB762686	large subunit ribosomal RNA gene	1.00E-86	gi|56555563	genbank nr nucleotide
ILS-1236	EB762687	Pfam identified Ribosomal protein S17	2.00E-08	SNOG_13626	*Stagonospora nodorum genes*
ILS-1271R	EB762688	Pfam identified Ribosomal protein L21e	7.00E-51	SNOG_00986	*Stagonospora nodorum genes*
ILS-1377	EB762691	Pfam identified KOW motif; Ribosomal protein L14	2.00E-42	SNOG_07798	*Stagonospora nodorum genes*
ILS-1405	EB762692	cytosolic large ribosomal subunit protein L30	3.00E-40	gi|71001145	genbank nr nucleotide
ILS-197	EB762696	Pfam identified Ribosomal protein S26e	3.00E-90	SNOG_12195	*Stagonospora nodorum genes*
ILS-204	EB762698	Pfam identified Ribosomal L38e protein family	1.00E-98	SNOG_03650	*Stagonospora nodorum genes*
ILS-217	EB762699	Pfam identified Ribosomal L38e protein family	1.00E-98	SNOG_03650	*Stagonospora nodorum genes*
ILS-235F	EB762742	Pfam identified Ribosomal L40e family; Ubiquitin family	2.00E-43	SNOG_05870	*Stagonospora nodorum genes*
ILS-180	EB762694	ribosomal protein S7	3.00E-50	SNOG_00241	*Stagonospora nodorum genes*
ILS-306R	EB762705	18S ribosomal RNA gene	3.00E-133	gi|45181629	genbank nr nucleotide
ILS-35F	EB762707	18S ribosomal RNA gene, partial sequence	5.00E-138	gi|55585866	genbank nr nucleotide
ILS-587R	EB762718	28S large subunit ribosomal RNA gene	2.00E-125	gi|31415582	genbank nr nucleotide
ILS-702	EB762747	Pfam identified Ribosomal L40e family; Ubiquitin family	1.00E-43	SNOG_05870	*Stagonospora nodorum genes*
ILS-711	EB762724	40S RIBOSOMAL PROTEIN S22 (S15A) (YS24)	3.00E-81	SNOG_02003	*Stagonospora nodorum genes*
ILS-832	EB762731	Pfam identified Ribosomal L10	2.00E-35	SNOG_03153	*Stagonospora nodorum genes*
ILS-867	EB762734	Pfam identified Ribosomal protein L6e	2.00E-11	SNOG_14064	*Stagonospora nodorum genes*
ILS-976	EB762739	28S large subunit ribosomal RNA gene	6.00E-91	gi|31415582	genbank nr nucleotide
ILS-274R	EB762703	28S ribosomal RNA gene, partial sequence	9.00E-71	gi|45775613	genbank nr nucleotide
ILS-336	EB762706	28S ribosomal RNA gene	2.00E-128	gi|82398629	genbank nr nucleotide
ILS-339	EC091595	ribosomal RNA	5.00E-104	gi|82504131	genbank nr nucleotide
ILS-569	EC091599	ribosomal RNA	3.00E-56	gi|82504130	genbank nr nucleotide
ILS-591	EC091600	ribosomal RNA	3.00E-108	gi|82504131	genbank nr nucleotide
	translation				
ILS-656	EB762722	Pfam identified NAC domain	2.00E-75	SNOG_07434	*Stagonospora nodorum genes*
	translational control				
ILS-1276	EC091588	cytochrome oxidase	7.00E-140	gi|31790108	genbank nr nucleotide
ILS-1490F	EC091591	cytochrome oxidase	2.00E-140	gi|31790108	genbank nr nucleotide
INTERACTION WITH THE ENVIRONMENT					

(Table 1) contd…..

EST	GenBank Accession Number	Match Description	E-Value	Match Accession Number	Match Database
ILS-1220	EB762685	glycoprotein CIH1	1.00E-12	gi|2369760	genbank nr protein
CELL CYCLE AND DNA PROCESSING					
	cell cycle				
ILS-1484	EB762743	histone H4	3.00E-22	gi|47027020	genbank nr protein
ILS-399R	EB762711	Pfam identified Core histone H2A/H2B/H3/H4	6.00E-73	SNOG_14611	*Stagonospora nodorum genes*
ILS-25R	EB762702	Pfam identified Core histone H2A/H2B/H3/H4	1.00E-165	SNOG_14611	*Stagonospora nodorum genes*
CELL RESCUE DEFENSE AND VIRULENCE					
	disease virulence and defense				
ILS-312F	EB762744	sirodesmin biosynthetic gene cluster	2.00E-140	gi|46403045	genbank nr nucleotide
ILS-598	EB762745	Exopolygalacturonase precursor (ExoPG)	8.00E-16	gi|2499717	genbank nr protein
ILS-718R	EB762725	Woronin body major protein	1.00E-29	gi|2197050	genbank nr protein
CELLULAR TRANSPORT, TRANSPORT FACILITATION AND TRANSPORT ROUTES					
	transport facilitation				
ILS-374	EB762708	mitochondrial gene for ATPase subunit 6	1.00E-45	gi|12869	genbank nr nucleotide
ILS-564F	EB762717	AHA2; ATPase AT4G30190	4.00E-12	gi|30688579	genbank nr nucleotide
ILS-877	EB762736	mitochondrial gene for ATPase subunit 6	5.00E-73	gi|12869	genbank nr nucleotide
ILS-772R	EB762728	Pfam identified ABC transporter	0	SNOG_05861	*Stagonospora nodorum genes*
TRANSCRIPTION					
	RNA sysnthesis				
ILS-409R	EB762712	Pfam identified Fungal Zn(2)-Cys(6) binuclear cluster domain	4.00E-12	SNOG_12086	*Stagonospora nodorum genes*
	mRNA processing				
ILS-1265	EC091587	Pfam identified KH domain	7.3269E-24	SNOG_07507	*Stagonospora nodorum genes*
ILS-1479	EC091590	Pfam idenified Poly-adenylate binding protein	4.51E-43	SNOG_14350	*Stagonospora nodorum genes*
PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)					
	nucleic acid binding				
ILS-1123	EB762680	single-stranded DNA binding protein, putative	4.00E-16	gi|57222972	genbank nr protein
	protein binding				
ILS-644R	EB762720	Pfam identified WD domain, G-beta repeat	3.00E-59	SNOG_01594	*Stagonospora nodorum genes*
ILS-644F	EC091601	Pfam identified WD domain, G-beta repeat	0	SNOG_01594	*Stagonospora nodorum genes*

