Generation of a wheat leaf rust, *Puccinia triticina*, EST database from stage-specific cDNA libraries

GUANGGAN HU1†, ROB LINNING1, BRENT MCCALLUM2, TRAVIS BANKS2, SYLVIE CLOUTIER2, YARON BUTTERFIELD3, JERRY LIU3, ROBERT KIRKPATRICK3, JEFF STOTT3, GEORGE YANG3, DUANE SMAILUS3, STEVEN JONES3, MARCO MARRA3, JACQUELINE SCHEIN3 AND GUUS BAKKEREN1*

1Pacific Agri-Food Research Centre, Agriculture and Agri-Food Canada, Highway 97, Summerland, BC, V0H 1Z0, Canada
2Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Road., Winnipeg, MB, R3T 2M9, Canada
3Michael Smith Genome Sciences Centre, 600 W 10th Avenue, Vancouver, BC, V5Z 4E6, Canada

SUMMARY

Thirteen cDNA libraries constructed from small amounts of leaf rust mRNA using optimized methods served as the source for the generation of 25 558 high-quality DNA sequence reads. Five life-cycle stages were sampled: resting urediniospores, urediniospores germinated over water or plant extract, compatible, interactive stages during appressorium or haustorium formation just before sporulation, and an incompatible interaction. mRNA populations were subjected to treatments such as full-length cDNA production, subtractive and normalizing hybridizations, and size selection methods combined with PCR amplification. Pathogen and host sequences from interactive libraries were differentiated *in silico* using cereal and fungal sequences, codon usage analyses, and by means of a partial prototype cDNA microarray hybridized with genomic DNAs. This yielded a non-redundant unigene set of 9760 putative fungal sequences consisting of 6616 singlets and 3144 contigs, representing 4.7 Mbp. At an *E*-value $\leq 10^{-5}$, 3670 unigenes (38%) matched sequences in various databases and collections but only 694 unigenes (7%) were similar to genes with known functions. In total, 296 unigenes were identified as most probably wheat and ten as rRNA sequences. Annotation rates were low for germinated urediniospores (4%) and appressoria (2%). Gene sets obtained from the various life-cycle stages appear to be remarkably different, suggesting drastic reprogramming of the transcriptome during these major differentiation processes. Redundancy within contigs yielded information about possible expression levels of certain genes among stages. Many sequences were similar to genes from other rusts such as *Uromyces* and *Melampsora* species; some of these genes have been implicated in pathogenicity and virulence.

INTRODUCTION

*Puccinia triticina* Eriks. (syn. *P. recondita* Rob. Ex Desm. f. sp. tritici Eriks. and Henn.) is a basidiomycete, biotrophic, obligate plant pathogen and the causal agent of leaf or brown rust on wheat (*Triticum aestivum*). It is an economically important pathogen in all wheat-growing areas of the world. Yield losses caused by leaf rust over a large area can be up to 20%, while individual fields can be destroyed when the disease is severe prior to heading (Kolmer, 2005; Saari and Prescott, 1985; Samborski, 1985). Chemical methods can be used to control the disease, but the most cost-effective strategy to reduce the incidence of the disease has been through the development of resistant wheat cultivars. Classical genetic data have demonstrated gene-for-gene interactions underlying incompatibility between *P. triticina* and its host plant (Browder, 1971) and 56 leaf rust resistant genes or alleles (*Lr*) are currently recognized world-wide (http://www.cdl.umn.edu; McCallum and Seto-Goh, 2005). However, due to the emergence of new virulent races, resistant cultivars eventually become susceptible to the rust. Novel methods to obtain durable resistance are needed, which requires a better understanding of fungal development and the interaction with its host to identify new targets for disease control.

*P. triticina* has a complex life cycle which includes five different spore types and two hosts: wheat and meadow rue, *Thalictrum speciosisissum*, its alternate host on which it completes its sexual stage (Horton et al., 2005; Littlefield, 1979). The asexual uredial stage on wheat is the economically important part of the life cycle which can progress from initial infection to sporulation.
within 10 days under warm, humid conditions, potentially leading to epidemics. During infection on wheat, the pathogen undergoes a high degree of morphological and physiological differentiation (Hu and Rijkenberg, 1998a). After landing on a wheat epidermis, the dikaryotic urediniospore germinates within hours and an emerging germ tube extends until a stomatal pore is encountered. An appressorium is formed on the stoma and an infection peg is produced to enter the plant. Topographical features of the stomatal cells and chemical signals may be involved in the infection structure formation (Collins and Read, 1997). Once the rust has forcibly gained entrance into the underlying cavity, the germ tube differentiates into a substomatal vesicle followed by a haustorial mother cell. This latter fungal structure makes the first significant contact with a host cell between 12 and 24 h post infection (hpi). This entails breaching the cell wall, forming an intercellular haustorium and establishing an intimate feeding relationship. In compatible interactions, the fungus ramifies through the tissue producing intercellular hyphae and intracellular haustoria (Hu and Rijkenberg, 1998b). During incompatible interactions, resistant reactions are mainly observed only after the first haustorium is formed (Niks and Dekens, 1991; Niks, 1983).

For plant fungal pathogens there is limited genome sequence information available, although the total genome sequence of several fungi has been made public and major new projects are underway or are being promoted (Yarden et al., 2003; Fungal Genome Initiative, http://www.broad.mit.edu/annotation/fungi/fgi/; Microbial Genome Sequencing Projects by the US Department of Energy, http://microbialgenome.org/organisms.shtml; suggested Microbial Genome Sequencing Priority List of the American Phytopathological Society, http://www.apsnet.org/). As a prelude to genome sequencing, the generation of an expressed sequence tag (EST) database represents a relatively inexpensive and rapid approach for gene discovery. In addition, ESTs can be used to design molecular markers such as microsatellites or simple sequence repeats (SSRs), to design and construct expression microarrays, and can aid with genome annotations or be used in comparative approaches to reveal relatedness between species (Marra et al., 1999; Skinner et al., 2001; Yarden et al., 2003). Gene discovery and expression profiling through the generation of an EST database is particularly useful for organisms with little or no genetic research history or those that are experimentally difficult to handle. EST collections from several plant fungal pathogens have been compiled (e.g. COGEME repository, http://cogeme.ex.ac.uk/; Soanes et al., 2002). However, collections from basidiomycete fungi are under-represented (Abu et al., 2004; Guettler et al., 2003; Karlsson et al., 2003; Lee et al., 2002; Loftus et al., 2005; Ospina-Giraldo et al., 2000). Collections from basidiomycete plant pathogens are restricted to the corn smut Ustilago maydis (Sacadura & Saville 2003; Nugent et al., 2004) and small sets from a number of rust fungi.

