RESEARCH ARTICLE

Proteome of monoclonal antibody-purified haustoria from *Puccinia triticina* Race-1

Christof Rampitsch[1], *Aslıhan Günel*[2], *Eva Beimcik*[1] *and Wayne Mauthe*[1]

[1] Agriculture and Agrifood Canada, Cereal Research Centre, Morden MB, Canada
[2] Department of Chemistry, Ahi Evran University, Kırşehir, Turkey

Puccinia triticina causes leaf rust, a disease that causes annual yield losses in wheat. It is an obligate parasite that invades the host leaf and forms intracellular structures called haustoria, which obtain nutrients and suppress host immunity using secreted proteins called effectors. Since effector proteins act at the frontier between plant and pathogen and help determine the outcome of the interaction, it is critical to understand their functions. Here, we used a direct proteomics approach to identify effector candidates from *P. triticina* Race 1 haustoria isolated with a specific monoclonal antibody. Haustoria were >95% pure and free of host contaminants. Using high resolution MS we have identified 1192 haustoria proteins. These were quantified using normalized spectral counts and spanned a dynamic range of three orders of magnitude, with unknown proteins and metabolic enzymes as the most highly represented. The dataset contained 140 candidate effector proteins, based on the presence of a signal peptide and the absence of a known function for the protein. Some of these candidates were significantly enriched with cysteine, with up to 13 residues per protein and up to 6.8% cysteine in composition.

Received: May 28, 2014
Revised: November 17, 2014
Accepted: December 19, 2014

Keywords:
Basidiomycete / Brown rust / Plant proteomics / *Puccinia recondita* f. sp. *tritici*

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

The rust fungus *Puccinia triticina* Eriks. is an obligate parasite that causes wheat leaf rust (WLR), leading to significant annual yield losses to the wheat harvest worldwide [1, 2]. WLR is wide spread and is managed through the application of fungicides and by deploying host resistance (*R*) genes in wheat cultivars. *R* genes confer resistance to one or more races (pathotypes) of rust and at present more than 60 such genes have been identified in wheat [3]. WLR races are closely related and are defined by their ability to overcome *R* genes. This trait is encoded by avirulence (*avr*) genes in the pathogen, and these can interact with *R* genes in a manner originally described by Flor [4] in flax rust. Flor's model is called the gene-for-gene interaction and it has proven itself robust, although numerous refinements have been made, notably by Jones and Dangl who proposed four phases of the defense response [5]. In the first, the basal response, the host recognizes pathogen-associated molecular patterns (PAMPs) and activates PAMP-triggered immunity by producing reactive oxygen species, defense proteins, and depositing callose. The pathogen must overcome this, and attempts to do so through secreting effector proteins. In the case of rust pathogens, many of these are produced and secreted by haustoria [6]. The host then activates a second level of immunity, called effector-triggered immunity, and this level includes the *R-avr* gene-for-gene interaction. Ultimately immunity is the result of a compatible *avr-R* gene interaction and is manifested as a localized, hypersensitive cell death with no further fungal spread, and no sporulation. If a compatible *R* gene product is not encountered, the fungus completes its life cycle and

Correspondence: Dr. Christof Rampitsch, Cereal Research Centre, Agriculture and Agrifood Canada, 101 Route 100, Morden MB R6M 1Y5, Canada
E-mail: chris.rampitsch@agr.gc.ca
Fax: +1204-8227602

Abbreviations: dNSAF, distributed normalized abundance factor; **PAMP**, pathogen-associated molecular pattern; **WLR**, wheat leaf rust

disease results. Rust races are genetically defined based on *avr* genes and up to 100 or more individual leaf rust races are routinely identified in countries that conduct surveys [1] with new races emerging regularly, rendering much of the known *R* gene arsenal as no longer practically useful. New races emerge either through mutation, sexual reproduction or they are introduced from other areas. The rust life cycle is complex with both sexual and asexual modes of reproduction, although the sexual stage—an important source of genetic diversity—can be eliminated by removing the alternate host (meadow rue, *Thalictrum speciosissimum*). In areas where this has been done the emergence of new races is rarer [2].