genome sequence data were used for comparison [15]. In total, 7% of the Infected leaf library ESTs were identified as fungal. Although a subtraction step was included in the construction of this library, the intent was to enrich for infection specific sequences. Two driver RNA pools, one from uninfected *B. napus* leaf tissue and one from cultured *L. maculans*, were used and consequently there has not been significant enrichment for fungal sequences. The proportion if fungal sequences observed in this library is within the range observed (1.8-8.6%) from a number of other plant-fungal interaction EST collections [20].

Attempts were made to gather large scale expression data using microarrays generated from the same cDNAs that produced the ESTs discussed here. Unfortunately this analysis yielded little information, possibly due in part to the relatively low amounts of fungal tissue and thus fungal RNA recovered from the infected leaf tissues used in this analysis. Only five fungal transcripts were found to be significantly upregulated at any time over an eight day timecourse (LM6-306; LM6-329F; LM6-371; LM9-298; ILS-1227; unpublished results). Only ILS-1227 had a predicted function, it exhibits similarity to a large subunit ribosomal RNA gene. It may prove more productive to track expression changes of a few candidate genes using an alternate technique such as quantitative RT-PCR.

Functional annotation based on sequence similarity was attempted for all ESTs described here. Functions could be putatively assigned to ~one-quarter of the ESTs collected from cultured fungus. The sequences collected from cultured mycelia represent genes involved in a diverse array of cellular processes and provide a resource for studying various aspects of *L. maculans* biology.

Table 2. ***Leptosphaeria maculans*** **ESTs Grouped into Functional Categories**