In rusts, despite detailed cytological research, knowledge about molecular mechanisms underlying infection and development within host tissue is only recently emerging (Hahn, 2000; Vogele, 2006). The major factor limiting molecular genetic analyses is the essentially obligate nature of fungal development and the lack of a stable DNA transformation system, although progress has been made with respect to the latter (Barja et al., 1998; Bhaiari and Staples, 1992; Li et al., 1993; Schillberg et al., 2000; Webb et al., 2005). An additional handicap is the size of the rust genomes; by flow cytometry, genome sizes of rust fungi were measured to range from 54 Mbp for P. coronata to 415 Mbp for Uromyces appendiculatus, while that of P. triticina was estimated at 124 Mbp (Eilam et al., 1994). The genome size of P. graminis f. sp. tritici was determined to be 67 Mbp by reassociation kinetics (Backlund and Szabo, 1993; Leonard and Szabo, 2005). Despite these hurdles, initial genomics work on several rusts has yielded interesting results over the last few years. The best-studied is the bean rust, U. fabae, from which haustorium-specific cDNAs have been described (Hahn and Mendgen, 1997; Jakupovic et al., 2006). The subsequent study of several newly discovered genes and their expression has yielded insight into the subtle biotrophic nature of the interaction with its host (Hahn et al., 1997; Link et al., 2005; Mendgen and Hahn, 2002; Struck et al., 1996, 1998; Vogele, 2006; Vogele et al., 2001, 2005; Wirsel et al., 2001, 2004). From the cereal pathogen wheat stem rust, P. graminis f. sp. tritici, 68 cDNA clones were generated from germinating urediniospores (Liu et al., 1993) and 203 unigenes were developed from infected wheat leaves using an SSH technique (Broeker et al., 2006). Recently, 488 unigenes were described for the soybean rust Phakopsora pachyrhizi (Posada-Buitrago & Frederick 2005).

To gain more insight into the molecular basis of P. triticina wheat biotrophism and to identify genes involved in pathogenicity and virulence, we have generated and analysed ESTs from 13 cDNA libraries constructed by using different techniques and representing four life-cycle stages including resting and germinated urediniospores and during appressorial and haustorial development; an existing library representing an incompatible interaction was mined for fungal sequences. Previous reports described small P. triticina EST collections from wheat-infected leaves generated through a cDNA-AFLP (Zhang et al., 2003) and an SSH technique (Thara et al., 2003) aimed at revealing in-planta expressed rust genes. A proteomics approach revealed up-regulated wheat and fungal proteins (Rampitsch et al., 2006). Genes expressed during the interaction with the host are generally considered to be more relevant to understanding the molecular mechanism of disease development. This presents a challenge to construct fungal-specific cDNA libraries from ‘interactive’ stages or to discriminate between cDNA sequences from pathogen and host. Haustorial structures can be isolated from heavily infected tissues (Catanzariti et al., 2006; Hahn and Mendgen, 1997) but the isolation procedure
risks generating major changes in expression profiles. Prior to the construction of cDNA libraries, the mRNA populations from infected plant tissues can be physically subtracted with unchallenged host RNA. Alternatively, sequenced cDNA clones can be distinguished on the basis of G+C content (Qutob et al., 2000), codon bias (Hsiang and Goodwin, 2003; Li et al., 2004; Maor et al., 2003), through identification by regular DNA blots (Hahn and Mendgen, 1997; Rauyaree et al., 2001; Thara et al., 2003; Zhang et al., 2003), using PCR (Sugui and Deising, 2002) and/or subtracted in silico with available host sequences (Kruger et al., 2002).

Assembly of a unigene set of putative fungal origin

Due to the manner in which our cDNA libraries were constructed, some of the clones contained chimeric inserts; that is, certain inserts were composed of cDNA fragments originating from different mRNA molecules ligated erroneously 'head-to-head' or 'tail-to-tail' around ApaI or NotI restriction sites used to generate the inserts. Potential chimeric inserts were identified, and subsequently separated in silico by finding sequences based on the presence of internal ApaI or NotI sites; non-chimeric sequences erroneously separated in this manner were identified by virtue of their joint appearance in the same contig (see below) coupled with similar BLAST homology results and were consequently reassembled (see Experimental procedures).

DNA sequences derived from resting urediniospores (stage Ptu) and urediniospores germinated in vitro (stage Ptg) were considered purely fungal; BLAST comparisons to the 'cereal database' did not indicate plant contaminants (other than the occasional leaf rust sequences that match both fungal and cereal/
plant sequences with similar E-values, discussed below). However, sequences that were derived from ‘interactive’ stages (appressorial and haustorial) needed to be distinguished from genes that were ‘most likely wheat’. To achieve this, we employed three methods. Two in silico analyses used similarity searches through BLAST and codon bias (see Experimental procedures). A third method employed comparative genome hybridizations (CGH) using a prototype cDNA microarray representing an initial 4484-unigene subset representing libraries PT001 to PT013 (G.B., unpublished data) and hybridizations to wheat and rust total genomic DNAs (data not shown). We did not want to discard sequences that were potentially fungal and therefore employed rather conservative criteria; unigenes that could not be assigned to these tests were considered potentially fungal. A subtraction method based solely on G+C content (e.g. Qutob et al., 2000) proved unsuitable because G+C content of 400 analysed wheat genes ranged broadly from 40 to 70 mol% with most around 54 mol%, basically encompassing that of the set of pure rust sequences which formed a much narrower Gaussian distribution with an average calculated content of 48.2 mol% (data not shown).

The cDNA library from an incompatible interaction between Thatcher Lr1 and race 1 is not expected to yield many fungal sequences as little fungal biomass will have accumulated in host tissue undergoing a hypersensitive response triggered within 24–48 hpi. Nevertheless, fungal ESTs from this stage will be very informative in shedding light on the pathogen response to an active HR resistance. The previously generated 7772 TaLr1 quality sequences (Table 1, B. Fofana et al., unpublished data) yielded a unigene set of 3769 sequences (2772 singlets and 997 contigs), which was mined for potential fungal sequences using conservative criteria (see Experimental procedures). A first set of similarity searches to collected Gramineae and fungal databases extracted 651 putative fungal unigenes from this stage, Pt1 (28 most likely fungal, 408 possibly fungal and 215 without any matches at E ≤ 10–5). Including these 651, we started analyses with a total set 25 558 quality sequences. A non-redundant set of 9760 unigenes resulted from the same clone were classified as singlets. As a result, the final unigene set consisted of 6616 singlets and 3144 contigs with an average length of 485 bp.

**Similarity to databases, annotation and classification of ESTs according to putative biological functions**

Various BLAST similarity searches were performed with the complete putative fungal data set of 9760 unigenes resulting in matches and annotations with various levels of confidence.

### Table 2

<table>
<thead>
<tr>
<th>Similarity*</th>
<th>e-value</th>
<th>NR</th>
<th>FUD</th>
<th>EO</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>10–20</td>
<td>594</td>
<td>407</td>
<td>1480</td>
</tr>
<tr>
<td>moderate</td>
<td>10–5–10–20</td>
<td>667</td>
<td>858</td>
<td>2271</td>
</tr>
<tr>
<td>low</td>
<td>10–3–10–5</td>
<td>309</td>
<td>1412</td>
<td>1978</td>
</tr>
<tr>
<td>total</td>
<td>1570 (16%)</td>
<td>2677 (27%)</td>
<td>5729 (59%)</td>
<td></td>
</tr>
<tr>
<td>no match</td>
<td>&gt; 10–3</td>
<td>8190 (84%)</td>
<td>7083 (73%)</td>
<td>4031 (41%)</td>
</tr>
</tbody>
</table>

* Amino acid sequence similarity based on the BLAST software to NR (GenBank non-redundant protein) or the TBLASTX software to FUD (custom collection of fungal DNA sequences; supplementary Table S7) or EO (est_others) databases.

