Analysis of Expressed Sequence Tags from *Uromyces appendiculatus* Hyphae and Haustoria and Their Comparison to Sequences from Other Rust Fungi

D. P. Puthoff, A. Neelam, M. L. Ehrenfried, B. E. Scheffler, L. Ballard, Q. Song, K. B. Campbell, B. Cooper, and M. L. Tucker

First, second, third, sixth, seventh, eighth, and ninth authors: Soybean Genomics and Improvement Lab, United States Department of Agriculture-Agriculture Research Service (USDA-ARS), Beltsville, MD 20705; and fourth and fifth authors: MSA Genomics Laboratory, USDA-ARS, Stoneville, MS 38776. First and second authors contributed equally to this manuscript.

Current address of first author: Department of Biology, 212 Compton Science Center, Frostburg State University, 101 Braddock Road, Frostburg, MD 21532.

Current address of second author: Monsanto Company, Mail Stop: GG5B, 700 Chesterfield Parkway North, Chesterfield, MO 63017.

Accepted for publication 9 July 2008.

ABSTRACT

Puthoff, D. P., Neelam, A., Ehrenfried, M. L., Scheffler, B. E., Ballard, L., Song, Q., Campbell, K. B., Cooper, B., and Tucker, M. L. 2008. Analysis of expressed sequence tags from *Uromyces appendiculatus* hyphae and haustoria and their comparison to sequences from other rust fungi. Phytopathology 98:1126-1135.

Hyphae, 2 to 8 days postinoculation (dpi), and haustoria, 5 dpi, were isolated from *Uromyces appendiculatus* infected bean leaves (*Phaseolus vulgaris* cv. Pinto 111) and a separate cDNA library prepared for each fungal preparation. Approximately 10,000 hyphae and 2,700 haustoria clones were sequenced from both the 5′ and 3′ ends. Assembly of all of the fungal sequences yielded 3,359 contigs and 927 singletons. The *U. appendiculatus* sequences were compared with sequence data for other rust fungi, *Phakopsora pachyrhizi*, *Uromyces fabae*, and *Puccinia graminis*. The *U. appendiculatus* haustoria library included a large number of genes with unknown cellular function; however, summation of sequences of known cellular function suggested that haustoria at 5 dpi had fewer transcripts linked to protein synthesis in favor of energy metabolism and nutrient uptake. In addition, open reading frames in the *U. appendiculatus* data set with an N-terminal signal peptide were identified and compared with other proteins putatively secreted from rust fungi. In this regard, a small family of putatively secreted RTP1-like proteins was identified in *U. appendiculatus* and *P. graminis*.

Additional keywords: expressed sequence tags, gene expression, secretome.

Rust fungi, order Uredinales, cause considerable damage to a broad range of crop plants (25,35,38). These obligate plant pathogens have intricate relationships with their hosts. After penetrating the plant surface through a stomatal pore or directly through the epidermis, an infective hypha grows from a subepidermal vesicle and upon recognition of an appropriate plant cell differentiates into a haustorium mother cell (HMC) (25). The HMC initiates a process that allows it to grow through the plant cell wall and invaginate the host plasma membrane to produce a haustorium that expands inside the host cell. The haustorium is bounded by its own plasma membrane, an extracellular gelatinous-like matrix and the haustorial extracellular membrane, which is an intact but modified derivative of the host plasma membrane (25). The haustorium is responsible for uptake and transport of nutrients from the host to the hyphae (40,41). The hyphae continue to grow intercellularly and initiate secondary infections until finally differentiating into uredia, asexual reproductive bodies. When conditions are right, the uredia can differentiate into telia, sexual reproductive bodies (38). The rusts do not survive by degrading dead cells; they derive their nutrients from a living plant cell (25). Therefore, the interactions with the host to support nutrient exchange at the haustoria must be well orchestrated and highly evolved to reduce natural plant defenses from destroying the haustorium.

The secretion of proteins is crucial to the development of a rust fungus. Secreted proteins could serve several functions: (i) modification of the plant cell wall during penetration by the fungal body, (ii) integration into the host plasma membrane to aid in exchange of plant nutrients and fungal molecules, (iii) enhancement of the synthesis or conversion of plant substrates in the host cell for uptake into fungus, or (iv) suppression of plant defense responses (9,16,24). In addition to being important to the development of the fungus in a susceptible host, the secreted proteins are strong candidates for being avirulence factors that trigger a host resistance response (3,17).

To begin to understand some of the processes associated with the host–pathogen interactions between dry beans (*Phaseolus vulgaris*) and dry bean rust (*Uromyces appendiculatus*), we isolated hyphae from inoculated bean leaves at 2 to 8 days postinoculation (dpi) and from haustoria after 5 dpi. cDNA libraries were prepared using RNA from each of the two preparations. The 5′ and 3′ expressed sequence tags (ESTs) were assembled and compared to assembled ESTs from soybean rust (*Phakopsora pachyrhizi*) germlings, broad bean rust (*Uromyces fabae*) haustoria and in vitro induced infection structures (15,19,24), and open reading frames (ORFs) in the genomic sequence for wheat stem rust (*Puccinia graminis* f. sp. *tritici*). In addition to an

Corresponding author: M. L. Tucker; E-mail address: mark.tucker@ars.usda.gov

*The *e*-Xtra logo stands for "electronic extra" and indicates that the online version contains a link to M. L. Tucker's website where additional data are presented.

doi:10.1094/PHYTO-98-10-1126

analysis and comparison of sequences in the cDNA libraries, we identified ORFs in the *U. appendiculatus* data set with an N-terminal signal peptide that make them eligible for secretion from the fungus. This subset of sequences was analyzed with regard to other putatively secreted rust proteins (21,24).

MATERIALS AND METHODS

Inoculation of plants. Primary leaves of *P. vulgaris* cv. Pinto 111 were sprayed with a water suspension of *U. appendiculatus* race 41 uredospores (30) containing 0.1% Tween 20 and 0.025% β-ionone (13). Spore density yielded 2 to 4 pustules/cm^2 of leaf surface. Inoculated plants were placed in an 18°C dew chamber (Percival Scientific, Perry, IA) overnight and the plants transferred to a growth room at 23°C under fluorescent lighting.

Isolation of hyphae. Inoculated plant leaves were collected at 2, 4, 6, and 8 dpi. Fungal hyphae were extracted from infected leaves using a protocol adapted from Dekhuijzen et al. (7). Briefly, 10 g of leaves were homogenized in a blender with 80 ml of 0.3 M sucrose, 3 mM $MgSO_4$, and 0.1 M KH_2PO_4, adjusted to pH 6.5 with Na_2HPO_4. The homogenate was filtered sequentially through 600-, 250-, 90-, and 45-μm sieves and the mycelia collected from the 45-μm mesh screen. Material that did not pass through the 600-, 250-, and 90-μm screens was collected and homogenized again at a higher blade speed and passed through the same series of sieves to increase the amount of mycelia collected from the 45-μm screen. These fractions consisted of mostly hyphae, some uredia in the 8 dpi sample, chloroplasts, and other plant debris such as xylem and broken cells (6). Further separation of the fractions through sucrose density gradients as described by Dekhuijzen et al. (7) did not work well in our hands; however, the number of contaminating chloroplasts were reduced by extensive washing of the hyphae with homogenization buffer while still on the 45-μm mesh screen. The hyphal fraction was concentrated by centrifugation and the pellets frozen in liquid nitrogen and stored at –70°C.

Isolation of haustoria. Leaves were collected at 5 dpi and haustoria isolated using a protocol adapted from Hahn and Mendgen (15). Leaves were rinsed thoroughly to remove non-germinated spores on the surface and 15 g of leaf tissue homogenized in a blender at high speed for 30 s in 100 ml of 0.3 M sorbitol, 20 mM MOPS, 0.2% PVP, 0.1% BSA, and 0.2% mercaptoethanol adjusted to pH 7.2. The mixture was filtered sequentially through 600-, 250-, 90- and 45-μm sieves and the filtrate that passed through the 45-μm sieve was filtered twice through double layers of miracloth to further remove contaminating hyphae and cell debris. The suspension of chloroplasts and haustoria was centrifuged at 5,000 g for 10 min at 4°C. The pellet was suspended in 3 ml of 0.3 M sorbitol, 10 mM MOPS, 0.2% BSA, 1 mM $CaCl_2$, and 1 mM $MnCl_2$ adjusted to pH 7.2, and loaded onto a column of Con-A Sepharose 6MB equilibrated in 0.15 M $NaCl_2$, 10 mM Tris, 1 mM $CaCl_2$, and 1 mM $MnCl_2$ adjusted to pH 7.2. After slowly loading sample into the column bed, the column was plugged and allowed to stand for 15 min to allow the haustoria to bind to the Con-A. The column was gently washed with the pellet suspension buffer at a slow flow rate to wash out chloroplasts. The column was then plugged and 3 ml of pellet suspension buffer added to the surface of the column bed. The Sepharose beads in the column were agitated with a plastic wide bore Pasteur pipette to release the haustoria from the beads. Beads were allowed to settle and the top suspension containing haustoria was collected. This release step was repeated seven to nine times, each time adding 3 ml of suspension buffer. The pooled haustoria suspension was concentrated by microcentrifugation at high speed for 2 min. The pellets were frozen in liquid nitrogen and stored at –70°C.