EST	Description	Related Taxon	E-Value
METABOLISM			
	Amino acid metabolism		
LM6-5R	glutamine synthetase	*Tuber borchii*	6.E-94
LM6-115R	glutamine synthetase	*Tuber borchii*	3.E-85
LM6-136R	glutamine synthetase	*Tuber borchii*	6.E-45
LM9-206R	putative dimethyl-allyl-tryptophan-synthase	*Claviceps purpurea*	3.E-05
LM6-269R	methionine synthase	*Cladosporium fulvum*	2.E-47
LM6-291R	methionine synthase	*Cladosporium fulvum*	2.E-39
LM9-335R	related to BENZOYLFORMATE DECARBOXYLASE	*Neurospora crassa*	3.E-47
LM9-210R	Homogentisate 1,2-dioxygenase	*Aspergillus nidulans*	4.E-56
LM9-210F	Homogentisate 1,2-dioxygenase	*Aspergillus nidulans*	2.E-45
LM9-206F	dimethylallyl tryptophan synthase	*Penicillium paxilli*	2.E-08
LM6-34	S-ADENOSYLMETHIONINE SYNTHETASE	*Neurospora crassa*	3.E-21
LM9-364	Homocysteine S-methyltransferase	*Escherichia coli K12*	5.E-21
	nucleotide metabolism		
LM9-89	Uricase (Urate oxidase)	*Aspergillus flavus*	1.E-112
	C-compound and carbohydrate metabolism		
LM9-250F	Aldehyde reductase I (Alcohol dehydrogenase [NADP+])	*Sporidiobolus salmonicolor*	9.E-47
LM9-250R	Aldehyde reductase I (Alcohol dehydrogenase [NADP+])	*Sporidiobolus salmonicolor*	7.E-35
LM9-150R	beta-galactosidase	*Cloning vector pWD42-09*	6.E-44
LM9-27	beta-galactosidase alpha fragment	*Shuttle vector pSET1*	3.E-54
LM9-279	beta-galactosidase alpha fragment	*Shuttle vector pSET2*	4.E-54
LM9-123R	beta-galactosidase alpha fragment	*Shuttle vector pSET3*	2.E-49
LM6-195R	MALATE SYNTHASE, GLYOXYSOMAL	*Neurospora crassa*	3.E-98
LM6-199R	Enolase	*Alternaria alternata*	1.E-103
LM6-199F	Enolase	*Alternaria alternata*	2.E-74
LM6-140R	Etf1p: Elongation factor 2	*Saccharomyces cerevisiae*	1.E-74
LM6-97R	GLUCOSE-6-PHOSPHATE 1-DEHYDROGENASE (G6PD)	*Neurospora crassa*	1.E-47
LM6-230F	beta-glucosidase 5	*Coccidioides posadasii*	4.E-23
LM6-193F	beta-glucosidase 5	*Coccidioides posadasii*	4.E-14
ILS-251F	beta-glucosidase 5	*Coccidioides posadasii*	8.E-16
LM6-304	chitin binding protein	*Magnaporthe grisea*	5.E-61
LM6-92R	chitinase	*Aspergillus fumigatus*	4.E-06
LM6-12R	phosphomannomutase homolog	*Schizosaccharomyces pombe*	1.E-59
LM9-280R	putative mannitol dehydrogenase	*Alternaria alternata*	2.E-79
LM9-280F	putative mannitol dehydrogenase	*Alternaria alternata*	2.E-59
LM6-104F	UDP-glucose 4-epimerase	*Bacillus anthracis str. Ames*	6.E-09
	lipid, fatty acid and isoprenoid metabolism		
LM9-159	delta-9 fatty acid desaturase	*Ajellomyces capsulatus*	2.E-78
LM9-171R	delta(24)-sterol C-methyltransferase	*Neurospora crassa*	1.E-65
LM9-171F	delta(24)-sterol C-methyltransferase	*Neurospora crassa*	6.E-25
LM9-256F	C-4 methyl sterol oxidase	*Candida albicans*	5.E-22
LM6-36F	phospholipase	*Aspergillus oryzae*	4.E-18
LM9-175R	Glyoxaloxidase 2	*Ustilago maydis*	1.E-13
LM9-376F	enoyl reductase	*Aspergillus terreus*	2.E-05
LM6-74	acyl-Coenzyme A thioesterase 2, mitochondrial	*Mus musculus*	2.E-13
LM6-138R	3-hydroxybutyryl-CoA dehydratase	*Aeropyrum pernix*	2.E-07
	secondary metabolism		
LM6-161F	amine oxidase	*Aspergillus oryzae*	6.E-46
LM6-161R	copper amine oxidase	*Aspergillus niger*	5.E-66
LM9-164F	CalO1	*Micromonospora echinospora*	4.E-17
LM9-245F	CalO2	*Micromonospora echinospora*	1.E-15
LM6-151	carotene cyclase	*Gibberella fujikuroi*	1.E-47
LM9-164R	desmethyl DIF-1 methyltransferase DmtA	*Dictyostelium discoideum*	2.E-05
ENERGY			
	glycolysis and gluconeogenesis		
LM6-281R	fructose-1,6-bisphosphatase	*Aspergillus nidulans*	1.E-60
LM6-281F	fructose-bisphosphatase	*Aspergillus oryzae*	2.E-47

(Table 2) contd…..