**Table 3**

<table>
<thead>
<tr>
<th>Stages</th>
<th>Input ESTs</th>
<th>Unigenes ( %)*</th>
<th>Hits (%)‡</th>
<th>Annotated (%)§</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pt1</td>
<td>3917</td>
<td>920 (24)</td>
<td>372 (41)</td>
<td>214 (23)</td>
</tr>
<tr>
<td>Pt2</td>
<td>11 000</td>
<td>3711 (34)</td>
<td>1178 (32)</td>
<td>153 (4)</td>
</tr>
<tr>
<td>Pta</td>
<td>5988</td>
<td>2910 (49)</td>
<td>877 (30)</td>
<td>56 (2)**</td>
</tr>
<tr>
<td>Pth</td>
<td>4002</td>
<td>1631 (41)</td>
<td>985 (61)</td>
<td>428 (26)**</td>
</tr>
<tr>
<td>P1t</td>
<td>7727</td>
<td>651†</td>
<td>176 (27%)</td>
<td>127 (20)**</td>
</tr>
<tr>
<td>total</td>
<td>25 558</td>
<td>9760 (38)</td>
<td>3670 (38)</td>
<td>1000 (10)**</td>
</tr>
</tbody>
</table>

* Number of unique sequences (singlets plus contigs) representing one transcribed gene are given with percentage of total number of quality sequences for that stage in parentheses; CAP3 compilation within the respective stage-specific sets; total set not equal to numbers for individual stages combined due to overlap of extra sequences not assigned to stages.
‡ Potential fungal unigenes extracted from TaLr1 set (according to criteria outlined in Experimental procedures).
§ Number of unigenes finding a match at E ≤ 10–5 in all available databases (excluding human and mouse ESTs) are given with percentage of unigenes in parentheses.
† Excluding data for ‘est_others’ database to correct for self-identification of previously deposited TaLr1 ESTs.
** Number of unigenes annotated with putative gene function at E ≤ 10–5 are given with percentage of unigenes in parentheses.
*** Includes unigenes most likely wheat (based in addition on annotation results) in stage Pta: 3; Pth: 211; P1t: 79; for total set: 296 (supplementary Table S2).

Table 2 shows the result of an initial analysis to assess the overall level of similarity in three major databases. It became immediately clear that the majority of sequences did not produce significant matches at E ≤ 10–5. To address the question of whether levels of similarity were associated with specific life-cycle stages, we pooled the quality sequences in each stage and compiled unigene sets within each of them (Table 3). Using this approach, roughly one-third of the unigenes matched other sequences at E ≤ 10–5; the stage Pth produced significantly more matches.
(61%) but it is likely that our conservative criteria left too many host sequences in this set and because databases contain many more cereal (plant) than fungal sequences, this would yield a biased set. A low annotation rate for the unigenes from stages Ptg and Pta suggested the presence of many novel or highly divergent genes.

Using the BLASTn algorithm, the total unigene set revealed a small set of ten members representing rRNA ‘contaminants’. A further 990 unigenes having similarity matches with $E \leq 10^{-5}$ according to translated BLAST analyses were organized into 19 main functional categories (based on the Munich Information Centre for Protein Sequences’ FunCat system; http://mips.gsf.de/projects/funcat; Mewes et al., 1999), assigning each unigene to a single, most-appropriate category; class 99, ‘unclassified proteins’, was omitted. Supplementary Table S1 presents the annotation of 694 unigenes (7%) most likely to be of fungal origin. Overall, the ESTs represented a broad range of genes encoding proteins involved in general metabolism, energy, protein synthesis and cell fate. In particular, genes involved in cell rescue, defence, and virulence are of interest because these are thought to play a role in pathogenicity and pathogen defence against its host. Several gene sequences from leaf rust-infected wheat plants (Thara et al., 2003) deposited in the public databases almost perfectly matched our *P. triticina* ESTs; some had close homologues in flax rust, *Melampsora lini* (Catanzariti et al., 2006), soybean rust, *Phakopsora pachyrhizi* (Posada-Buitrago & Frederick 2005), or bean rust, *Uromyces fabae* (Jakupovic et al., 2006).

Most of the translated *P. triticina* ESTs yielding significant hits (at $E \leq 10^{-20}$) showed similarity to fungal protein sequences, particularly to those from the related plant pathogenic fungus *U. maydis* and other basidiomycetes such as the *Cryptococcus* species and the higher mushrooms, *P. chrysosporium* and *C. cinereus*. Interestingly, 42 sequences matched ESTs from *U. maydis* and i: 574.

Annotation results for the remaining 296 unigenes suggested that they could be derived from host tissues. Based on all available criteria, they were deemed most likely host sequences and are given in supplementary Table S2. Some entries yielded annotations from fungal species, but in these cases a sequence from wheat or a related cereal with unknown function provided a better match. The advantage of constructing and analysing cDNA libraries from ‘interactive tissues’ in both compatible and incompatible interactions is the discovery of host genes potentially involved in accommodating or resisting fungal development. Changes in the transcription levels of such genes upon host infection would have to be verified, which has been done by comparing a compatible and an incompatible interaction using microarrays and quantitative RT-PCR (S.C. et al., unpublished data). Wheat genes potentially involved in disease progression as judged from their elevated occurrence in the Pth stage include a carboxic anhydrase (Contig5720), a glycine decarboxylase (Contig6034), a furmarylacetocetate hydratase (Contig5861), a ubiquitin-conjugating enzyme E2 (Contig5874), an aquaporin-like protein (Contig6585), a metacaspase (Contig6141), a metallothionein (Contig6001) and an α-tubulin protein (Contig5647). Of interest is a group of resistance gene-like proteins (supplementary Table S2, category 32) that might be induced in an incompatible interaction, as well as a heat shock protein (Hsp70, TaLr.Contig419). In general, we identified many wheat genes involved in energy housekeeping, photosynthesis, protein synthesis (ribosomal proteins) and protein degradation (ubiquitination), some of which have been shown to be differentially expressed upon host infection and subsequent defence initiation (B. Fofana et al., unpublished data).

### Abundance of ESTs and comparison between stages

The large set of unigenes allowed a search for sequences common to several stages and to identify gene sets that are possibly unique to certain stages. Figure 1 shows a Venn diagram summarizing this initial analysis. The highest number of common transcripts (160) was found between resting and germinated urediniospores, suggesting transcriptome overlap between these stages. Overall, however, little overlap was found between the various stages, suggesting the induction of rather different gene
complements in these differentiated fungal structures. To assess which of the annotated genes were found in these various unique and overlapping stages, data from supplementary Table S1 were reanalysed to obtain supplementary Table S3. Among observed trends were a possibly increased number of genes involved in energy conversion, including six weak hits to fructose-bisphosphate aldolase, in germinated urediniospores. Also, in the haustorial stage, a large number of ribosomal protein genes was revealed, suggesting very active protein synthesis and metabolism, similar to what has been reported in earlier studies in the bean and wheat rust pathosystems (Broeker et al., 2006; Jakupovic et al., 2006; Thara et al., 2003; Zhang et al., 2003). Several transcripts for proteins involved in the biosynthesis of vitamins and some other compounds were only discovered in the haustorial stage, some of which have been identified as in planta-induced (PIG) genes in haustorial cDNA libraries from the bean rust system (Voegele, 2006) and the leaf rust-wheat system (TR genes, Thara et al., 2003; see Discussion). Very few fungal unigenes with significant hits specific to the Pti stage were recovered; of interest might be a P450 monooxygenase (TaLr.1170H09.F.pti, TaLr.1170H09.R.pti) possibly involved in detoxification of defence compounds, and a U. maydis homologue (TaLr.1174C01.R.pti), similar to Sec24C, which codes for a component in a coat protein complex responsible for COPII vesicle budding from the endoplasmic reticulum (Miller et al., 2003). Sec24C was also identified with lower similarity, however, through Contig6655, which represented other stages as well and expression of which was analysed by qPCR (Fig. 2).