RNA isolation, cDNA library construction, sequencing, and cleanup. RNA was purified from the material using a standard guanidinium isothiocyanate protocol (22). Hyphae were collected at 2, 4, 6, and 8 dpi and the RNA isolated and pooled for preparation of a cDNA library at The Biotechnology Center, University of Illinois, Urbana, IL. Briefly, first- and second-strand cDNA was synthesized from 2 μg of RNA using an oligo d(T) primer with a *Not*I restriction site. *Eco*RI adaptors were ligated to both ends of the cDNAs and the cDNAs directionally inserted into pBluescript (Invitrogen, Carlsbad, CA). The library was normalized using previously described methods (31). The haustorial library was prepared using the Creator Smart cDNA Synthesis kit (Clontech, Mountain View, CA). The *Sfi*I-digested cDNA was inserted into the *Sfi*I site of the pDNR-LIB Donor Vector (Clontech). Clones from both libraries were arrayed into 384-well plates and plasmid DNA isolated and sequenced using ABI BigDye Terminator chemistry (version 3.1; Applied Biosystems, Foster City, CA) on an ABI 3730xl DNA sequencing machine with 50-cm arrays.

We sequenced approximately 10,700 hyphae cDNAs and 2,600 haustoria cDNAs from both the 5′ and 3′ ends. Phred was used to convert the electropherogram files to base and quality files (11,12). These files were examined with LUCY, which removed vector sequence and confirmed that each sequence was of high quality and read length (4). Approximately 10% of the sequences were removed because they had low Phred scores or because the cDNA insert was shorter than 100 bp. The LUCY output was transformed to FASTA format using a modified version of the TIGR Babaloo script. The sequences were assembled using Seqmerge in the Genetics Computer Group (GCG) software package (Accelrys, San Diego, CA). To assure an efficient assembly of contiguous sequences, the Seqmerge assembly application was iterated three times using the same settings for each cycle (word 7, stringency 0.85, overlap 20, identity 20).

In order to identify and remove plant sequences in the ESTs, the hyphal and haustorial sequences were assembled together and the contigs (consensus sequences) and singletons used in BLAST comparisons against the National Center for Biotechnology Information (NCBI) nucleotide and protein sequences. Each search of the NCBI databases was limited to flowering plants (division Magnoliophyta) or fungi (kingdom Fungi). If the BLAST bit score was ≥100 in the plant search and the plant bit score was greater than any fungal match, the contig, including all of the aligned ESTs, or singleton was removed from the *U. appendiculatus* data set. If the plant and fungal scores were nearly equal, as occurs for the highly conserved sequences (e.g., ubiquitin), the sequence was marked as fungal. Approximately 10% of the contigs and singletons matched plant sequences and were removed from the data set leaving 4,286 contigs and singletons. The cleaned assemblies were then disassembled and submitted to GenBank as individual ESTs (accessions EH294895-EH306219 and FE660223-FE668377).

Assembly, annotation, acquisition of other rust ESTs, and software applications used. The soybean rust, *P. pachyrhizi*, (34,394 ESTs) (32) and *U. fabae* (601 ESTs) (15,19,24) were downloaded from NCBI in December 2007 and assembled in Seqmerge using the same settings used for the *U. appendiculatus* ESTs. In the *U. fabae* ESTs there were 86 sequences identified as "stage specific secretomes" (24). These putative secretome sequences were included in a separate UfSecretome data set. ORFs discovered in the recently completed wheat stem rust, *P. graminis*, genomic sequence were downloaded in December 2007 from the *P. graminis* Sequencing Project at the Broad Institute of Harvard and MIT. The GCG GenPept protein database (5,176,369 sequences) used for global BLAST comparisons was compiled by Accelrys from the NCBI nonredundant nucleotide database (release 16.0, 6/2007).

As a means of ranking the quality and uniqueness of the BLAST alignments, the bit score was used rather than the expect value because the bit score is less effected by the database size,

which varied from 500 for the *U. fabae* data set to over 5 million in the GenPept data set. To put this into perspective with the more familiar expect value, a bit score of 100 is approximately an expect value of –20 for a large database like the GenPept database. A bit score of 100 or greater was chosen because it is reasonably conservative, requiring a fairly good match between two sequences.

Cellular process annotation of the *U. appendiculatus* EST assemblies was completed using a BLASTX comparison of UaAll, ESTs to the yeast (*Saccharomyces cerevisiae*) and UniprotSwiss protein databases that included gene ontology (GO) slim annotation. The yeast database included 6,200 yeast sequences downloaded on October 2007 and the UniprotSwiss comparisons were completed at the GOanna website using a BLAST bit score cutoff of 70. When there was a discrepancy between the yeast and UniprotSwiss annotation and the bit scores were similar, the yeast annotation was given preference. Some of the GO slim categories were combined for simplicity. For example, the processes for cellular respiration, tricarboxylic acid cycle, and electron transport were combined into an energy metabolism category. ESTs with homology to retroelements were not well represented in the GO databases and were therefore assigned a function based on the most similar sequence in the GenPept database. The retrotransposon category includes sequences with a BLASTX hit description that included the words or phrases polyprotein, pol protein, retroelement, retrotransposon, retrotransposable, reverse transcriptase, retro-virus related, RT-like, integrase, RNA-directed DNA polymerase, transposase, and transposon.

The GCG software Frames and SPScan were used to identify ORFs and putative N-terminal signal peptides, respectively, in the assembled and cleaned *U. appendiculatus* data set. The bean, soybean, and wheat stem rust files were converted into BLAST searchable databases with GCG FormatDB. Other non-GCG software packages used with default settings were: (i) iPSORT (1) and SignalP (10) for signal peptide identification, (ii) WoLF PSORT (18) and LOCtree (28) were used for subcellular localization in a plant cell, (iii) NucPred (2) for nuclear localization prediction, and (iv) PredictNLS (5) to determine if a nuclear localization signal (NLS) was present in the protein sequence.

Reverse transcription-polymerase chain reaction (RT-PCR). RT-PCR was performed as previously described (36). Gene-specific primers used to amplify *P. vulgaris* and *U. appendiculatus* translation elongation factors 1α and the *U. appendiculatus* RTP1-like sequence were: PvEF1a-f, GAGAAGGAGCCTAAGTTCTTGAAGAATG; PvEF1a-r, TGACACAAGAGTACAAGCAAGAAACC; UaEF1a-f, TGGTTCAAGGGATGGTCTAAGGAG; UaEF1-r, CATTGTCACCAGGCATACCAGC; UaRTP1-f, ACCGGAAGTCTACAAGTTTCTCCA; UaRTP1-r, GGCCGAGATAAGTATCGAACACAG.

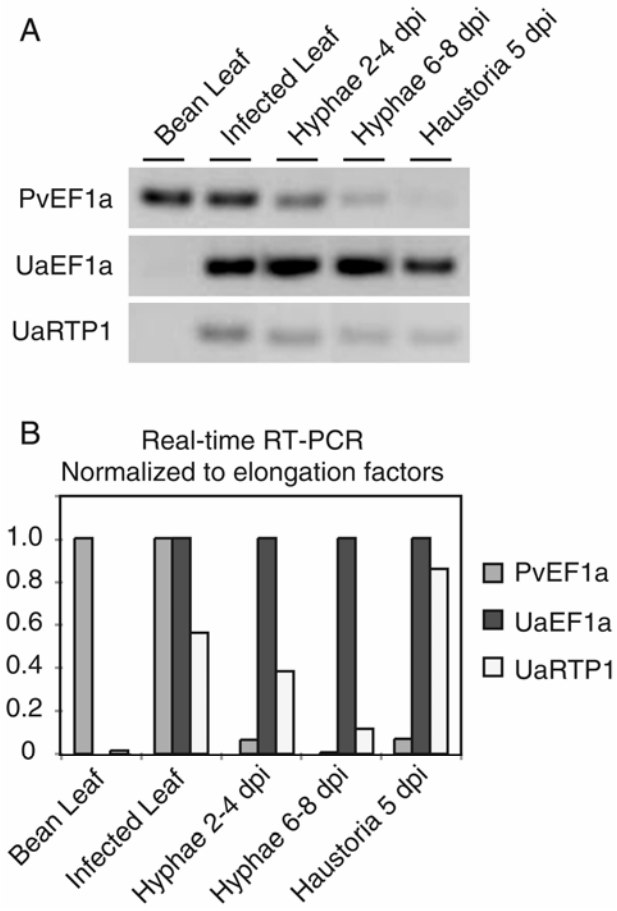

Fig. 1. Determination of fungal isolate purity by reverse transcription-polymerase chain reaction (RT-PCR). Gene-specific primers were prepared for *Phaseolus vulgaris* elongation factor 1α (PvEF1a), *Uromyces appendiculatus* elongation factor 1α (UaEF1a) and *Uromyces appendiculatus RTP1* (UaRTP1). **A,** Agarose gel separation of PCR amplification after 30 cycles. **B,** Real-time RT-PCR normalized to PvEF1a or UaEF1a. The relative concentration of the PvEF1a in each sample was normalized to the concentration of PvEF1a in the noninfected bean leaf. The UaEF1a and the UaRTP1 concentrations were normalized to the concentration of UaEF1a in each sample, i.e., UaEF1a concentration equals 1.