EST	Description	Related Taxon	E-Value
LM9-257F	alcohol dehydrogenase 3	*Yarrowia lipolytica*	7.E-45
LM9-257R	alcohol dehydrogenase I	*Neurospora crassa*	7.E-60
LM6-303F	alcohol dehydrogenase I	*Neurospora crassa*	4.E-55
LM6-319F	alcohol dehydrogenase I	*Neurospora crassa*	9.E-41
LM6-303R	alcohol dehydrogenase I	*Neurospora crassa*	1.E-60
LM6-319R	alcohol dehydrogenase I	*Neurospora crassa*	2.E-35
LM6-277	benzyl alcohol dehydrogenase	*Acinetobacter calcoaceticus*	7.E-33
LM6-118R	phosphoenolpyruvate carboxykinase	*Emericella nidulans*	4.E-74
LM6-40R	phosphoenolpyruvate carboxykinase	*Emericella nidulans*	2.E-67
LM6-120R	phosphoenolpyruvate carboxykinase	*Emericella nidulans*	6.E-59
LM6-208R	phosphoenolpyruvate carboxykinase	*Emericella nidulans*	8.E-94
LM9-144R	phosphoenolpyruvate carboxykinase	*Emericella nidulans*	1.E-40
	fermentation		
LM6-6	pyruvate decarboxylase	*Aspergillus oryzae*	1.E-159
LM9-75R	pyruvate decarboxylase	*Aspergillus oryzae*	1.E-73
LM6-192F	pyruvate decarboxylase	*Aspergillus oryzae*	2.E-61
LM6-111F	pyruvate decarboxylase	*Aspergillus oryzae*	3.E-57
LM6-270F	pyruvate decarboxylase	*Aspergillus oryzae*	3.E-46
LM6-310F	pyruvate decarboxylase	*Aspergillus oryzae*	2.E-28
LM6-111R	pyruvate decarboxylase	*Emericella nidulans*	9.E-81
	Fermentation (continued)		
LM6-270R	pyruvate decarboxylase	*Emericella nidulans*	1.E-78
LM9-270R	pyruvate decarboxylase	*Emericella nidulans*	1.E-71
LM6-13	pyruvate decarboxylase	*Emericella nidulans*	1.E-116
	respiration		
LM6-96R	URF 1	*Emericella nidulans*	2.E-49
LM6-135R	URF 2	*Emericella nidulans*	4.E-18
	conversion and regeneration		
LM9-330F	NADH-dependent flavin oxidoreductase	*Deinococcus radiodurans*	1.E-18
CELL CYCLE AND DNA PROCESSING			
	cell cycle		
LM9-166R	ubiquitin-like fusion protein SonA	*Dictyostelium discoideum*	3.E-13
LM9-177R	14-3-3-like protein	*Paracoccidioides brasiliensis*	6.E-39
LM6-171R	spindle assembly checkpoint protein SLDA	*Emericella nidulans*	6.E-61
LM6-253R	histone H1	*Emericella nidulans*	8.E-24
ILS-399R	histone H2A	*Aspergillus niger*	2.E-43
ILS-24R	histone H2A	*Aspergillus niger*	7.E-42
	DNA repair		
LM6-184R	homologue of the REX1-S protein which is involved in DNA repair	*Chlamydomonas reinhardtii*	5.E-07
TRANSCRIPTION			
	RNA sysnthesis		
LM9-76R	Dod ND5 i1 grp ID protein	*Podospora anserina*	7.E-32
LM9-263R	TRI15	*Fusarium sporotrichioides*	6.E-22
LM9-97F	Sulfur metabolite repression control protein sconB	*Emericella nidulans*	1.E-36
LM6-274F	myc-type bHLH transcription factor Cph2	*Emericella nidulans*	4.E-09
LM6-95R	SipA3	*Aspergillus nidulans*	1.E-34
ILS-409R	C6 transcription factor (Fcr1)	*Aspergillus fumigatus Af293*	7.E-45
ILS-325R	transcription factor	*Hypocrea jecorina*	6.E-24
ILS-952R	related to RNA helicase MSS116	*Neurospora crassa*	5.E-36
PROTEIN SYNTHESIS			
	aminoacyl-tRNA-synthetases		
LM9-343F	seryl-tRNA synthetase; Ses1p	*Saccharomyces cerevisiae*	2.E-22
	Ribosome biogenesis		
LM9-281	60S acidic ribosomal protein P1	*Alternaria alternata*	3.E-26
LM9-342	60S acidic ribosomal protein P1	*Alternaria alternata*	2.E-23
LM6-45F	60S ribosomal protein L3	*Neurospora crassa*	1.E-73
LM9-72	60S ribosomal protein L44	*Pichia jadinii*	1.E-51
	Ribosome biogenesis (contimued)		

(Table 2) contd…..