Within the collection of 9760 non-redundant unigenes, the overall level of sequence abundance could be assessed based on the numbers of ESTs included in the 3144 contigs. Supplementary Table S4 shows a list of the 35 most frequently occurring ESTs; many of these contigs did not yield annotations with high confidence levels nor had hits in the databases we used, indicating novel or highly diverged genes. Most ESTs in this project were produced by PCR-based approaches that were designed to ensure linear amplification of all cDNA molecules. However, there is a possibility that the abundance of ESTs may not necessarily represent the real transcript level. Because libraries PT001 to PT004 were produced from subtracted and normalized cDNAs and were therefore expected not to represent true mRNA levels, they were excluded from the stage-specific abundance analysis. Therefore, numbers for the respective stages were generated out of total pools of 1836 ESTs for Ptu, 8093 for Ptg, 5259 for Pta, 2160 for Pth and 651 for Pti. The data nevertheless allowed a preliminary look at possible gene expression levels based on transcript abundance in certain stage-specific libraries (supplementary Tables S1–3: column ‘contig_info’). No statistical analysis was performed (such as described in Soanes and Talbot, 2005; Stekel et al., 2000) but we assessed the correlation of the
Served families of SKP1 genes essential for many cell functions, presumably as a result of their function during protein Inf8p from the same organism, seemed highly abundant whereas Contig2736, possibly coding for differentiation-related gene, probably in preparation for appressorium development. The ratio, however, indicated that the normalization procedure contained no information yet an elevated level of mRNA was measured in this stage by qPCR correlated with its occurrence.

In Contig6655, EST abundance did not seem to be correlated with the low transcript level in germinated urediniospores as measured by qPCR. Similarly, in Contig6890, no large transcript accumulation was measured by qPCR 6 days after infection. It is possible that the latter gene represents a wheat homologue of a very conserved gene given that no mixed-source sequences were apparent in the contig pile-up; however, the gene-specific primers did not produce a PCR product from wheat total genomic DNA. For both anomalies, no easy explanation can be offered but these results indicate that the presented ‘contig_info strings’ are mere ‘indications of transcription rates’. In another example, Contig2511 (Fig. 2) consisted of four ESTs sequenced from library PT001, which was constructed from resting urediniospores using a normalizing procedure. Consequently, the ‘contig_info string’ contained no information yet an elevated level of mRNA was measured in this stage by qPCR correlated with its occurrence (four ESTs out of 1166 quality reads revealed in library PT001). The ratio, however, indicated that the normalization procedure might not have worked as anticipated for this library.

From the annotated contigs (supplementary Table S1) that had at least five EST members between two compared stages, we attempted to glean tentative differential gene expression values by comparing the respective EST abundance in normalized sets (supplementary Table S5). A few warrant special mention, such as a ‘negative regulator of sulphur metabolism’ sconCp (in particular Contig5823, among several contigs), most similar to a Botryotinia fuckeliana homologue. Many transcripts for this gene were found in resting urediniospores and upon germination but seemed significantly more abundant in planta. SconCp from A. nidulans was recently described to belong to the highly conserved family of SKP1 genes essential for many cell functions, including cell cycle regulation (Piotrowska et al., 2000). Two genes coding for a tropomyosin, involved in lateral binding and stabilizing actin cables (Contig109, Contig6817), appeared to be induced during spore germination, as was a gene similar to HIS4 (Contig6884) involved in histidine biosynthesis. A transcript (Contig5386) with weak similarity to infection structure protein Inf24p from U. appendiculatus seemed induced upon spore germination, probably in preparation for appressorium development whereas Contig2736, possibly coding for differentiation-related protein Inf8p from the same organism, seemed highly abundant in urediniospores, presumably as a result of its function during sporulation. A similar function in leaf rust urediniospore genesis and/or during its subsequent germination could also explain the relative abundance of a homologue to conidiation-specific protein, con6p, from N. crassa (Contig6730). We encountered a number of ESTs with similarity to heat-shock protein family members (HSP40, 70, 78, 80 and 90) and their co-chaperones. These are important factors involved in alleviating stress on proteins (folding, protection, assisting mutated forms in attaining their function; Heitman, 2005). They were identified in several stages, sometimes in large numbers such as HSP90 (Contig2530), a homologue identified in a haustorium-specific cDNA library from the bean rust U. viciae-fabae. This latter HSP seemed similarly abundant in leaf rust haustoria and in germinating spores. This could indicate that the organism experiences stress during various developmental stages or suggests a common accessory, protein-stabilization role for this protein.

Some genes may have specific roles in certain stages. For example, an endopolygalacturonase possibly encoded by Contig731 and expressed in the appressorium stage is known to aid plant infections in other pathosystems by degrading host cell walls (Deising et al., 1995). A phenylalanine ammonia lyase (Contig5951) seems significantly expressed in haustoria (supplementary Table S5) and, although known to be induced in plants upon pathogen attack, its function in fungi is not known but could indicate increased amino acid metabolism or a need for phenylpropanoid compounds. Very high mRNA abundance also specific to the haustorial stage was found for a metallothionein-like gene known to be expressed in organisms under oxidative and metal stress (Contig2474_part1); the same annotation was attributed to Contig23 and Contig6120, which were encountered only once in this stage. Similar results were reported for in planta stages for several other fungi (Broeker et al., 2006; Ebbole et al., 2004; Jakupovic et al., 2006; Kim et al., 2001; Sexton et al., 2006). However, metallothionins were also found in the resting urediniospore stage (PT00111b.F07.77.ptu.chim1, homologue to PIG11, and PT0281.A23.C21.ptm.chim2) and in resting and germinated urediniospores as well as haustorial stages (Contig2745, homologue to PIG13).

**EST variants and marker design**

The comparison of constituent sequences in the contigs revealed variations among the members in many contigs. Identical changes seen in at least two independent clones in a contig, to exclude possible errors during cDNA library construction or sequencing, were analysed further. Figure 3 illustrates examples of contigs in which we observed 3'-end (Fig. 3A) and internal mRNA variation (Fig. 3B). Whether such variants originate from different members of gene families, illustrate alternative mRNA processing or represent polymorphic alleles coming from the two different parental nuclei must await analysis of the corresponding genomic sequences (Wahle and Ruegsegger, 1999; Zhang and Min, 2005). The possibility of differential splicing, in particular in
different developmental or infection stages, would be of great interest. Many differences were found to reside in single nucleotide stretches (SNPs) or in di-, tri-, or more nucleotide repeats (SSRs or microsatellites; see Fig. 3C for an example) and are currently being investigated for use as markers in our mapping project (Wang et al., 2006).