RESULTS

RNA purity and assembly of ESTs from rust infected bean leaves. Fungal hyphae were isolated from *U. appendiculatus*-infected leaves between 2 to 8 dpi. Greater amounts of hyphae were isolated from leaves at the later time points due to greater colonization of the leaves at these times. Haustoria were isolated from leaves 5 dpi. Attempts to isolate haustoria at earlier time points were not successful because of very low yields. RNA was extracted from the preparations and RT-PCR used to determine the relation between the amounts of fungal material and any residual plant cell material. Primers specific to the *P. vulgaris* and *U. appendiculatus* translation elongation factors (*PvEF1α* and *UaEF1α*, respectively) were used to estimate the amount of plant RNA relative to fungal RNA. RT-PCR analysis of the amounts of *PvEF1α* transcript suggested that the RNA population from the fungal preparations were comprised of 2 to 8% plant RNA (Fig. 1). This is approximately equivalent to the number of plant sequences identified by BLAST comparisons.

RNA isolated from the hyphal and haustorial preparations were used to make two separate cDNA libraries and several thousand clones from each library were sequenced. After assembly of all EST sequence data and removal of sequences with identity to plant genes, there remained 4,286 contigs and singletons, which were presumed to be of fungal origin (data set named UaAll, supplemental data UaAll.fa). The UaAll data set was then disassembled into individual ESTs and the hyphae and haustoria library-derived sequences reassembled separately, yielding 3,136 contigs and singletons from the hyphal library (named UaHyp, supplemental data UaHyp.fa) and 1,300 sequences from the haustorial library (named UaHaust, supplemental data, UaHaust.fa). The average length of the UaAll, UaHyp and UaHaust sequences were 772, 846, and 577 nts, respectively.

Categorization by cellular processes. The UaAll, UaHyp, and UaHaust ESTs were annotated by BLASTX comparisons to the yeast (*Saccharomyces cerevisiae*) and UniprotSwiss protein databases. The number of ESTs (fragments) in each assembly was summed and the percent of fragments in each cellular process was tabulated to illustrate the number of ESTs linked to each category (Fig. 2). Approximately 0.5% of the cDNA clones in the hyphae library encoded fungal rRNAs, whereas the haustorial library contained approximately 30% (supplemental data UaAll.xls). To

better illustrate differences in the nonribosomal RNA categories, rRNA was excluded from the graphical representation and the statistical analysis (Fig. 2A). A student's *t* test was applied to the grouped data to determine if the fragment number (sequences) in each category was significantly different between the hyphal and haustorial libraries at a probability of 0.05 with or without a Bonferroni correction (26,33). If all of the categories were included in the statistical analysis, most of the categories were determined to be significantly different between the two libraries (supplemental data UaAll.xls). However, the number of sequences that could not be assigned a cellular process (category unknown) was considerably greater in the haustorial library compared with the hyphal library, 85% compared with 58%, respectively (Fig. 2). In addition to the unknown category, the category for retrotransposons was also disproportionate between the two libraries. The large differences in these categories tend to distort comparisons for the more biologically relevant categories. Therefore, the data was further analyzed excluding the unknown and retrotransposon categories. In this case, fewer categories were significantly different between the hyphal and haustorial libraries (Fig. 2B). Transcripts in the categories for amino acid metabolism, cellular regulation, protein metabolism and translation significantly decreased in haustoria relative to hyphae whereas transcripts in cell structure, energy metabolism, and vitamin metabolism significantly increased in haustoria. The biological relevance of these differences will be discussed below.

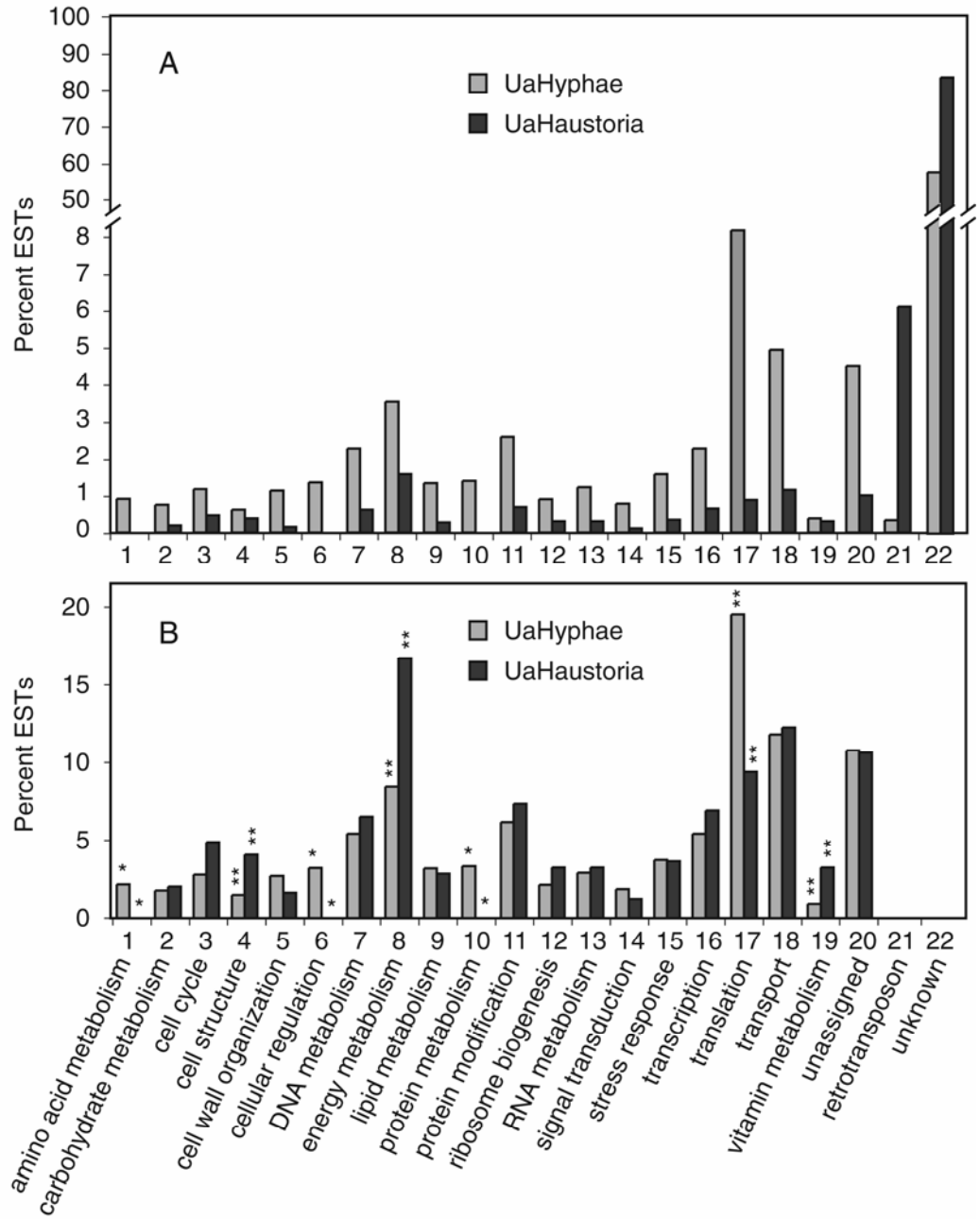

Fig. 2. Percentage of expressed sequence tags (ESTs) in the *Uromyces appendiculatus* data set categorized by cellular process. Categorization was performed as described in the text using the consensus sequence for each assembly (contig) and the total number of sequences in each assembly used for calculation of the percent of ESTs in each category. **A,** The category for ribosomal RNA was excluded from the tabulation of percents. The retrotransposon category includes any GenPept match with the words or phrases polyprotein, pol protein, retroelement, retrotransposon, retrotransposable, reverse transcriptase, retro-virus related, RT-like, transposase, and transposon. **B,** Categories for ribosomal RNA, unknown, and retrotransposon were excluded from the tabulation of percents. *Categories in which the hyphae and haustoria were significantly different in a *t* test at $P < 0.05$. **Categories in which the hyphae and haustoria were significantly different in a *t* test with a Bonferroni correction for $P < 0.05$.

BLAST search of other rust libraries. As a means of determining both shared and unique sequences in our *U. appendiculatus* ESTs with respect to other rusts, we performed a series of BLAST comparisons. The ESTs (NCBI, 1 December 2007) for soybean rust (*P. pachyrhizi*) uredospores and germlings (32) were assembled with Seqmerge to produce 5,902 contigs and singletons with an average length of 846 nt (named PpAll, supplemental file PpAll.fa). The ESTs (NCBI, 1 January 2008) prepared from the haustoria of broad bean rust (*U. fabae*) (15) and a putative *U. fabae* secretome (24) were similarly assembled (named UfAll). The average length of the UfAll sequences was 621 nt. The Broad Institute recently released nucleotide and amino acid sequence for ORFs in the *Puccinia graminis* f. sp. *tritici* genomic sequence. This data set (PgAll) includes 20,567 ORFs with an average length of 1,104 nt (368 aa). The average length of the predicted *Puccinia* ORFs was approximately 40% longer than the assembled *U. appendiculatus* (UaAll), *U. fabae* (UfAll), and *P. pachyrhizi* (PpAll) data sets. This suggests that a large fraction of the assembled EST sequences do not encompass the full-length transcript.