Urediniospores, generated asexually, are the main source of WLR. They germinate on the surface of the leaf, and germ-tubes enter the leaf through stomata [7, 8]. Once inside the apoplastic space of the leaf, hyphae spread, and form haustorial mother cells that invaginate host cells and form haustoria [7]. Haustoria are intracellular structures that secrete proteins and metabolites into the host cells to suppress their defenses and to reprogramme the host into supporting fungal growth [9]. Among the secreted proteins are effectors [10, 11] that include the *avr* gene products.

The role of fungal effector proteins was the subject of a recent review [12]. In haustoria-forming fungi, good progress has been made with flax rust (*Melampsora lini*) with four effector proteins cloned [13]. These were all small haustorially expressed proteins (<130 amino acids in length), rich in cysteine that were able to induce a hypersensitive response in flax lines carrying their cognate *R* gene, even in the absence of *M. lini* [13]. None of the cloned proteins had any sequence similarity to other known proteins. Two effector protein genes, *Avra10* and *Avrk1*, have been cloned from *Blumeria graminis* f.sp. *hordei* (powdery mildew of barley) [14].

Given the central role of haustoria in the infection process, several attempts have been made to purify them to a degree suitable for proteomic and genetic analyses. Enrichment has relied mainly on concavalin A sepharose affinity chromatography [15, 16] and density gradient centrifugation [17, 18]. Although both methods enrich haustoria to a degree, they suffer from the drawback that the resulting enrichments are contaminated with both plant material and hyphae, and it is therefore difficult to identify haustoria proteins unequivocally, although a subtractive approach can be used [16]. Here, we present findings on the haustoria proteome using haustoria that were purified to near homogeneity by immunoprecipitation with monoclonal antibodies prepared specifically to haustoria of *P. triticina*. We have used a high-resolution, high mass accuracy mass spectrometer to analyze tryptic peptides derived from these haustoria and have tentatively identified 1192 proteins with two or more peptides matched ($p < 0.05$). Using bioinformatics-based algorithms and characteristics of known effectors, 140 effector protein candidates have been tentatively identified.

2 Materials and methods

2.1 Biological materials

Wheat (*Triticum aestivum*) cultivar 'Thatcher' (RL6101) was inoculated with *Puccinia triticina* Eriks. urediniospores of virulence phenotype BBBD (Race-1, isolate 90–1) 9 days after germination exactly as described in Rampitsch et al. [19]. Inoculated plants were grown in a growth cabinet until harvest, under 16 h daylight, 20°C. Plants were harvested after 7 days by cutting primary leaves directly onto ice. At this stage flecks were visible, but spores had not yet formed.

2.2 Preparation of monoclonal antibodies

Enriched haustoria to be used as antigen were prepared exactly as described by Song et al. [18] following a modified procedure of Tiburzy et al. [20]. Briefly, leaves were homogenized in a blender with extraction buffer (0.1 M K-phosphate pH 6.7, 0.1 M sucrose, 1.5 mM $MgSO_4$, 1.0 mM phenylmethylsulfonyl fluoride), filtered successively through 100 and 25 μm pore size nylon mesh and then enriched by sucrose density gradient centrifugation, using the gradient-steps detailed in Song et al. [18]. The final fraction, enriched in haustoria but also contaminated with hyphae, chloroplasts, and other debris, was used to immunize three Balb/c mice for the production of monoclonal antibodies. The mouse with the strongest response, as determined visually from a dot blot using serum, was sacrificed for a fusion that was performed using the standard polyethylene glycol procedure. After the fusion, hybridomas were screened initially by ELISA using the tissue culture supernatant directly as primary antibody solution. For ELISA, the antigen was either a plant extract or the injected antigen mixture used to immunize the mice, and was bound directly to 96-well plates (Maxisorp: Nunc, ThermoFisher Scientific, NJ). The plant extract was obtained in the same way as haustoria, but using mock-inoculated plants. Only hybridomas reacting to the injected antigen mixture alone, and not to the plant extract, were retained. All others were discarded. Retained hybridomas were screened further by microscopy as follows. Tissue culture supernatant was used to decorate haustoria in the enriched haustoria samples by mixing the two in equal volumes for 30 min at room temperature. Haustoria were then collected by centrifugation at 13 000 *g*, 5 min and washed three times with standard PBS. A fluorescent-labeled anti-mouse conjugate (donkey anti mouse-whole antibody-FITC-labeled: Jackson Immuno Research, West Grove, PA) was then incubated with the decorated haustoria for 30 min at room temperature, and washed as before. The screen was completed by observing the haustoria under a fluorescence microscope and overlaying phase contrast bright-field images with fluorescent images to identify hybridomas producing anti-haustoria-specific antibody. Such clones were retained and single-cell cloned twice by