DISCUSSION

We report here the generation of a large P. triticina EST database covering several life-cycle stages during the infection on wheat with the primary goal being that of gene discovery. The P. triticina EST analysis has identified many genes expressed during urediniospore germination, infection structure formation such as appressorial and haustorial development, and in the early interaction with an incompatible wheat host. The number of identified unigenes is most likely an overestimate of the number of expressed genes discovered as independent unigenes potentially could represent 5′- and 3′-parts of the same transcribed gene; in some cases, similarity searches corroborated this.

We attempted to derive a unigene set restricted to fungal sequences, but it is likely that a few wheat contaminants are still present due to relatively conservative contaminant filtration criteria. For example, using BLAST, sequences lacking similarity to existing databases were considered to be of fungal origin but could potentially be derived from wheat genes, although given the sheer magnitude of cereal sequences in public databases, this is rather unlikely. It is interesting that many sequences show BLAST alignments with similar E-values to proteins in both plants and fungi. For example, TaLr1022F09R matches a sequence from Triticum aestivum in ‘nr’ with an E-value of 8e–83, suggesting that this EST is a contaminant, while a match in our fungal genomic database is to C. albicans at 1e–58 and in the fungal EST database to M. grisea at 9e–57. It is tempting to speculate that during coevolution of this pathosystem horizontal gene transfer has occurred from host to fungus. However, comparing sequences across several complex databases and extracting biological relevance based on E-values is neither trivial nor automatic. We must use great caution when trying to distinguish fungal from wheat host sequences in interactive libraries using in silico approaches. The comparative genome hybridizations using total, labelled gDNA from both P. triticina and wheat on our prototype cDNA microarray worked rather well but only sampled 4484 initial putative fungal transcripts. Calculations using codon-usage tables generated from sets of well-annotated rust ESTs and published wheat genes (essentially as described in Li et al., 2004) were useful, although a large number of unigenes remained unassigned using this method. Bioinformatic methods and training or reference sets (Hsiang and Goodwin, 2003; Maor et al., 2003; Thomas et al., 2001) will need to be optimized to discriminate between pathogen and host sequences from interactive libraries. Until complete genome sequences of both host and pathogen are available, only indirect methods can be used to distinguish sequences from mixed sources. However, to evaluate the source of the ESTs, we included results of BLASTn comparisons against genomic sequences of the stem rust fungus, P. graminis (see Experimental procedures).

The number of unigenes matching other sequences across various databases at $E \leq 10^{-5}$ was in the 40% range, which, although low, was not very different from what has been reported for other fungal EST projects. Nevertheless, it is obvious that there is a paucity of fungal (especially basidiomycete) sequences in public databases. More surprising was the low number of unigenes (10%) in our total set that matched proteins with known function, compared with 62% for U. maydis (Sacadura & Saville, 2003) and 40% for the bean rust, U. fabae (Jakupovic

**Fig. 3** Examples of observed sequence variation among members of contigs. A, 3′-end variation; B, mRNA-internal variation; C, single nucleotide (SNP) and tri-nucleotide repeat (SSR) polymorphisms.
et al., 2006). However, when considering the different stages we tested, the level of functional annotation varied more widely with very low numbers for the germinated urediniospore and appressorium stage libraries (Table 3). These highly differentiated and specialized stages could express very specific gene complements but since the number of overall database hits found for these stages is comparable with the other stages, it is more likely that they just express many genes of yet unknown function. It is also possible that our set contains many partial sequences that do not produce significant hits using the BLAST software.

It was revealing to see that various life-cycle stages seemed to express rather different gene complements (Fig. 1). Because our approach has probably resulted in the identification of the most prevalent mRNAs, expanding our non-comprehensive unigene set would undoubtedly reveal rarer transcripts and probably increase the occurrence of ESTs common to the various stages. Our results nevertheless suggest the occurrence of major transcriptome shifts during the profound developmental programmes the fungus undergoes during infection. Also of note was the high number of transcribed sequences obtained from resting urediniospores. Apparently, a large pre-existing pool of mRNA is available probably for essential functions during the early stages of germination. This was suggested for conidial germination in A. nidulans (Osherov & May, 2001). It is also possible that these mRNAs reflect (additionally) the transcriptome present at spore maturation and dessication, therefore revealing proteins involved in these processes.

Ten unigenes matched ‘glucose repressible protein GRG1’ of which eight produced similar E-values (Contig147, Contig5538, Contig2506, Contig2539, Contig2764, Contig2793, Contig2664 and Contig5533). A more precise nucleotide comparison between these unigenes confirmed that their consensus sequences were related. Their divergence, however, caused the CAP3 program under the parameter settings used to assign them as originating from separate transcripts (data not shown). This is probably an example of a gene family whose members are possibly differentially transcribed between stages; for example, Contig2506 had 16 ESTs from stage Ptu, six from Ptg and three from Pth, which, given the total EST pool sizes in the respective stages (see legend to supplementary Table S1), seems significant. In contrast, Contig5533 harboured a single EST in stage Ptu, three in stage Ptg and two in stage Pth.

We identified two ESTs with similarity to the gEgh16 and gEgh16/MAS1 proteins from B. graminis (Contig5593 and Contig5424, respectively). In this fungus, both genes seem to constitute a large family potentially involved in plant interaction (Grell et al., 2003). Similar to the situation in B. graminis, these two transcripts seemed more abundant in P. triticina urediniospores and seemed induced even further upon germination; this is corroborated by the finding of homologues in the soybean rust, P. pachyrhizi (Posada-Buitrago & Frederick, 2005).

In the interactive libraries we found several unigenes with similarity to sequences identified as in planta-induced genes (PIGs), in particular PIG1p (thiamine biosynthesis protein THI1p), PIG4p (thiazole biosynthetic enzyme), PIG16p (cytochrome P-450 monooxygenase), PIG17p and PIG6p in U. fabae (Hahn and Mendgen, 1997; Jakupovic et al., 2006), and TR4 (thiamine biosynthesis protein, homologue to THI1p), TR29 (eF1α), TR33 (microsomal signal peptidase) and TR40 (unknown) in P. triticina (Thara et al., 2003). However, in our set, several homologues were clearly expressed in other stages as well, such as PIG11p and PIG13p (metallothioneins), PIG28p = TR24 (cyclophilin), TR21 (thiazole biosynthetic enzyme), TR14 (histone H4), TR27 (Gbeta-like protein), TR32 (mRNA polyadenylate-binding protein), TR34 (unknown), TR42 (ADP/ATP translocase-antipporter), TR51 (histone H2A) and TR56 (HSP70). These might be transcribed at a higher level in planta but must have roles during other developmental stages as well; for some of the homologues this has been corroborated in the bean rust (Jakupovic et al., 2006) and stem rust systems (Broeker et al., 2006). In this context, examination of the proteome of wheat leaves inoculated 9 days prior with P. triticina, identified 22 (possibly 25) consistently up-regulated fungal proteins out of 32 analysed. Six of the 22 were directly identified using the P. triticina EST database described in this study, one of which was similar to PIG4p from U. fabae (Rampitsch et al., 2006). Several of the bean rust PIG genes are thought to contribute to the functionality of the haustoria and could therefore impinge on virulence, but no direct evidence for this is available (Voegele, 2006).