The assembled *U. appendiculatus* (UaAll, UaHyp, and UaHaust), *U. fabae*, and *P. pachyrhizi* ESTs and the *P. graminis* ORFs were converted into BLAST searchable databases and cross comparisons were made. In addition, the rust sequences were used as queries to search the GCG GenPept database. The number of matches with a bit score greater than 100 as a percent of the total number of query sequences in the database are provided in Table 1 and the numbers of unique matches with a bit score greater than 100 are shown in Figure 3. It is worth noting that using a threshold bit score of 100 for a tBLAST(X) search of a data set against itself did not return a 100% similarity score (Table 1). This is because the alignment of shorter sequences of 100 to 200 nt that include repetitive sequences result in bit scores of less than 100. The 96% matching for the tBLASTX comparison of UaHaust to itself indicates that this data set includes more repetitive sequences than the others (Table 1). Also of importance in our tabulation of matches is that if two or more different query sequences matched the same subject sequence, then the multiple subject matches were counted as one unique match. This tabulation criteria results in a lower percentage of matches counted for members of a gene family (Table 1).

Based on BLAST comparison between the data sets for *U. appendiculatus* and *U. fabae*, which are closely related species (37), there was approximately 30% overlap between the two data sets (Table 1). Comparison of the *U. appendiculatus*, *U. fabae*, and *P. pachyrhizi* data sets to the *P. graminis* genomic ORFs indicated that approximately 38% of the sequences are related using the threshold bit score of 100. However, when the *U. appendiculatus* haustoria sequences (UaHaust) were compared separately to the *P. graminis* data set (PgAll), the percent of similar sequences common to both dropped to 9%. The percentage dropped even further when the haustoria sequences were compared to the GenPept and *P. pachyrhizi* data sets (Table 1). The significance of the low percentage of matches for the haustoria ESTs will be discussed in greater detail below.

Identification of N terminus signal peptides. To identify genes for proteins that might be secreted from the hyphae and haustoria and thereby potentially serve as virulence and avirulence factors, we searched the UaAll data set for ORFs that included a putative N-terminal signal peptide. The UaAll se-

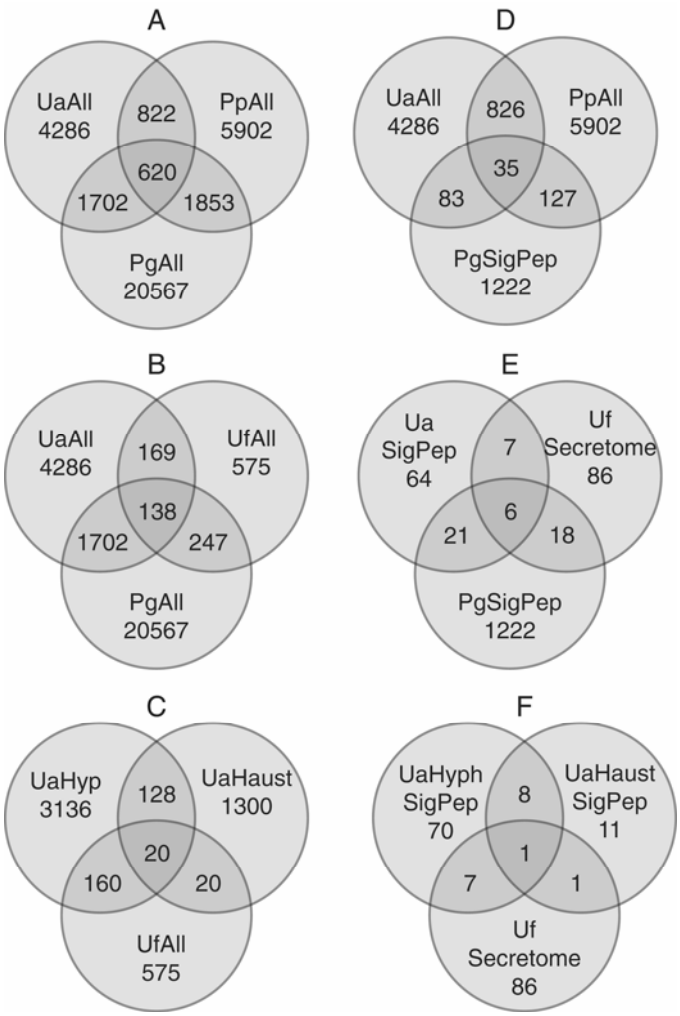

Fig. 3. Venn diagrams displaying the number of query sequences in each data set and BLAST matches with a bit score of 100 or greater shared in two or three of the data sets. When the same sequence in the searched database (subject) was matched by multiple queries, the subject hit was only counted once. Ua, *Uromyces appendiculatus*; Uf, *U. fabae*; Pp, *Phakopsora pachyrhizi*; and Pg, *Puccinia graminis* f. sp. *tritici*.

TABLE 1. BLAST matches with a bit score ≥100 as a percentage of the total number of query sequences that match a unique sequence in the subject database[a]

Query (total sequences)	Subject matches with ≥100 bit score as a percentage of query total						
	UaAll	UaHyp	UaHaust	UfAll	PpAll	PgAll	GenPept[b]
UaAll (4,286)	98	71	28	5	19	40	24
UaHyp (3136)	97	99	4	5	21	52	33
UaHaust (1,300)	93	10	96	2	4	9	5
UfAll (575)	30	29	4	99	22	43	25
PpAll (5,902)	12	12	1	2	99	31	26
PgAll (20,567)	8	8	1	1	9	99	18

[a] Each data set was converted into a BLAST searchable database. All BLAST searches were for a protein query compared to a protein subject. In other words, the nucleotide query and/or the nucleotide subject were translated and compared in all six frames (tBLASTX or BLASTX). The PgAll and GenPept databases were already protein and therefore not translated (BLASTP or tBLASTN). Only the highest scoring hit for all the frames was counted. When the subject match was the same for two or more different queries, the match was only counted once.

[b] The GenPept database is the translated nonredundant GenBank core nucleotide database as of June 2007.

quences were translated in all six reading frames and any ORF greater than 100 aa was then searched with the GCG SPScan software for the presence of a signal peptide. Predicted signal peptides longer than 35 aa were eliminated from the output as were sequences where the signal peptide did not start with the first ATG in the ORF or were more than 100 nt from the beginning of the sequence. This operation yielded 85 sequences with putative signal peptides. The selected ORFs were then tested

TABLE 2. Open reading frame (ORFs) greater than 100 amino acids in the *Uromyces appendiculatus* data set that passed all three tests for the presence of a signal peptide[a]