EST	Description	Related Taxon	E-Value
LM9-366R	homology to rat L11 and E. coli L5; Rpl11bp ribosomal protein	*Saccharomyces cerevisiae*	5.E-64
LM6-45R	large subunit ribosomal protein L3	*Aspergillus fumigatus*	1.E-90
LM6-46	putative ribosomal protein S28	*Aspergillus fumigatus*	1.E-73
LM6-206F	ribosomal L10 protein	*Paracoccidioides brasiliensis*	2.E-63
LM9-172	ribosomal L10 protein	*Paracoccidioides brasiliensis*	9.E-41
LM9-359R	ribosomal L10 protein	*Paracoccidioides brasiliensis*	4.E-57
LM6-43R	ribosomal protein P0	*Podospora anserina*	8.E-60
LM9-333R	40s ribosomal protein S29	*Schizosaccharomyces pombe*	6.E-17
LM6-308R	probable ribosomal protein 10, cytosolic	*Neurospora crassa*	1.E-64
ILS-1309	ribosomal protein L34-like protein	*Neurospora crassa*	1.E-37
ILS-204	probable ribosomal protein L38	*Neurospora crassa*	2.E-19
ILS-832	ribosomal L10 protein	*Paracoccidioides brasiliensis*	5.E-92
ILS-183	40S ribosomal protein S17	*Neurospora crassa*	1.E-51
ILS-197	40S ribosomal protein S26E	*Neurospora crassa*	4.E-51
ILS-711	40S ribosomal protein S22	*Neurospora crassa*	2.E-38
ILS-53	40S ribosomal protein S3	*Neurospora crassa*	4.E-19
ILS-180	homology to rat S4 and human S4; Rps4bp	*Saccharomyces cerevisiae*	2.E-48
LM9-366F	ribosomal Protein, Large subunit (22.8 kD) rpl-11.2	*Caenorhabditis elegans*	2.E-14
	translation		
LM6-318R	elongation factor 2 (EF-2)	*Neurospora crassa*	3.E-08
LM6-185R	elongation factor 1-alpha (EF-1-ALPHA)	*Schizophyllum commune*	3.E-88
LM6-185F	elongation factor 1-alpha (EF-1-ALPHA)	*Ajellomyces capsulata*	4.E-55
LM6-276	elongation factor 1-alpha (EF-1-ALPHA)	*Aureobasidium pullulans*	1.E-136
LM9-195R	elongation factor 1-alpha (EF-1-ALPHA)	*Aureobasidium pullulans*	1.E-90
LM9-138R	elongation factor 1-alpha (EF-1-ALPHA)	*Puccinia graminis*	3.E-77
LM9-231R	elongation factor 2 (EF-2)	*Neurospora crassa*	3.E-90
LM9-231F	elongation factor 2 (EF-2)	*Neurospora crassa*	3.E-29
LM6-140F	elongation factor 2 (EF-2)	*Neurospora crassa*	5.E-25
PROTEIN FATE (folding, modification, destination)			
	protein folding and stabilization		
LM6-154	related to Hsp90 associated co-chaperone	*Neurospora crassa*	5.E-39
LM9-276R	protein disulfide isomerase precursor (PDI)	*Humicola insolens*	5.E-41
	protein targeting, sorting and translocation		
LM6-194F	probable VpsA protein	*Neurospora crassa*	1.E-43
LM6-261R	related to dynamin-like protein	*Neurospora crassa*	1.E-12
	protein modification		
LM6-290F	poly-ubiquitin	*Magnaporthe grisea*	3.E-47
LM6-33F	alanyl dipeptidyl peptidase	*Aspergillus oryzae*	1.E-39
LM6-33R	alanyl dipeptidyl peptidase	*Aspergillus oryzae*	6.E-32
LM9-96	protein-L-isoaspartate (D-aspartate) O-methyltransferase 1	*Mus musculus*	4.E-56
LM9-178	UBI1	*Emericella nidulans*	5.E-66
LM9-278R	UBI4 locus	*Kluyveromyces marxianus*	1.E-81
LM6-290R	similar to polyubiquitin	*Bos taurus*	6.E-75
	protein degradation		
LM9-349F	alkaline protease	*Aspergillus fumigatus*	2.E-22
LM9-85F	elastase-like serine protease	*Metarhizium anisopliae*	5.E-45
LM9-30F	elastase-like serine protease	*Metarhizium anisopliae*	1.E-30
LM9-66F	elastase-like serine protease	*Metarhizium anisopliae*	4.E-17
LM9-209R	peptidase	*Streptomyces coelicolor*	3.E-42
LM9-215F	probable 26S ATP/ubiquitin-dependent proteinase chain S4	*Neurospora crassa*	7.E-26
LM9-215R	probable 26S ATP/ubiquitin-dependent proteinase chain S5	*Neurospora crassa*	2.E-61
LM9-349R	protease	*Aspergillus fumigatus*	2.E-28
LM9-240R	serine protease precursor	*Fusarium oxysporum*	7.E-23
LM9-85R	serine protease precursor	*Fusarium oxysporum*	4.E-24
LM9-205	trypsin-like protease	*Phaeosphaeria nodorum*	5.E-85
LM6-233	trypsin-like protease	*Phaeosphaeria nodorum*	5.E-85
LM6-243R	MepB	*Aspergillus fumigatus*	3.E-77
LM6-243F	MepB	*Aspergillus fumigatus*	1.E-39

(Table 2) contd…..