Potential virulence and pathogenicity genes

Of great interest are potential virulence and pathogenicity genes, particularly those coding for products essential for early infection processes. Components of signal transduction cascades impinging on virulence have been the focus of many studies because they represent possible targets for fungicides. These include components of rather conserved pathways such as cAMP-dependent protein kinases and MAP kinases. We have identified homologues of several such components and a P. triticina MAPK1 (Contig2832) has been studied in detail (Hu et al., 2007). Another pathway features calcineurin, an essential, serine/threonine-specific protein phosphatase that is controlled by calcium and calmodulin, acts through cyclophilin, and plays a role in morphogenesis and virulence of most fungi, including C. albicans, C. neoformans, M. grisea and B. cinerea (Viala et al., 2003). Many Ca2+-signalling proteins have been described in a comparative genomic analysis of N. crassa, M. grisea and S. cerevisiae (Zelter et al., 2004). Some homologues involved in calcium binding and calmodulin-cyclophilin signalling were identified in P. triticina including a putative calcium transporting ATPase/translocase related to VCX1.
which acts in the degradation of arginine, citrulline and ornithine, and ornithine aminotransferase CAR2 (PT0065c.G08.BR.ptg), catalyses the fourth step in ornithine and arginine biosynthesis, polyamine metabolism, ornithine decarboxylase Odc (Contig2661), five ESTs representing a glyoxysomal malate synthase (Contig6672) for spore germination and pathogenicity, possibly through peroxisome oxidation, probably involving carnitine-

Indeed, ultrastructural observations suggested that large lipid bodies in the cytoplasm of several spore types seemed to degrade during germination (Mendgen, 1984). Recent reports indicate the need for rapid increase of the acyl-CoA pool through lipolysis, i.e. peroxisomal fatty acid β-oxidation, probably involving carnitine-dependent metabolic activities, to build up the turgor necessary in the appressorium for penetration (Bhambra et al., 2006; Ramos-Pamplona and Naqvi, 2006). Four unigenes identified as having acyl-coenzyme A dehydrogenase activities were generated from resting and germinating spore stages (Contig2609, Contig2901, Contig2773 and Contig5548). In addition, a homologue of mitochondrial carnitine/acylcarnitine translocase (Contig6743) seemed up-regulated upon urediniospore germination (supplementary Table S5). Lipid metabolism via the glyoxylate cycle requires malate synthase and it was recently shown that this enzyme was important for infection and vital for pathogenicity of Stagonospora nodorum (Solomon et al., 2004). We sequenced five ESTs representing a glyoxysomal malate synthase (Contig6672) from the resting urediniospore stage indicating that this spore type is ‘primed’ for this metabolic activity. Moreover, Contig2729, sequenced at least three times from the same spore stage, codes for an isocitrate lyase which also has been shown to be important for spore germination and pathogenicity, possibly through peroxisomal metabolic pathways, in the blackleg fungus Leptosphaeria maculans (Idnurm & Howlett, 2002) and the cucumber anthracnose fungus, Colletotrichum lagenarium (Asakura et al., 2006); an EST having low similarity was also found in the incompatible stage (TaLr.1170H07.F.ptt).

Interestingly, we identified a sequence with similarity to a member of a group of recently described ‘haustorially expressed secreted proteins’ (HESPs) in the flax rust, M. lini (Catanzariti et al., 2006). Several of these ‘HESPs’, identified in a bioinformatics approach, are proven elicitors of the hypersensitive response in flax and co-segregate with known avirulence genes. Direct sequence comparison revealed that P. tritici Contig6739 is homologous to M. lini/HESP379 (and U. viciae-fabae haustorium-specific cDNA clone Uf074; data not shown). It will be interesting to see whether this homologue elicits the HR in wheat and/or co-segregates with one of the many known avirulence genes in our populations.

Fungal components neutralizing host defence responses are virulence factors. Such host responses frequently involve the production of reactive oxygen species (ROS). The metabolism/de toxification of ROS is therefore essential for the establishment of the pathogen. We found a (manganese) superoxide dismutase precursor (Contig6132) in the haustorial stage. SOD was identified as a differentiation protein expressed in U. appendiculatus gernmings during early appressorium development (Lamboy et al., 1995). Hydrogen peroxide as a byproduct of SOD must be detoxified by catalase, transcripts for which we discovered ESTs in the appressorial (Contig266), haustorial (Contig6007) and also germinated spore stages (PT0065d.F05.BR.ptg). Other antioxidants such as thioredoxin and glutathione can also enhance protection and detoxify xenobiotics, apart from fulfilling many other functions (Pocsi et al., 2004). We found several ESTs coding for activities generating this potential, such as thioredoxin and glutathione reductase/lyase/hydrolase (glutaredoxin) and glutathione-S-transferase (Contig6033, PT00110c.D12.B7.ptu.chim1, Contig2851, PT0065b.C01.BR.ptg, PT0065b.F03.BR.ptg, PT0191.O19.CPTR.ptm.chim2, PT0055a.F04.BR.ptu, Contig5460), but of these, Contig6033 was the only one uniquely found in the haustorial stage. Contig1886 (and PT0141.K04.C21.ptm) bore similarity to a β-glucosidase precursor from U. viciae-fabae, BGL1p, identified in a haustorial-specific library but later shown to be expressed in all early infection stages; it is involved in cellulose/cellobiose utilization but conceivably in defence because of its similarity to avenacinate with a potential phytoalexin-detoxifying activity (Haerter and Voegele, 2004). Our unigene was discovered in germinated urediniospores and appressoria, which would not preclude a function in defence. Some fungal factors might be involved in evasion of the host response; U. appendiculatus changes the surface-exposed chitin moieties of its germ tubes and appressoria to chitosan using highly induced chitin deacetylases when penetrating host cells to prevent triggering immune responses (Deising...
and Siegrist, 1995; Mendgen et al., 1985). In *P. triticina* we identified several chitin synthases and a homologue of CD2A, a sporulation-specific chitin deacetylase and precursor (Contig6649, Contig2903) in germinated spores.

A rough estimate puts the *P. triticina* genome coverage achieved through this project at over 4 Mbp (9760 unigenes with an average size of 485.3 bp), revealing an average G+C content of 48.2 mol%. The generated sequences could therefore represent 3–4% of the *P. triticina* genome estimated to contain 100–120 Mbp. Having generated over 9000 unigenes, we are uncertain what percentage coverage of the leaf rust transcriptome these ESTs might represent. Based on the analysis of complete fungal genome sequences, *U. maydis* has an estimated 6902 genes (Kamper et al., 2006) whereas another basidiomycete, *P. chrysosporium*, contains a calculated 10,048 (Wymelenberg et al., 2006). Among ascomycetes, *N. crassa* was recently estimated to possess 10,082 genes (Borkovich et al., 2004) and the related plant pathogen *M. grisea*, 11,109 (Dean et al., 2005). It is conceivable that macrocyclic rusts possess more genes than the aforementioned fungi because of their extraordinarily complex life cycles including a sexual stage on an alternate host.