U. appendiculatus assembly	ORF length	BLASTX of NCBI nr protein database			BLASTX of *P. graminis*			tBLASTX of *U. fabae*		
		Protein hit	Hit description	Score	Hit	SP	Score	Hit	Sec	Score
UaHaust_02c_f_d11	231	AAA34217	Infection structure-specific	66	PGTG_11182	Y	50	gi_164245285	Y	71
UaHaust_03d_f_f08	197	NP_588156	Protease inhibitor	47	PGTG_14060		104			
UaHaust_05c_f_c01	115	YP_001538288	Alpha-1,2-mannosidase	31	PGTG_08763		28			
UaHyp_02b_f_f12	247	CAE85538	Probable gEgh 16	67	PGTG_12331	Y	311			
UaHyp_02c_f_c07	202	AAW40663	COPII-coated vesicle	206	PGTG_11514		279			
UaHyp_02c_f_g02	104			39	PGTG_12874		32	gi_164246332	Y	62
UaHyp_02d_f_e07	218	XP_747144	Cell surface MAS1 protein	54	PGTG_12330	Y	341			
UaHyp_04b_f_a06	197	XP_764729	Hypothetical protein	39	PGTG_19738	Y	107	gi_164245272	Y	123
UaHyp_05b_f_d12	127	AAW45754	Fumarate reductase (NADH)	167	PGTG_17753	Y	224			
UaHyp_05b_r_d04	174	YP_595309	Transcription antiterminator	32	PGTG_03412		30	gi_164245275	Y	64
UaHyp_05d_f_f03	473	AAW46674	Expressed protein	208	PGTG_01123		514			
UaHyp_06a_f_c04	231	AAA34217	Infection structure-specific	48	PGTG_11182	Y	39	gi_164245285	Y	53
UaHyp_06a_f_f02	111	YP_546645	Membrane autotransporter	32	PGTG_03694	Y	98	gi_164246325	Y	112
UaHyp_06b_f_f08	138	XP_001616827	AAA family ATPase	34	PGTG_18157		31			
UaHyp_06c_f_a11	132	XP_716953	Protein disulfide isomerase	120	PGTG_04854	Y	169	gi_66652163		259
UaHyp_06c_f_b10	103	ZP_02163015	b-N-acetylglucosaminidase	35	PGTG_04027		32	gi_164242089	Y	42
UaHyp_06c_f_e09	121	ZP_00955518	Hypothetical protein	32	PGTG_06590		30	gi_164242089	Y	42
UaHyp_07c_f_e01	186	YP_001373580	Extracellular solute-binding	32	PGTG_08993		32			
UaHyp_08c_f_a09	296	XP_712007	Putative ER co-chaperone	118	PGTG_20008	Y	328			
UaHyp_09a_f_f05	229	XP_001555760	Hypothetical protein	95	PGTG_18569		288			
UaHyp_09b_f_d01	203	XP_778235	Hypothetical protein	91	PGTG_16691	Y	114	gi_164243166	Y	44
UaHyp_10a_f_h09	127	YP_434834	Endoglucanase C-terminus	73	PGTG_19856	Y	162	gi_164244213	Y	100
UaHyp_10d_f_g05	355	XP_572915	Lectin	244	PGTG_12132		286			
UaHyp_11c_f_b10	103							gi_164246332	Y	63
UaHyp_11c_r_e12	127	XP_001523188	Hypothetical protein	33	PGTG_08470	Y	41	gi_164277370	Y	68
UaHyp_14a_f_g10	181	NP_009160	Protein kinase C	36	PGTG_18238	Y	132			
UaHyp_14b_r_d07	232	AAD38996	Differentiation-related prot	241	PGTG_17906	Y	129	gi_164246329	Y	40
UaHyp_14d_f_a08	124	AAA34218	INF56	142	PGTG_15800		32			
UaHyp_15c_f_g09	185	ABN44257	Platelet-binding glycoprotein	40	PGTG_18012		32	gi_164245276	Y	58
UaHyp_16c_f_h09	220	XP_750110	Exo-beta-1,3-glucanase	94	PGTG_17056	Y	341			
UaHyp_16d_f_b06	336	AAW41996	Glycotransferase	216	PGTG_01612	Y	503			
UaHyp_17d_f_d10	241	NP_001093267	Phosphatidyleth. binding	45	PGTG_11450		192			
UaHyp_19a_f_d09	208				PGTG_18138	Y	37			
UaHyp_19a_f_f11	104									
UaHyp_19a_f_g01	189	AAA34217	Infection structure-specific	70	PGTG_11320	Y	65	gi_164245285	Y	160
UaHyp_19a_r_h01	247	NP_849913	TOPLESS-RELATED 1	34	PGTG_06969	Y	213	gi_164277366	Y	182
UaHyp_19b_f_f10	121	YP_595309	Transcription antiterminator	34	PGTG_13020		27	gi_164245275	Y	62
UaHyp_19b_r_d06	392	AAW47222	Glucan 1,3 beta-glucosidase	279	PGTG_02069	Y	516			
UaHyp_19b_r_g08	377	BAE18209	DNA polymerase III alpha	39	PGTG_17067		72			
UaHyp_19c_f_h02	107				PGTG_01812		30	gi_164242089	Y	39
UaHyp_19d_r_a11	256	CAA43289	PriA	90	PGTG_00898	Y	314	gi_164244226	Y	152
UaHyp_20c_f_f11	189	AAA34217	Infection structure-specific	71	PGTG_04810	Y	64	gi_164245285	Y	163
UaHyp_20c_r_a03	174				PGTG_16543		32	gi_164245275	Y	60
UaHyp_20d_f_g07	213	CAI96536	RTP1, rust transferred prot	136	PGTG_17519	Y	181	gi_66651749		85
UaHyp_21b_r_a05	100	XP_571431	Glycogen phosphorylase	31	PGTG_06549		34	gi_164242089	Y	43
UaHyp_21c_f_b12	175	CAJ81221	Unnamed protein product	74	PGTG_05076	Y	481			
UaHyp_22a_f_a09	332	CAD21425	Related to stress response	165	PGTG_06207		436	gi_164244227	Y	234
UaHyp_22b_f_a09	196	AAA34217	Infection structure-specific	71	PGTG_11320	Y	58	gi_164245285	Y	122
UaHyp_22b_f_e09	154	AAW43537	Conserved hypothetical	49	PGTG_16086		31			
UaHyp_22b_f_h06	312	CAD27464	SPAPB15E9.01c	50	PGTG_04691	Y	47	gi_164246331	Y	57
UaHyp_22d_r_e01	101									
UaHyp_23c_r_g10	122	AAK31230	Variable surface protein 14d	38	PGTG_03694	Y	137	gi_164246325	Y	161
UaHyp_23d_r_d10	174	NP_987559	Dihydroorotate dehydro.	34	PGTG_16543		29	gi_164245275	Y	61
UaHyp_24c_r_g06	174	YP_595309	Transcription antiterminator	32	PGTG_03412		30	gi_164245275	Y	64
UaHyp_24d_f_h05	187	ZP_00531579	ApbE-like lipoprotein	34	PGTG_14762	Y	132			
UaHyp_24d_r_b01	181							gi_164245276	Y	56
UaHyp_25a_r_c11	105				PGTG_18970		26	gi_164246333	Y	39
UaHyp_25b_r_c09	150	AAW41301	Conserved hypothetical	62	PGTG_17690	Y	108	gi_164245278	Y	120
UaHyp_25d_f_a05	188	YP_341415	Glucosyltransferase	32	PGTG_03412		30	gi_164245275	Y	64
UaHyp_25d_f_e02	210	CAA96325	AGA1	41	PGTG_02990	Y	31			
UaHyp_25d_f_g08	105	YP_022987	Membrane serine protease	38	PGTG_06549		37			
UaHyp_27c_r_g04	271	XP_001271461	CFEM domain protein	41	PGTG_01207	Y	72			
UaHyp_27d_r_c10	173	XP_001015884	Nucleoside diphos. kinase	36	PGTG_16543		32	gi_164245275	Y	64
UaHyp_28a_r_d04	153	XP_570934	FK506 binding protein 2	149	PGTG_00244	Y	155	gi_66652124		231

[a] The best BLAST match is listed for the National Center for Biotechnology Information (NCBI) nonredundant protein database as of January 2008 and the *Puccinia. graminis* and *Uromyces fabae* data sets. If the *P. graminis* ORF was predicted to possess an N-terminal signal peptide (SP), it is indicated with a Y. Similarly, if the best *U. fabae* match was in the putative secretome (sec) data set, it is marked with a Y.

again for a signal peptide with iPSORT (1) and SignalP (10). Only 64 ORFs passed all three predictive tests for the presence of a signal peptide at the N terminus (Table 2). This data set was named UaSigPep.

For comparison, in addition to *U. appendiculatus*, all of the *P. graminis* ORFs were searched for signal peptides with SPScan, iPSORT, and SignalP. Out of the 20,567 ORFs in *P. graminis*, 1,222 ORFs were predicted by all three algorithms to include an N-terminal signal peptide (data set named PgSigPep). The *U. appendiculatus* signal peptide data set (UaSigPep), the *P. graminis* signal peptide data set (PgSigPep), and the 86 *U. fabae* NCBI ESTs labeled as "stage specific secretomes" (UfSecretome) were compared using BLAST (Fig. 3E and F; Table 2). If we were fairly conservative about deciding which proteins in the data sets are similar, (i.e., we used a bit score of 100 or greater), then only a few proteins were common to each of the rusts (Fig. 3D, E, and F). However, if we simply listed the best BLAST match with an expect value greater than 0.0001 (an approximate bit score ≥25), then many more of the *U. appendiculatus* ORFs matched a *P. graminis* or *U. fabae* ORF with a predicted signal peptide (Table 2).

RTP1-like sequences. When we started this project, we relied on the sequence information for *U. fabae* and *U. striatus* RTP1, a protein that is secreted from haustoria into the host (21), to estimate by RT-PCR the purity of our *U. appendiculatus* haustoria preparation. We prepared PCR primers to conserved regions in *UfRTP1* and *UsRTP1* and, after sequencing an amplified fragment, prepared new primers specific for the UaRTP1-like transcript. RT-PCR analysis using these specific primers suggested that the haustoria purification steps resulted in a clear enrichment of these cell structures compared with the more generic hyphal purification procedure (Fig. 1). It also indicated that there was a higher concentration of the RTP1 transcript in the 2 to 4 dpi hyphal preparation compared to the 6 to 8 dpi preparation. This was likely due to the fact that by 8 dpi uredia were beginning to form on the bottom of the leaves and this additional fungal material diluted the amounts of haustorial RNAs relative to RNAs from other fungal structures.