EST	Description	Related Taxon	E-Value
LM6-90F	MepB	*Aspergillus fumigatus*	6.E-36
LM6-90R	MepB	*Aspergillus fumigatus*	2.E-10
LM9-34R	proteasome regulatory subunit 12	*Hypocrea jecorina*	7.E-83
LM9-274R	protease 1	*Pneumocystis carinii*	5.E-05
PROTEIN WITH BINDING FUNCTION OR COFACTOR REQUIREMENT (structural or catalytic)			
	nucleotide binding		
LM9-145R	rho3 protein	*Hypocrea jecorina*	1.E-78
LM9-259R	Ras-related protein ypt1	*Schizosaccharomyces pombe*	2.E-65
CELLULAR TRANSPORT, TRANSPORT FACILITATION AND TRANSPORT ROUTES			
	transported compounds (substrates)		
LM6-256F	potassium transporter	*Kluyveromyces lactis*	5.E-13
LM6-160F	aromatic amino acid and leucine permease	*Penicillium chrysogenum*	4.E-30
LM9-273R	putative peptide transporter	*Aspergillus fumigatus*	1.E-27
LM6-31F	putative peptide transporter	*Aspergillus fumigatus*	2.E-20
LM6-31R	probable peptide transporter Ptr2	*Schizosaccharomyces pombe*	6.E-21
LM9-137	thioredoxin	*Emericella nidulans*	2.E-13
LM9-76F	cytochrome b	*Mycosphaerella fijiensis*	8.E-38
LM6-149	cytochrome b	*Mycosphaerella fijiensis*	9.E-38
	transported compounds (substrates) (continued)		
LM6-238R	cytochrome b	*Venturia inaequalis*	7.E-23
LM9-344R	cytochrome b	*Venturia inaequalis*	8.E-18
LM6-109	cytochrome c	*Curvularia lunata*	4.E-50
LM9-49F	ADP-ATP translocase	*Gaeumannomyces graminis*	9.E-43
LM6-162F	Dal5p subfamily of the major facilitator family; Yil166cp	*Saccharomyces cerevisiae*	8.E-06
LM6-152F	cytochrome P450, putative	*Arabidopsis thaliana*	3.E-14
LM6-285F	cytochrome P450, putative	*Arabidopsis thaliana*	2.E-14
	transport facilitation		
LM6-189F	ATP synthase A chain	*Cochliobolus heterostrophus*	1.E-55
LM9-239F	ATP synthase A chain	*Cochliobolus heterostrophus*	4.E-57
LM6-106F	ATP synthase A chain	*Cochliobolus heterostrophus*	7.E-40
LM6-189R	ATP synthase A chain	*Cochliobolus heterostrophus*	2.E-10
LM9-188R	H+-transporting two-sector ATPase	*Emericella nidulans*	4.E-16
LM6-8F	vacuolar ATP synthase catalytic subunit A	*Neurospora crassa*	4.E-58
ILS-374	ATP synthase A chain (Protein 6)	*Cochliobolus heterostrophus*	1.E-106
ILS-877	ATP synthase A chain (Protein 6)	*Cochliobolus heterostrophus*	1.E-47
ILS-718R	probable V-ATPase, 20K chain	*Neurospora crassa*	4.E-33
ILS-203	H+-transporting two-sector ATPase	*Emericella nidulans*	6.E-16
LM9-113	plasma membrane H+-ATPase	*Aspergillus fumigatus*	3.E-40
LM6-150R	putative sugar transporter	*Emericella nidulans*	1.E-05
LM6-203R	putative sugar transporter	*Emericella nidulans*	8.E-09
LM6-8R	VmaA protein	*Emericella nidulans*	2.E-35
LM6-251R	mitochondrial F1 ATP synthase beta subunit	*Arabidopsis thaliana*	5.E-05
LM9-190R	mitochondrial F1 ATP synthase beta subunit	*Arabidopsis thaliana*	7.E-06
LM9-266	mitochondrial F1 ATP synthase beta subunit	*Arabidopsis thaliana*	2.E-07
CELLULAR COMMUNICATION/SIGNAL TRANSDUCTION MECHANISM			
	intracellular signalling		
LM9-259F	secretion related GTPase (SrgB)	*Aspergillus niger*	5.E-20
LM9-147	GTPase rho1	*Blumeria graminis*	1.E-100
LM6-196R	WSC4 homolog	*Kluyveromyces marxianus*	1.E-15
ILS-1096	probable protein phosphatase 2C	*Neurospora crassa*	3.E-06
CELL RESCUE, DEFENSE AND VIRULENCE			
	stress response		
LM6-147F	catalase B	*Aspergillus fumigatus*	1.E-20
LM6-147R	catalase	*Blumeria graminis*	4.E-47
LM9-44R	opsin	*Leptosphaeria maculans*	5.E-95
LM9-44F	opsin	*Leptosphaeria maculans*	3.E-34
LM9-227R	glutathione S-transferase	*Botryotinia fuckeliana*	8.E-18
LM6-47F	thioredoxin reductase	*Neurospora crassa*	1.E-16

(Table 2) contd…..

EST	Description	Related Taxon	E-Value
LM6-173F	heat shock 70 kDa protein	*Ajellomyces capsulatus*	3.E-14
	stress response (continued)		
LM6-173R	heat shock 70 kDa protein	*Ajellomyces capsulatus*	2.E-82
LM6-170R	heat shock 70 kDa protein	*Ajellomyces capsulatus*	4.E-75
LM6-245R	heat shock protein CLPA	*Paracoccidioides brasiliensis*	2.E-42
LM6-245F	heat shock protein CLPA	*Paracoccidioides brasiliensis*	4.E-22
LM6-25F	heat shock protein CLPA	*Paracoccidioides brasiliensis*	5.E-21
LM6-122R	putative glutathione S-transferase	*Salmonella typhimurium*	8.E-29
	disease, virulence and defense		
LM6-126R	TOXD protein	*Cochliobolus carbonum*	2.E-34
LM9-54	TOXD protein	*Cochliobolus carbonum*	2.E-49
LM9-376R	TOXD protein	*Cochliobolus carbonum*	2.E-19
LM9-204	cyclophilin	*Ricinus communis*	2.E-66
	detoxification		
LM9-357	manganese superoxide dismutase	*Emericella nidulans*	2.E-82
LM6-181R	cytochrome P450 oxidoreductase	*Gibberella fujikuroi*	6.E-68
LM9-202	putative epoxide hydrolase	*Rhodotorula mucilaginosa*	3.E-06
LM6-220F	multidrug resistance protein 1	*Filobasidium floriforme*	8.E-47
LM6-127F	multidrug resistance protein 2	*Filobasidium floriforme*	3.E-31
LM6-220R	multidrug resistance protein 1	*Aspergillus fumigatus*	8.E-64
LM6-127R	multidrug resistance protein 2	*Aspergillus fumigatus*	3.E-12
LM9-194R	putative cyanide hydratase	*Leptosphaeria maculans*	1.E-100
LM6-122F	glutathione S transferase	*Schizosaccharomyces pombe*	2.E-38
LM9-42	putative thiol-specific antioxidant protein Tsa1	*Ajellomyces capsulatus*	4.E-74
LM6-225R	epoxide hydrolase	*Caulobacter crescentus*	4.E-23
INTERACTION WITH THE ENVIRONMENT			
ILS-1220	glycoprotein CIH1	*Colletotrichum lindemuthianum*	1.E-16
ILS-1006	putative immunoreactive protein	*Magnaporthe grisea*	3.E-07
BIOGENESIS OF CELLULAR COMPONENTS			
LM6-328F	actin	*Neurospora crassa*	1.E-26
LM6-328R	cytoplasmic actin	*Cavia porcellus*	7.E-47
LM6-218	alpha-tubulin	*Rhynchosporium secalis*	2.E-44
LM9-368R	GPI-anchored membrane protein; Dcw1p	*Saccharomyces cerevisiae*	7.E-08
LM6-148R	GEL1 protein	*Aspergillus fumigatus*	8.E-40
LM6-267R	GEL1 protein	*Aspergillus fumigatus*	8.E-19
LM6-24	lectin-B - Virginian pokeweed	*Phytolacca americana*	3.E-11
LM6-315R	similar to p48 eggshell protein	*Rattus norvegicus*	5.E-24