To increase functionality of our EST database, we have made use of ESTIMA, a general EST management system (Kumar et al., 2004), which manages gene ontology (GO; Camon et al., 2004) information. Currently, we are in the process of extending the database by generating ESTs from a haustoria-specific stage (G.B., J. Fellers), a teliospore stage isolated from senescent wheat, and two spore stages (pynicio- and aeciospores) from the alternate host (G.B., J. Xu, Y. Ankster). *P. triticina* sequences from these collaborative projects, as well as those acquired from public resources, will be merged into a comprehensive collection for analyses and the annotation of the related stem rust (*P. graminis* f. sp. *tritici*) genome. The data presented here should contribute significantly to the generation of general genomic resources for the cereal rusts. Careful analyses of these resources should allow in time for the identification of genes essential for the disease process. Such fungal genes might prove to be their ‘Achilles heel(s)’, the targeting of which will help protect our cereal crops.

**EXPERIMENTAL PROCEDURES**

**Plant materials and rust strains**

Wheat cultivar Thatcher, which is susceptible to *P. triticina* race 1 (designation BB8; Long and Kolmer, 1989) carrying avirulence gene *Avr1*, and near-isogenic line Thatcher *Lr1*, which is resistant to this race, were used in this study. Seedlings were grown in 10-cm pots maintained at 18 °C (night) and 20 °C (day) in a leaf rust-free greenhouse with a photoperiod of 12 h. Inoculation was done on 10-day-old seedlings as described (Hu and Rijkenberg, 1998a). Materials used to prepare cDNA libraries included: (1) prefix *Ptu*—resting urediniospores; (2) *Ptg*—urediniospores germinating in vitro after dusting them on the surface of sterile double-distilled water in Petri dishes (for 6 h or pooled with those after 12 h)—for one library ‘wheat extract’ made from sterilized, surface-sterilized 2-day-old wheat germinals, ground-up in water, was added; (3) *Pta*—germinating urediniospores on compatible Thatcher leaf surfaces, at the stage of appressorium formation, before penetration, at 5 hpi; (4) *Pth*—haustorium developmental stage before sporulation, at 6 days post inoculation (dpi); and (5) *Pti*—incompatible interaction between Thatcher *Lr1* and race 1 at 24 hpi; see Table 1.

**Preparation of mRNA and cDNA library construction**

Freshly harvested resting urediniospores, germinating urediniospores or infected plant tissues were ground to a fine powder in liquid nitrogen using a mortar and pestle. Appressoria at 5 hpi were embedded by spraying with a solution of 20% cellulose acetate in acitone, which solidified within minutes. This film was quickly peeled from the epidermis and frozen in liquid nitrogen. Total RNA was extracted in a solution of phenol-guanidine isothiocyanate (‘TRI Reagent’, BIO/CAN Scientific, Mississauga, ON, Canada) according to instructions provided by the manufacturer. RNA was precipitated overnight at 4 °C with 1/8 vol 10 m lithium chloride to remove polysaccharides and other contaminating substances. PolyA RNA was purified using biotinylated oligo-dT/streptavidin-coated magnetic beads (‘PolyA-Tract’; Promega, Madison, WI). To allow for directionally cloning and 5′- or 3′-end sequencing of the cDNA products, we designed primers incorporating rare-cutting restriction enzyme sites. *NotI* is embedded at the 3′-end and *ApaI* at the 5′-end of cDNAs. First-strand cDNA synthesis was performed as described (Ausubel et al., 2003) using SuperScript II RT (INvirogen Canada Inc., Burlington, ON, Canada) and a poly T primer, dTNotI: 5′-GCGCCGCGGCCGCGC-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT (restriction enzyme site underlined). Single strand cDNA-3′-ends were extended by poly(dG)-tailing. Samples in 25 µL were incubated at 94 °C for 3 min, cooled on ice after which 5 µL 5×TdT buffer, 1 µL 40 mM dGTP, and 1 µL (15 units) terminal deoxynucleotidyl transferase (Invirogen) was added. After incubating at 37 °C for 1 h, the sample was heat-inactivated at 65 °C for 10 min.

Due to the biotrophic nature of the leaf rust fungus, only limited amounts of mRNA could be obtained, except from infected leaf tissues. We therefore employed a PCR-based protocol for cDNA library construction from limited starting material (Froussard, 1992; Revel et al., 1995). This included the use of a combination of ‘full-length’ *dNotI* (above) and *dCaph* (5′-GGCCCCGGGCCCCCTCCCCCCCCCCCCCCC) (primers at 0.2 µM each, and ‘truncated’ (tr) primers, trNotI (5′-GGGCCTGGCCGCGCTT) and trApaI (5′-GGCCCCGGGCCCCCTCC) at 1.8 µM each. To reduce the polyA/T tract to limit subsequent non-specific hybridization during
subtractions and slippage during sequencing (Wang et al., 2000), we modified this protocol to replace the dTNotI primer with a set of five ‘anchored’ T primers (equal amounts of oligo-dT primers: 5′-GGCCGCGGCCGC-TTTTTTTTTTTA; 5′-GGCCGCGGCCGGCCG-TTTTTTTTTTTTCA; 5′-GGCCGCGGCCCGCTTTTTTTTCC; 5′-GGCCGCGGCCCGCGCTTTTTTTTG; and 5′-GGCCGCGGCCCGCGC-GTTTTTTTTTTTCC) at 0.2 µM total. PCR reactions were performed directly on first strand cDNAs in 50 µL containing 60 mm Tris-HCl, pH 8.5, 17 mm ammonium sulphate, 6.5 mm MgCl2, 6.7 µM EDTA, 0.07% β-mercaptoethanol, 0.017% BSA, dNTPs at 0.5 mM, and 5 units Taq polymerase (Invitrogen), or ‘Herculase’ polymerase (Stratagene), or a mixture of 1 : 10 Pfu (Stratagene)/Taq polymerase (Invitrogen). The PCR protocol was designed to ensure linear amplification representative of the transcription profile (Revel et al., 1995) and consisted of a first round: preheating at 95 °C for 10 min followed by 15 cycles of 2 min at 95 °C, 50 s at 68 °C and 6 min at 72 °C. For a second round the PCR mix was divided over four new tubes, 12.5 µL in each, and 37.5 µL of the same, fresh 1× PCR mix including all primers was added. The PCR was continued for an additional 15 cycles with a final 12 min-elongation step at 72 °C.

Full-length cDNA production
We modified an existing protocol (Maruyama and Sugano, 1994) based on the protocol from the GeneRacer Kit (Invitrogen): Calf Intestinal Alkaline Phosphatase (Invitrogen) was used to dephosphorylate all molecules except mRNA molecules protected by an intact 5′-m7Gppp-cap structure. Tobacco Acid Pyrophosphatase (Epiconcept Technologies, Madison, WI) was used for subsequent de-capping of remaining mRNAs, leaving a 5′-phosphate. To these 5′-ends, an ApaI-restriction site-containing RNA-oligonucleotide, 5′-GGGGGCGGCCCCUC (equivalent to the trApaI DNA primer), was ligated using T4 RNA ligase (Epiconcept). Subsequent first strand cDNA synthesis and PCR were as described above.

Normalization
For normalization, 10/11 of a particular cDNA pool was randomly labelled with biotin according to the manufacturer’s protocol (Biotin-Chem-link, cat#1-812-149, Boehringer Mannheim/Roche Diagnostics, Laval, Quebec, Canada). This fraction was mixed with the unlabelled 1/11, heated to 94 °C and hybridized for 24–48 h at 68 °C. Hybrids were removed by binding to streptavidin-coated magnetic beads (Promega) in three repeated steps. The remainder of the cDNA molecules were re-amplified for 15 cycles as above.