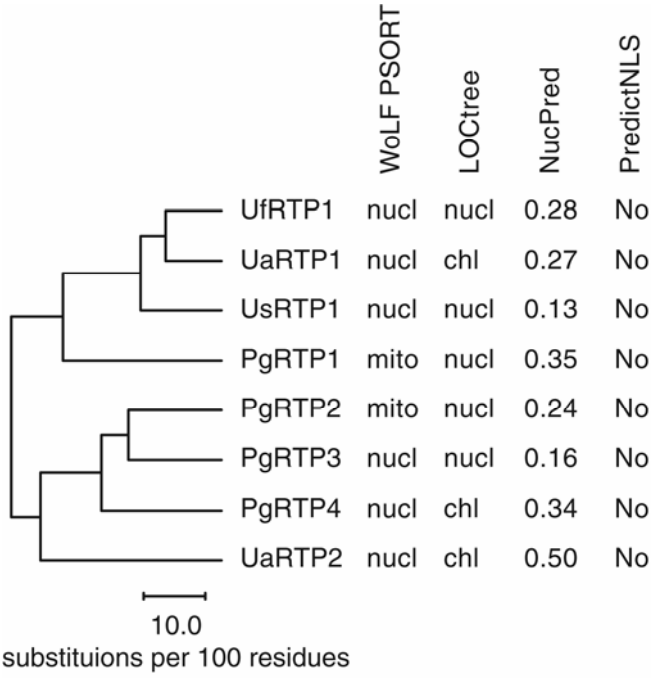

Fig. 4. Dendrogram displaying the sequence relationship of proteins in *Uromyces striatus*, *U. appendiculatus*, and *P. graminis* that have similarity to the *U. fabae* rust transferred protein (UfRTP1). The dendrogram was prepared using the Genetic Computer Group (GCG) software Distances and Growtree with the uncorrected and unweighted pair-group method with arithmetic average (UPGMA) options selected. Cellular localization and nuclear targeting website predictions (WoLF PSORT, LOCtree, NucPred, and PredictNLS) were performed using the peptide sequences with the predicted signal peptides removed. Abbreviations: nucl, nuclear; chlo, chloroplast; mito, mitochondria. Only the top score for subcellular localization is given (UsRTP1 had an equal score for nuclear and chloroplast). A NucPred score of 0.1 or 0.5 indicates that approximately 45 or 70%, respectively, of the proteins predicted to be nuclear are actually nuclear.

In addition, after library construction, sequencing, and EST assembly the *U. appendiculatus*, *P. pachyrhizi*, and *P. graminis* sequences were searched for translated ORFs similar to the *U. fabae* RTP1 protein sequence. We also searched the NCBI *Phaseolus vulgaris* ESTs, which includes ESTs from a cDNA library of *P. vulgaris* leaves infected with *U. appendiculatus* (B. Cooper and G. Stacey, *unpublished data*). One RTP1-like sequence was found in the assembled *U. appendiculatus* ESTs, another in the *P. vulgaris* ESTs, and four different genes in *P. graminis* (Figs. 4 and 5). No sequence similar to RTP1 was found among the *P. pachyrhizi* ESTs, but this was not surprising since these ESTs were derived from germinated spores, which should not express haustoria-specific genes. Alignment and phylogenetic analysis of the sequences indicated that there are at least two clades for the RTP family of proteins (Fig. 4). As reported previously for the alignment of UfRTP1 and UsRTP1 (21), the C-terminal halves of the proteins are well conserved while the N-terminal halves are more diverse (Fig. 5). The RTP1-like sequence identified in the leaf-infected library (accessions FE684103 and FE684104 for 5′ and 3′ ends, respectively) is missing part of the 5′ end of the sequence presumed to include a putative signal peptide (Fig. 5, UaRTP1). However, all of the other RTP1-like protein sequences passed at least two of three test programs for the presence of a signal peptide.

Kemen et al. (21), using polyclonal antibody, demonstrated that UfRTP1 accumulates in the host nucleus. To determine if the other RTP1-like proteins might also be targeted to the host nucleus we submitted each of the predicted ORFs with the signal peptide removed to multiple online algorithms that predict the subcellular localization or presence of nuclear localization signal motifs. WoLF PSORT (18), LOCtree (28), and NucPred (2) make predictions of subcellular localization based on similarity comparisons to proteins of known subcellular localization, amino acid composition, and sequence motifs. PredictNLS (5) searches for documented nuclear localization signals in the proteins. The most probable cellular localization predicted by WoLF PSORT and LOCtree for the RTP1-like proteins was mostly nuclear but for some the WoLF PSORT prediction was different from the LOCtree prediction (Fig. 4). The NucPred scores indicated that all of the RTP1-like proteins might be nuclear, i.e., a score of 0.1 accurately predicts nuclear localization 45% of the time (Fig. 4). PredictNLS, however, did not find any typical nuclear localization signals (NLS) in any of the RTP1-like proteins (Fig. 4).

The RTP1 primers we designed match perfectly with *UaRTP1* but do not match well with the nucleotide sequence for another RTP1-like protein discovered in the hyphal ESTs that we named UaRTP2 (Figs. 4 and 5). RT-PCR suggested that our haustoria preparations were enriched for the *UaRTP1* mRNA relative to the hyphal preparation (Fig. 1) but no *UaRTP1* EST was sequenced from either the haustoria or hyphal libraries. This may be due to random but limited selection of clones from the library. Curiously, all 59 copies of the *UaRTP2* ESTs came from the hyphal library and none from the haustorial library.

DISCUSSION

The genome size of rust fungi varies markedly (8). The genome of *P. graminis* has one of the smaller genomes (8). The Broad Institute estimated the size of the *P. graminis* genome at 81.5 Mb. The DOE Joint Genome Institute (JGI) estimated that the *P.*

pachyrhizi genome is close to 800 Mb, 10-fold larger than *P. graminis*. Eilam et al. (8) estimated that the genome of *U. appendiculatus* is 6.9 times larger than *P. graminis*, i.e., 560 Mb. Until these large genomes are sequenced, genomic comparisons between important rusts with unsequenced genomes are limited to EST analysis, which in addition to sequence information, can provide important insight regarding rust biology.

The June 2007 GenPept database we used for comparisons did not include ORFs for the *P. graminis* genomic sequence or any EST data. Nevertheless, it was surprising that only 24% of the *U. appendiculatus* ESTs matched sequences in GenPept. Subsequent searches with the other rust data sets also produced a low number of matches (Table 1). This suggests that the genomic and EST sequence data compiled for rusts is unique. This observation became more pronounced when we limited our examination to ESTs from the haustorium library and implies that haustoria specific genes are more unique to biology.