Although the LM EST library was derived from cultured fungal mycelia, it is possible that some of the sequences presented here may still be for successful infection. Genes potentially involved ion protection form host defenses (e.g. LM9-194R; cyanide hydratase) or reactive oxygen species (e.g. LM6-147; catalase) have been identified, and general metabolic processes may also be have a vital role in providing nutrition in the resource limited plant apoplast [21-23]. The ILS infected leaf library can be used to identify genes which are more likely to be required for pathogenesis. A handful of ESTs identified in this study may be involved in pathogenesis and would be good candidates for further study These include: ILS-1220, similar to a proline rich glycoprotein from *Colletotrichum lindemuthianum* involved in preventing recognition by its host [24]; ILS-312F, a near perfect match for the SirD gene of the sirodesmin biosynthetic cluster, which may contribute to virulence [25, 26]; ILS-598, similar to an exopolygalacturonase precursor from *Cochliobolus carbonum* [27]; ILS-718R, similar to a Woronin body major protein from *N. crassa* involved in plugging septal pores [28] and a related protein in *Magnaporthe grisea* required for invasive growth and host colonization of fungal hyphae [29]. Another promising target is ILS409R, a putative zinc finger Zn(2)-Cys(6) transcription factor. Suggestively, a mutagenized strain of *L. maculans* bearing a T-DNA insert 1kb upstream of the ILS-409R EST sequence grows normally in culture but has lost the ability to successfully infect the susceptible *B. napus* cultivar Westar (unpublished results).

The libraries presented here provide a major increase in the amount of sequence information from *L. maculans* which is a prerequisite for all future functional genomic approaches in this important pathogen. A total of 551 new cDNAs from the LM library have been identified. The genes represented

in this set have putative functions that range over all aspects of fungal biology incuding structural (e.g. histones, ribosomal proteins), regulatory (transcription factors, translation initiation factors) and metabolic (proteases, alcohol dehydrogenases) roles. We tested various methods for distinguishing between plant and fungal sequences in mixed sequence libraries, and found a TBLASTX based comparison using custom reference sequence databases to be most robust. The procedures outlined here for separating sequences in a mixed library may be of use to others.

ACKNOWLEDGEMENTS

We would like to thank Dr. Andrew Sharpe for his advice and assistance in the preparation and sequencing of the cDNA libraries described here. We would also like to thank Paul Stothard at the Canadian Bionformatics Help Desk for assistance in performing BLAST comparisons. We acknowledge the assistance of Bayer Cropscience and especially Dr. Malcolm Devine for participating in an NSERC industrial postgraduate scholarship. Funding for this work was provided in part from grants from the Alberta Agricultural Research Institute, Farming for the Future program and NSERC to AGG.

REFERENCES

[1] Howlett BJ, Idnurm A, Pedras MS. *Leptosphaeria maculans*, the causal agent of blackleg disease of *Brassicas*. Fungal Genet Biol 2001; 33: 1-14.

[2] Petrie GA. Occurrence of a highly virulent strain of blackleg (Leptosphaeria maculans) on rape in Saskatchewan (1975-77). Can Plant Dis Surv 1978; 65: 35-41.

[3] Hammond KE, Lewis BG. The establishment of systemic infection in leaves of oilseed rape by *Leptosphaeria maculans*. Plant Pathol 1987a.; 36: 135-147.

[4] Chen CY, Howlett BJ. Rapid necrosis of guard cells is associated with the arrest of fungal growth in leaves of Indian mustard (*Brassica juncea*) inoculated with avirulent isolates of *Leptosphaeria maculans*. Physiol Mol Plant Pathol 1996;48: 73-81.

[5] Hammond KE, Lewis BG. Variation in stem infections caused by aggressive and non-aggressive isolates of *Leptosphaeria maculans* on *Brassica napus* var. *oleifera*. Plant Pathol 1987b; 36: 53-65.

[6] Idnurm A, Howlett BJ. Isocitrate lyase is essential for pathogenicity of the fungus *Leptosphaeria maculans* on canola (*Brassica napus*). Eukaryot Cell 2002; 1: 719-724.