Subtraction
Plant host cDNAs from ‘interactive libraries’ (that is, fungus-infected wheat leaves) were subtracted by mixing in 10× excess (by weight), biotinylated cDNA molecules made from mRNA from uninfected wheat leaves at the same developmental age. This mixture was heated to 94 °C and hybridized for 48 h at 68 °C. Hybrids were removed by binding to streptavidin-coated magnetic beads (Promega) in four repeated steps. The remainder of the cDNA molecules were re-amplified for 15 cycles as above.

Standard protocol
This comprised the Stratagene protocol with modifications. After the first strand synthesis using a polyT primer with an embedded Xhol restriction site, cDNAs were subjected to the second strand synthesis and addition of an EcoRI adapter. Ligation products were digested with Xhol and EcoRI and directionally cloned into the corresponding sites of pBlueScriptIIKS (Stratagene, La Jolla, CA).

PCR products were purified over columns (QIAquick PCR Purification Kit, Qiagen, Mississauga, ON, Canada) and digested with ApaI and NotI. They were then subjected to two rounds of size selection over 1% agarose gels in 0.5x TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8.0) with intermittent direct isolation from agarose slices using a QIAquick Gel Extraction Kit (Qiagen). Fractions containing fragments of 0.4–1.0 kb, 1.0–2.0 kb, and > 2.0 kb were cloned separately and directionally in dephosphorylated pBlueScriptIIKS digested with NotI and ApaI. Ligation products were introduced into E. coli DH5α or DH10B/ir via electroporation. TaLr1 clones were generated using mRNA obtained from 14-day-old leaf tissue from the near-isogenic line Thatcher Lr1, 24 hpi with P. triticina, race 1. cDNA was prepared according to the standard Stratagene protocol using EcoRI and XhoI adapters and was cloned directionally into vector Lambda ZapII in E. coli strain XLOLR followed by mass-excision of the clones, now in plasmid vector pBK-CMV.

Quantitative RT-PCR
Two micrograms total RNA was treated with 2 units DNasel (Amplification Grade, Invitrogen), reverse-transcribed with 200 units SuperscriptIII (Invitrogen), and remaining RNA digested with RNaseH. The first-strand cDNA samples were diluted to 12.5 ng/mL (based on input RNA). Quantitative RT-PCR was performed on an Mx3000P qPCR System (Stratagene) with primer pairs as indicated in supplementary Table S6; the length and uniqueness of the amplicons was verified by agarose gel electrophoresis and melting curve comparison within the analysis software. Transcript levels were derived from the accumulation of SYBR green fluorescence using the following components: 20 µL volume containing 1× AmpliTaq Gold II buffer plus 0.5 unit AmpliTaq Gold (PE BioSystens), 2.5 mM MgCl2, 0.2 mM each dNTPs, 150 nmol each primer, 7.5% glycerol, 3% DMSO, 1 : 40 000 dilution of SYBR green and 1 : 30 000 dilution of ROX as reference dye (both from Stratagene), and 25 ng cDNA. PCR conditions were: 10 min at 95 °C, 40 cycles of 30 s at 95 °C, 30 s at 58 °C, 1 min at 72 °C.
Plasmids were isolated using a QIAprep Spin Miniprep Kit according to the manufacturer’s protocol (Qiagen) and at the CRC using the Eppendorf Direct Bind Kit. At the GSC, clones in E. coli DH10B/r were picked on a Genetix Q-Pix into 384-well plates containing 80 µL 2× YT medium with 7.5% glycerol and plasmids were isolated from 96-well plate cultures using a modified alkaline lysis method. Inserts representing the 3′-end of the mRNA were sequenced using the standard M13 forward primer (F: 5′-GTAAAACGACGGCCAG, or B21/C21: 5′-TGTAAAACGACGGCCAG), the T7 promoter primer (B7/C7: 5′-AATACGACTCACAATAG) or primer T8 (5′-(T)23 A/C/G) used to overcome stuttering and slippage during sequencing due to long polyT stretches. Inserts representing the 5′-end of the mRNA were sequenced using M13 reverse primers (BR/CR: 5′-AACAGCTATGACCATG). The 5′-end of the mRNA were sequenced using the modified Eppendorf Direct Bind Kit. At the GSC, clones in E. coli DH10B/r were picked on a Genetix Q-Pix into 384-well plates containing 80 µL 2× YT medium with 7.5% glycerol and plasmids were isolated from 96-well plate cultures using a modified alkaline lysis method. Inserts representing the 3′-end of the mRNA were sequenced using the standard M13 forward primer (F: 5′-GTAAAACGACGGCCAG, or B21/C21: 5′-TGTAAAACGACGGCCAG), the T7 promoter primer (B7/C7: 5′-AATACGACTCACAATAG) or primer T8 (5′-(T)23 A/C/G) used to overcome stuttering and slippage during sequencing due to long polyT stretches. Inserts representing the 5′-end of the mRNA were sequenced using M13 reverse primers (BR/CR: 5′-AACAGCTATGACCATG). The 5′-end of the mRNA were sequenced using the modified Eppendorf Direct Bind Kit. At the GSC, clones in E. coli DH10B/r were picked on a Genetix Q-Pix into 384-well plates containing 80 µL 2× YT medium with 7.5% glycerol and plasmids were isolated from 96-well plate cultures using a modified alkaline lysis method. Inserts representing the 3′-end of the mRNA were sequenced using the standard M13 forward primer (F: 5′-GTAAAACGACGGCCAG, or B21/C21: 5′-TGTAAAACGACGGCCAG), the T7 promoter primer (B7/C7: 5′-AATACGACTCACAATAG) or primer T8 (5′-(T)23 A/C/G) used to overcome stuttering and slippage during sequencing due to long polyT stretches. Inserts representing the 5′-end of the mRNA were sequenced using M13 reverse primers (BR/CR: 5′-AACAGCTATGACCATG).
Sequences (MIPS: http://mips.gsf.de) annotated U. maydis protein database and the corresponding MIPS functional category (FunCat) designation was then used as a basis for further classification. The parsed results of the BLAST reports against the informative NR, FUP and COGEME databases were then added and visually inspected in Microsoft Excel for final categorization.

ACKNOWLEDGEMENTS

We wish to thank Dr Barry Saville for sharing U. maydis EST sequences prior to publication, Dr Les Szabo for DNA probes and Mr Andrew Walsh and Mr Chad Matsalla for help with bioinformatic analyses.

REFERENCES


SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1 Supplementary results: construction of cDNA libraries from different stages of fungal development

Table S1 P. triticina ESTs annotated with various levels of confidence and grouped according to functional categories.

Table S2 Pt ESTs judged most likely to be wheat host-specific, annotated with various levels of confidence and grouped according to functional categories.

Table S3 Annotated P. triticina ESTs encountered uniquely in certain stages.

Table S4 Most abundantly present P. triticina ESTs.

Table S5 Tentative differential expression of annotated genes between several combinations of two indicated stages.

Table S6 Quantitative RT-PCR oligonucleotide primers used in this study.

Table S7 Public resources (with their respective Internet sites) used to generate our custom collections.

This material is available as part of the online article from:


Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.