limiting dilution. They were subsequently grown to confluence in 150 mL DMEM with 7% v/v FCS (Sigma Aldrich, St. Louis MO) in T-flasks, at 37°C, 10% v/v CO_2. The supernatant tissue culture solution from these was stored at 4°C in sterile test-tubes until needed, for up to 3 months without loss of antibody. All successful hybridomas were tested for serotype using a lateral-flow assay (Roche Applied Science, Laval PQ).

2.3 Immunoprecipitation of haustoria

Infected seedlings (typically 40 g) were harvested and crude haustoria prepared by sucrose density gradient centrifugation as described in Section 2.2. The haustoria resulting from this were washed in PBS and resuspended finally in 1 mL PBS. These were then separated into 250 μL aliquots and precipitated using magnetic beads as follows. A 50 μL suspension of magnetic beads charged with protein A was prepared separately for each 250 μL haustoria aliquot: following the instruction of the manufacturer (Dynabeads, Life Technologies, Burlington, ON), the beads were decorated with donkey anti mouse-IgM-μ-chain (Jackson Immuno Research) and the anti-haustoria monoclonal IgM in tissue culture solution, and washed in PBS before use. Each 250 μL haustoria aliquot was then mixed with 50 μL decorated bead suspension and allowed to cross-react for 30 min at room temperature with end-over-end mixing. The beads were then captured with a rare-earth magnet and washed three times with PBS. Haustoria were released from the beads using 0.1 M glycine, pH 2.5. The quality of each preparation was assessed by microscopy by staining a 1 μL sampling with calcofluor white and viewing it under a fluorescence microscope as described previously [18]. Haustoria were lyophilized and stored at –80°C until required.

2.4 Protein extraction

Proteins were extracted from purified lyophilysed haustoria as follows. Haustoria (typically 5 mg dry weight) were first homogenized using a glass Potter–Elvehjem homogenizer in liquid nitrogen. Upon thawing, 500 μL of 50 mM Tris.HCl, 20 mM DTT, pH 7.6 were added to the powder, which was homogenized further. Proteins were then extracted with 500 μL of Tris-buffered phenol, pH 7.6, heated to 75°C for 30 min with sonication for two 1 min bursts, and occasional vortexing. The extraction was centrifuged to separate phases, and the tube punctured to collect the organic phase. The aqueous phase was reextracted with 500 μL phenol and then with 500 μL chloroform. Proteins were precipitated from the pooled organic phases with four volumes of ice-cold methanol. The resulting pellet was washed with methanol, dried, and resuspended in ammonium bicarbonate buffer for trypsin digestion, which proceeded as recommended by the manufacturer (Invitrogen: Life Technologies).

2.5 MS

Samples of tryptic peptides in 0.1% v/v formic acid, 2% v/v aqueous acetonitrile were analyzed in a Triple-TOF mass spectrometer (5600: AB Sciex, Concord, ON) under a fee-for-service contract at the Manitoba Centre for Proteomics and Systems Biology (Winnipeg MB, Canada). Briefly, LC-MS runs used a standard 2–40% v/v acetonitrile gradient over 3 h while acquiring MS^2 data using a top-5 method. Two identical injections of 3 μg peptides per biological replicate were delivered to the LC-MS system. Output files were converted to Mascot Generic Format and queried against a *Puccinia* database using Mascot Server (v2.4, MatrixScience, London, UK). The database consisted of 52 198 sequences comprising the sequenced genomes of *P. triticina* (Race-1), and the related cereal rusts *P. graminis* f. sp. *tritici* and *P. striiformis* and was downloaded from the Broad Institute's website (www.broadinstitute.org) in September 2013. The spectra were also queried, simultaneously, against the *Aegilops tauschii* genomic sequence (29 796 sequences). This plant is the D-genome progenitor of hexaploid wheat and is a relevant substitute for *T. aestivum* in homology based searches [21]. Settings for Mascot were, fixed modifications CAM (C), + 57 Da; variable modifications DEAM (NQ) –1 Da; OX (M) + 64 Da; the enzyme was trypsin with one missed cleavage permitted; precursor ion mass tolerance ± 25 ppm; product ions ± 0.5 Da; decoy database searching (reversed sequence) was enabled. In addition a database of common mammalian contaminants (258 sequences) was included in each search.