We were also surprised to find so many retrotransposon-associated sequences in the haustoria library compared to the hyphal library (Fig. 2). Although it's possible that the haustoria library was contaminated with some intergenic rust genomic DNA, it's unlikely that these contaminations would account for more than a

```
           1                                                          60
   UfRTP1  ...msnLrlL ftiislAAia raqlvgsdvv lvrtshqsMg kASVsycreM trqhh.....
   UaRTP1  .......... .......... .......... .......... .Aldshstyt ahqlR.....
   UsRTP1  mlfnpnLrvf ltiiilAAva raltpg..vv lvrthhetMg LASVsicrae kdihlghpnv
   PgRTP1  mlacitLclL lskmtiAgiS dahlpayrhi ks....fPMg dAkVtfceai p.........
   PgRTP2  ....mfLthL aIalatAAtS lsnvealsdr sesamr.PMs LktIhdaknM aLlER.....
   PgRTP3  ..mflrtvhf aIalatAAsS fskaeslspr sdtallsPMs LkSIhdddsM vLlER.....
   PgRTP4  ..mflrmihL aItlviAAsS lsnaeslypg sd.tmfsPMs LqSIhgadsM qLlER.....
   UaRTP2  ..mflrriqv sIavciflaS rn...ltngs nlipmtglld mkaVttddnh iLvEk.....
Consensus  ------L--L -I----AA-S ---------- -------PM- LASV-----M -L-ER-----

           61                                                        120
   UfRTP1  ..kreldqda npghrrhkse pegvkpSnht SapTpssppl ttvdltpakl ntaCYPGTFQ
   UaRTP1  ..kreldsea tlehkhhkqK ledvksSnht tnpThsspsl ttvdltpakl itsCYPGTFQ
   UsRTP1  rlrkrdmdkf anTpqhskle ledakdSnyl SkaTpsspsl ttvdltpakl kssCYlGTFQ
   PgRTP1  ....grtAah halrkrglda gppgpsSnsn SnpTlPtrse PLDlTPakCT SitCnPGTFQ
   PgRTP2  YqmGrpgAqm KaTVeIyafK Ss.atlSkRn ntgevPGiTk PLDmTPStCT SqvCY.....
   PgRTP3  YqmGrpgAqt KaTVeIysvK SagtslakRn SttemPGiTk PLDmTPStCT SqvCYPGsFe
   PgRTP4  YqiGgagAsv KpTVeVystK StgssftkRd StteiPGiTk PLDlTPSkCk SrvCYPGsYg
   UaRTP2  FilGdsnAeq KtyVqIlqeK Ss.stlSkRs iqqddkGrT. iiDlTPSkCT SesCYeGsFQ
Consensus  Y--G---A-- K-TV-I---K S-----S-R- S--T-PG-T- PLD-TPS-CT S--CYPGTFQ

           121                                                       180
   UfRTP1  APlLeDCEVV IrAQLYNSTG SLqvsPGdYV FVSYGTCATV FQNPq...yS kYsLqYNWAE
   UaRTP1  APsLeDCEVV IrAQLYNSTG SLqvsPGdYV FVSYGTCATV FQNPh...yS nYsLeYNWAE
   UsRTP1  APlLeDCEVV IrAQLYNSTG SLqasPGdYV FVSYGTCATV FQNPq...nS nYTiqYNWAE
   PgRTP1  kPdLqDCEkV IqAQLYNSTG SLsaaPGtYV FVSYGTCATV FQNPe...eS sYpLqYNWAE
   PgRTP2  .......... .......... .....PGqYV lVyaGTCvvV FQhPmgkgdn kFTLeYNWAs
   PgRTP3  gPkktDCdVV VdAQLYhSvG SLrskPGqYV lVysGTCvvV FQhPmgkgdn kFvLeYNWAn
   PgRTP4  gPkktDCdaI VdAQLYhSvG SLtvkPGnYV lVyaGTCvvV FQhPmgkgdn nlTfdYNWsk
   UaRTP2  APkkkDCdsI IeAQmYNSTG SLrafPGtvV YVwsGTCAaV FQNPlknan. .YTLdYNWAE
Consensus  AP-L-DCEVV I-AQLYNSTG SL---PG-YV FVSYGTCATV FQNP-----S -YTL-YNWAE

           181                                                       240
   UfRTP1  LGyvgGKlaG RCLLPEDhSm GGtaVfDtYL grTYpDViIS LQRFDDRDfi iPe.......
   UaRTP1  LGyvgGKlaG RCLLPEDhSm GGtaVfDtYL grTYpDViIS LQRFDDRDfi iPe.......
   UsRTP1  LGylgGKlaG RCLLPEDhSm GGtaVfDtYL ghTYpDViIS LQRFDDRDfi iPe.......
   PgRTP1  LGaqaGKIaG kCLLeEDkSm GGsCIfkkYL qYTFdDVmIS LQRFDDRDie mPk.......
   PgRTP2  LGkiilgIqk qCnkPdaLSv GGvCkiDhYL kYnFenVlIS LQRYvapenp qtpmksgqpg
   PgRTP3  LGkimvdIqk qCdkPEsLSi GGvCkiehYL kYTFenVlIS LQRYvapekp qkpitsgqpd
   PgRTP4  LGldiieIqk kCdkPEaLSi GGaCrikrYL qYnFdnIlIS LQRYvapenp qtpikpapps
   UaRTP2  LGlavlnIqe RCLrPkqLSi GGiCqiknYL nYTlegVgIS LQRisps... ..........
Consensus  LG---GKI-G RCLLPEDLS- GG-CV-D-YL -YTF-DV-IS LQRFDDRD-- -P--------

           241                                                  292
   UfRTP1  .......... .......... .......... .......... .......... ..
   UaRTP1  .......... .......... .......... .......... .......... ..
   UsRTP1  .......... .......... .......... .......... .......... ..
   PgRTP1  .......... .......... .......... .......... .......... ..
   PgRTP2  kpansdksan sdks.aksdk spkqdkpakq dksgkpqtpa pqh....... ..
   PgRTP3  kpekpddkpa npdd.kpanp dkpqdpaakp .......... .......... ..
   PgRTP4  ipgnsekpgn sgnsekpgns kpvnskpgqp ekpgqpekpk nppqsqdaap kp
   UaRTP2  .......... .......... .......... .......... .......... ..
Consensus  ---------- ---------- ---------- ---------- ---------- --
```

Fig. 5. ClustalW alignment of peptide sequences for rust transferred protein (RTP1)-like proteins displayed in Figure 4. Predicted signal peptides are underlined with a single line. Conserved amino acids in four or more sequences are in capitol letters and aa conserved in all eight sequences are underscored with a double line.

few percent of the clones since the RNA was DNase treated and an oligo d(T) primer was used for first strand cDNA synthesis. Bacterial contamination is not a likely suspect since none of the rRNA sequences matched to bacterial rRNAs. It's possible that our strain of *U. appendiculatus* had a virus that was active only in the haustorium. However, if the retroelements were from a virus, we might have expected to find other ESTs that matched better to viruses, but we did not. Moreover, there were many different retrotransposon-associated assemblies (61) rather than just a few as would be expected for a contaminating virus. Thus, we propose that the sequence divergence in the haustoria ESTs is a true reflection of *U. appendiculatus* gene expression in haustoria at 5 dpi and may indicate a highly specialized host-pathogen relationship.

Retrotransposon-associated sequences can make up a large portion of the ORFs in genomes (23) and a BLAST search of the available *P. pachyrhizi* and *P. graminis* genomic sequence indicates that rust genomes also include large numbers of retrotransposon-associated sequences (data not shown). Why haustoria should express many more retrotransposon-associated genes than the hyphae is unclear. None of the retrotransposon-associated sequences in our *U. appendiculatus* ESTs or the *P. graminis* ORFs included a predicted N-terminal signal peptide that might indicate they were secreted into the host to alter or modify host metabolism or gene expression. Since siRNAs are used by many eukaryotic organisms including fungi to suppress the detrimental expression of transposons and retroelements that are often found at very high copy number in eukaryotic genomes (14,20,27), it's possible that the plant might deregulate the natural fungal system for silencing these repetitive regions of the genome (29). If true, host interference in processing of siRNA in haustoria would have the effect of increasing retrotransposon-associated sequences in the haustoria relative to hyphae.

In addition to the retrotransposon category, there were some other notable differences in the types of ESTs detected between the hyphae and haustoria libraries. If the categories for unknown mRNAs and retrotransposon are excluded from the comparison of the functional clustering of ESTs in haustoria and hyphae, then there were significantly lower relative amounts of ESTs associated with amino acid metabolism, protein metabolism, and translation in haustoria than in hyphae (Fig. 2B). This was in contrast to sequences classified as having rolls in energy metabolism, which were more abundant in haustoria library relative to hyphal library (Fig. 2B). The increase in mRNAs associated with energy metabolism and reduced concentration of mRNAs associated with protein synthesis suggests that the primary function of haustorium at 5 dpi is not to synthesize protein, but rather to serve as an interface with the host for uptake and transport of nutrient back into the hyphae that will feed growth and development of hyphae and uredia. These observations support earlier studies indicating that the role of haustoria is to take up nutrient from the host and the conversion and synthesis of substrate for fungal growth (34,39–41).

ESTs classified into cellular regulation were only found in the hyphal ESTs and most of these were linked to cell differentiation (supplemental data UaAll.xls), which is consistent with the role of the infective hyphae in formation of other fungal bodies. Other categories disproportionally represented in the haustoria were cell structure (primarily associated with an abundance of actin ESTs) and vitamin metabolism (Fig. 2B). Vitamin metabolism is interesting because transcripts for thiamine (vitamin B1) biosynthesis were previously reported to be most abundant in haustoria (34). In our data set there were 16 ESTs from the hyphal library linked to thiamine biosynthesis but only two from the haustorial library. However, six ESTs were found that matched a protein involved in pantothenic acid (vitamin B5) biosynthesis, and these were exclusive to the haustorial library. Pantothenic acid is needed to synthesize acetyl-CoA, which is critical in the metabolism and synthesis of carbohydrates, proteins, and fats. Again, the role of this vitamin is consistent with the predicted role of haustoria in nutrient conversion and carbohydrate synthesis (40).

Proteins that are secreted from the fungi are of special interest because they could potentially induce or suppress host defense responses, alter host carbohydrate or lipid metabolism, or effect host transcription or translation. A putative secretome for *U. fabae* has already been determined using a yeast secretion test (24). We identified ORFs in the *U. appendiculatus* ESTs and *P. graminis* genome that included putative N-terminal signal peptides and then compared this subset of sequences to each other and the *U. fabae* secretome (Fig. 3D, E, and F). This comparison identified only a few similar sequences among all the data sets. This suggests that similar rust proteins with putative signal peptides can be found in divergent rust species but that a subset of potentially secreted proteins have significantly diverged to accommodate a host-specific response.

In addition, we examined all 64 of the *U. appendiculatus* ORFs with a predicted signal peptide for cellular targeting using WoLF PSORT and NucPred. Approximately one-third of the proteins could be targeted to the nucleus (data not shown). Since nuclear proteins are not normally synthesized on the endoplasmic reticulum, these proteins are likely to be secreted and transported into the host where they are then targeted to the nucleus. Thus, these proteins are prime candidates for having effector function.