[7] Idnurm A, Howlett BJ. Analysis of loss of pathogenicity mutants reveals that repeat-induced point mutations can occur in the Dothideomycete *Leptosphaeria maculans*. Fungal Genet Biol 2003; 39: 31-7.

[8] Gout L, Fudal I, Kuhn M-L, *et al.* Lost in the middle of nowhere: the *AvrLm1* avirulence gene of the Dothideomycete *Leptosphaeria maculans*. Mol Microbiol 2006; 60: 67-80.

[9] Pandey A, Lewitter F. Nucleotide sequence databases: a goldmine for biologists. Trends Biochem Sci 1999; 24: 276-280.

[10] Qutob D, Hraber PT, Sobral BWS, Gijzen M. Comparative analysis of expressed sequences in *Phytophthora sojae*. Plant Physiol 2000; 123: 243-253.

[11] Hraber PT, Weller JW. On the species of origin: diagnosing the source of symbiotic transcripts. Genome Biol 2001; 2: research0037.1-0037.14.

[12] Kim S, Ahn IP, Lee YH. Analysis of genes expressed during rice-*Magnaporthe grisea* interactions. Mol. Plant-Microbe Interact 2001; 14: 1340-1346.

[13] Ronning CM, Stegalkina SS, Ascenzi RA, *et al.* Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 2003; 131: 419-429.

[14] Li R, Rimmer R, Buchwaldt L, *et al.* Interaction of *Sclerotinia sclerotiorum* with a resistant Bra*ssica napus* cultivar: expressed sequence tag analysis identifies genes associated with fungal pathogenesis. Fungal Genet Biol 2004; 41: 735–753.

[15] Cramer RA, La Rota CM, Cho Y, *et al.* Bioinformatic analysis of expressed sequence tags derived from a compatible *Alternaria brassicicola–Brassica oleracea* interaction. Mol Plant Path 2006; 7: 113-124.

[16] Fristensky B, Balcerzak M, He D, Zhang P. Expressed sequence tags from the defence response of *Brassica napus* to *Leptosphaeria maculans*. Mol Plant Path online [serial on the internet] 1999; [cited 2008 March 4] Available from: http://www.bspp.org.uk/mppol/1999/0301FRISTENSKY.

[17] Sharpe AG, Parkin IAP, Keith DJ, Lydiate DJ. Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 1995; 38: 1112-1121.

[18] Altschul SF, Madden TL, Schaeffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997; 25: 3389–3402.

[19] White O, Dunning T, Sutton G, Adams M, Venter JC, Fields C. A quality control algorithm for DNA sequencing projects. Nucleic Acids Res 1993; 21: 3829-3838.

[20] Hsiang T, Goodwin PH. Distinguishing plant and fungal sequences in ESTs from infected plant tissues. J Microbiol Methods 2003; 54: 339–351.

[21] Coleman M, Henricot B, Arnau J, Oliver RP. Starvation-induced genes of the tomato pathogen *Cladosporium fulvum* are also induced during growth *in planta*. Mol Plant-Microbe Interact 1997; 10: 1106-1109.

[22] Solomon PS, Nielsen PS, Clark AJ, Oliver RP. Methionine synthase, a gene required for methionine synthesis, is expressed *in planta* by *Cladosporium fulvum*. Mol Plant Pathol 2000; 1: 315-323.

[23] Solomon PS, Oliver RP. Evidence that gamma-aminobutyric acid is a major nitrogen source during Cladosporium fulvum infection of tomato. Planta 2002; 214: 414-420.

[24] Perfect SE, O'Connell RJ, Green EF, Doering-Saad C, Green JR. Expression cloning of a fungal proline-rich glycoprotein specific to the biotrophic interface formed in the *Colletotrichum*-bean interaction. Plant J 1998; 15: 273-279.

[25] Gardiner DM, Cozijnsenjavascript:popRef('a1') AJ, Wilsonjavascript:popRef('a1') LM, Pedras MSC, Howlett BJ. The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. Mol Microbiol 2004; 53: 1307-1318.

[26] Sock J, Hoppe HH. Pathogenicity of Sirodesmin-deficient Mutants of *Phoma lingam*. J Phytopathol 1999; 147: 169-174.

[27] Scott-Craig JS, Cheng Y, Cervone F, De Lorenzo G, Pitkin JW, Walton JD. Targeted Mutants of *Cochliobolus carbonum* Lacking the Two Major Extracellular Polygalacturonases. Appl Environ Microbiol 1998; 64: 1497-1503.

[28] Tenney K, Hunt I, Sweigard J, *et al. hex-1,* a Gene Unique to Filamentous Fungi, Encodes the Major Protein of the Woronin Body and Functions as a Plug for Septal Pores. Fungal Genet Biol 2000; 31: 205-217.

[29] Soundararajan S, Jedd G, Li X, Ramos-Pamploña M, Chua NH, Naqvi NI. Woronin Body Function in *Magnaporthe grisea* Is Essential for Efficient Pathogenesis and for Survival during Nitrogen Starvation Stress. Plant Cell 2004; 16: 1564-1574.

Received: January 17, 2008 Revised: March 5, 2008 Accepted: March 5, 2008