As a part of the project to discover secreted proteins, we examined in more detail sequences with similarity to *UfRTP1*, which is transported from the haustoria into the host where it accumulates in the nucleus (21). Our analysis indicates that RTP is comprised of at least two clades in a small family of genes in rusts (Fig. 4). Based on biological evidence for secretion and targeting of UfRTP1 (21) and our analyses of the other RTP1-like sequences for the presence of signal peptides and cellular targeting (Fig. 4), we hypothesize that the RTP-like proteins in both clades are transported into the host where they are targeted to the nucleus. This implies that the UaRTP2 proteins are secreted across the haustorial extracellular membrane. However, all 59 of the UaRTP2 ESTs were found in the hyphal library and none in the haustorial library. Moreover, since haustoria serve as the interface with the host, in addition to RTP-like sequences, we expected to find a higher percentage of ORFs with signal peptides in the haustoria than hyphae. A partial explanation for fewer sequences with signal peptides in the haustoria is that the haustoria cDNAs were on average not as long as the hyphal cDNAs and therefore may not have included as many full-length sequences that contained sequence for signal peptides. However, this does not explain the large number of ESTs for *UaRTP2* in the hyphal library. To best interpret this result we must be cognitive of the experimental procedures used to obtain the hyphal and haustorial isolates. The rough homogenization procedure we used to release the hyphae would be expected to break the connection between the HMC and the mature haustorium inside the plant cell. In our protocol, the separated haustoria would then have passed through the 45-μm sieve and thereby not be collected in the hyphal preparation. The HMC and possibly some young budding haustoria that weren't broken off by the rough treatment would be collected with the hyphae in the 45-μm sieve. Based on this interpretation of how our isolation protocols work, the HMC and young budding haustoria in the hyphal preparation must then be responsible for transcribing the genes that will be secreted into the host through the haustorial membrane. This prediction fits with earlier studies on the role of haustoria (34,39–41) and our hypothesis that at 5 dpi mature haustoria have redirected energy allocation from transcription and protein synthesis to nutrient uptake from the host and its conversion and transport to hyphae that will initiate and feed the development of uredia.

The EST data presented here for *U. appendiculatus* adds unique sequence information for rusts to the public databases and additional insight into the biology of rust development. In addition,

the characterization of a subset of genes whose proteins might be secreted from the fungus highlights potential virulence and avirulence genes for further study.

ACKNOWLEDGMENTS

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the United States Department of Agriculture.

LITERATURE CITED

1. Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., and Miyano, S. 2002. Extensive feature detection of N-terminal protein sorting signals. Bioinformatics 18:298-305.
2. Brameier, M., Krings, A., and MacCallum, R. M. 2007. NucPred-predicting nuclear localization of proteins. Bioinformatics 23:1159-1160.
3. Catanzariti, A. M., Dodds, P. N., Lawrence, G. J., Ayliffe, M. A., and Ellis, J. G. 2006. Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. Plant Cell 18:243-256.
4. Chou, H. H., and Holmes, M. H. 2001. DNA sequence quality trimming and vector removal. Bioinformatics 17:1093-1104.
5. Cokol, M., Nair, R., and Rost, B. 2000. Finding nuclear localization signals. EMBO Rep. 1:411-415.
6. Cooper, B., Neelam, A., Campbell, K. B., Lee, J., Liu, G., Garrett, W. M., Scheffler, B., and Tucker, M. L. 2007. Protein accumulation in the germinating *Uromyces appendiculatus* uredospore. Mol. Plant-Microbe Interact. 20:857-866.
7. Dekhuijzen, H. M., Singh, H., and Staples, R. C. 1967. Some properties of hyphae isolated from bean leaves infected with the bean rust fungus. Contrib. Boyce Thomps. Inst. 23:367-372.
8. Eilam, T., Bushnell, W. R., Anikster, Y., and McLaughlin, D. J. 1992. Nuclear DNA content of basidiospores of selected rust fungi as estimated from fluorescence of propidium iodide-stained nuclei. Phytopathology 82:705-712.
9. Ellis, J. G., Dodds, P. N., and Lawrence, G. J. 2007. The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi. Curr. Opin. Microbiol. 10:326-331.
10. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2:953-971.
11. Ewing, B., and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186-194.
12. Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175-185.
13. French, R. C., Graham, C. L., Gale, A. W., and Long, R. K. 1977. Structural and exposure time requirements for chemical stimulation of germination of uredospores of *Uromyces phaseoli*. J. Agric. Food Chem. 25:84-88.
14. Grewal, S. I., and Elgin, S. C. 2007. Transcription and RNA interference in the formation of heterochromatin. Nature 447:399-406.
15. Hahn, M., and Mendgen, K. 1997. Characterization of in planta-induced rust genes isolated from a haustorium-specific cDNA library. Mol. Plant-Microbe Interact. 10:427-437.
16. Hahn, M., and Mendgen, K. 2001. Signal and nutrient exchange at biotrophic plant-fungus interfaces. Curr. Opin. Plant Biol. 4:322-327.
17. Heath, M. C. 1997. Signalling between pathogenic rust fungi and resistant or susceptible host plants. Ann. Bot. 80:713-720.
18. Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. 2007. WoLF PSORT: Protein localization predictor. Nucleic Acids Res. 35:W585-587.
19. Jakupovic, M., Heintz, M., Reichmann, P., Mendgen, K., and Hahn, M. 2006. Microarray analysis of expressed sequence tags from haustoria of the rust fungus *Uromyces fabae*. Fungal Genet. Biol. 43:8-19.
20. Kasschau, K. D., Fahlgren, N., Chapman, E. J., Sullivan, C. M., Cumbie, J. S., Givan, S. A., and Carrington, J. C. 2007. Genome-wide profiling and analysis of Arabidopsis siRNAs. Plos Biol. 5:e57.
21. Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., and Hahn, M. 2005. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. Mol. Plant-Microbe Interact. 18:1130-1139.
22. Koehler, S. M., Matters, G. L., Nath, P., Kemmerer, E. C., and Tucker, M. L. 1996. The gene promoter for a bean abscission cellulase is ethylene-induced in transgenic tomato and shows high sequence conservation with a soybean abscission cellulase. Plant Mol. Biol. 31:595-606.
23. Kumar, A., and Bennetzen, J. L. 1999. Plant retrotransposons. Annu. Rev. Genet. 33:479-532.
24. Link, T. I., and Voegele, R. T. 2008. Secreted proteins of *Uromyces fabae*: Similarities and stage specificity. Mol. Plant Pathol. 9:59-66.
25. Mendgen, K., and Hahn, M. 2002. Plant infection and the establishment of fungal biotrophy. Trends Plant Sci. 7:352-356.
26. Mount, D. W. 2004. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
27. Murata, T., Kadotani, N., Yamaguchi, M., Tosa, Y., Mayama, S., and Nakayashiki, H. 2007. siRNA-dependent and -independent post-transcriptional cosuppression of the LTR-retrotransposon MAGGY in the phytopathogenic fungus *Magnaporthe oryzae*. Nucleic Acids Res. 35:5987-5994.
28. Nair, R., and Rost, B. 2005. Mimicking cellular sorting improves prediction of subcellular localization. J. Mol. Biol. 348:85-100.
29. Nakayashiki, H. 2005. RNA silencing in fungi: Mechanisms and applications. FEBS Lett. 579:5950-5957.
30. Pastor-Corrales, M. A. 2003. Sources, genes for resistance, and pedigrees of 52 rust and mosaic resistant dry bean germplasm lines released by the USDA Beltsville Bean Project in collaboration with the Michigan, Nebraska and North Dakota Agricultural Experiment Station. Annu. Rep. Bean Improv. Coop. 46:235-240.
31. Patanjali, S. R., Parimoo, S., and Weissman, S. M. 1991. Construction of a uniform-abundance (normalized) cDNA library. Proc. Natl. Acad. Sci. USA 88:1943-1947.
32. Posada-Buitrago, M. L., and Frederick, R. D. 2005. Expressed sequence tag analysis of the soybean rust pathogen *Phakopsora pachyrhizi*. Fungal Genet. Biol. 42:949-962.
33. Simonoff, J. S. 2003. Analyzing Categorical Data. Springer-Verlag, New York.
34. Sohn, J., Voegele, R. T., Mendgen, K., and Hahn, M. 2000. High level activation of vitamin B1 biosynthesis genes in haustoria of the rust fungus *Uromyces fabae*. Mol. Plant-Microbe Interact. 13:629-636.
35. Staples, R. C. 2000. Research on the rust fungi during the twentieth century. Annu. Rev. Phytopathol. 38:49-69.
36. Tucker, M. L., Burke, A., Murphy, C. A., Thai, V. K., and Ehrenfried, M. L. 2007. Gene expression profiles for cell wall-modifying proteins associated with soybean cyst nematode infection, petiole abscission, root tips, flowers, apical buds, and leaves. J. Exp. Bot. 58:3395-3406.
37. van der Merwe, M., Ericson, L., Walker, J., Thrall, P. H., and Burdon, J. J. 2007. Evolutionary relationships among species of *Puccinia* and *Uromyces* (Pucciniaceae, Uredinales) inferred from partial protein coding gene phylogenies. Mycol. Res. 111:163-175.
38. Voegele, R. T. 2006. *Uromyces fabae*: Development, metabolism, and interactions with its host *Vicia faba*. FEMS Microbiol. Lett. 259:165-173.
39. Voegele, R. T., Hahn, M., Lohaus, G., Link, T., Heiser, I., and Mendgen, K. 2005. Possible roles for mannitol and mannitol dehydrogenase in the biotrophic plant pathogen *Uromyces fabae*. Plant Physiol. 137:190-198.
40. Voegele, R. T., and Mendgen, K. 2003. Rust haustoria: Nutrient uptake and beyond. New Phytol. 159:93-100.
41. Voegele, R. T., Struck, C., Hahn, M., and Mendgen, K. 2001. The role of haustoria in sugar supply during infection of broad bean by the rust fungus *Uromyces fabae*. Proc. Natl. Acad. Sci. USA 98:8133-8138.