2.6 Data analyses

Mascot DAT files were loaded into Scaffold (v4.0: Proteomesoftware, Portland, OR). Scaffold was used to validate MS/MS-based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 99.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 90.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm [22]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. An X!Tandem search was also enabled using the default values set by Scaffold, and finally, the distributed normalized abundance factor (dNSAF) was calculated for each protein. This quantitative value is based on spectral counts and is used to estimate the relative abundance of proteins within a sample [23]. The dynamic range was estimated as being equal to 10^x of the difference between the highest and lowest dNSAF values.

GO annotations were generated using the BLAST2GO tool [24] (www.blast2go.org) by querying with BLASTp against the nonredundant database at NCBI, and annotating at $E < 10^{-6}$, cutoff 55, and GO weight 5. Annotation based on the NCBI description lines from the BLASTp hit with the smallest *E*-value (NCBI nonredundant protein database, November 2013) was also added. Further analyses were performed for potential signal-peptide signatures using SignalP (v4.1) [25] using the default settings for eukaryotes. Where known secretory signals were identified, proteins were filtered further to eliminate those with a known mitochondrial localization signal, using TargetP (v1.1) [26] using the default settings for nonplant organisms. This limits the set of potential secreted proteins, although missing proteins with noncanonical secretion signals cannot be ruled out.

3 Results and discussion

3.1 Monoclonal antibodies bind specifically to haustoria

Eleven stable hybridoma clones producing antibody were detected by ELISA against plant and/or fungal extracts. Of these, two monoclonal antibodies were shown by immunofluorescence microscopy to react specifically to haustoria (Fig. 1) without cross-reaction to other debris present in the crude preparation. For the remainder of this project a single monoclonal antibody was used, all others have been either discarded or stored under liquid nitrogen. Both of the reacting monoclonal antibodies, indeed all of the clones tested, were found to be of the IgM serotype, however, since their intended use was immunoprecipitation, this was not seen as a major drawback. IgM antibodies are generally less stable and have a lower binding affinity than other serotypes, and IgG serotypes are preferred. However, the IgM used produced high quality immunoprecipitates of haustoria as seen in Fig. 2. These preparations contained no chloroplasts that are typical of density gradient-enriched material as they autofluoresce red under the same excitation illumination as is required for calcofluor white staining [18]. Furthermore only very few hyphal contaminants, which are visibly septate and do not have the typical cigar shape of haustoria, were observed. The yield from 40 g of infected leaf material (fresh weight) was typically 5 mg of lyophilized haustoria.

The monoclonal antibody also cross-reacted to haustoria of *P. striiformis* (Supporting Information Fig. 1), *P. hieracii*, and *P. lagenophor* (unpublished observations) as determined by immuno-fluorescence microscopy indicating that the epitope is well conserved, although the epitope itself could not be identified. Since *Puccinia* spp. have a narrow host range and growth conditions were carefully controlled, cross-contamination of our preparations with other rusts can be ruled out.

3.2 Haustoria proteome contains more than 1000 proteins

In total, 1192 proteins were identified from *P. triticina* race 1 haustoria with 2 or more peptides matched ($p > 0.95$) and 1694 with one or more peptide matched using spectra from a high-resolution mass spectrometer (see Supporting Information Table 1 for a complete list). The false discovery rate (using reversed sequences) was 0.9%, there was 85% overlap in protein identifications between the two biological replicates and no significant matches were returned from the mammalian contaminants database. This contrasts favorably with protein numbers reported previously: 206 proteins were reported for *P. triticina* [18]; and 45 and 204 for *Blumeria graminis* f. sp. *hordei* (powdery mildew) [17, 27]. The increased numbers reported here are likely primarily due to improved purity of immunoprecipitated haustoria, but also to the availability of a full genomic sequence for *P. triticina* Race 1, and to improvements in MS instrumentation. In addition, the proteins previously reported by Song et al. [18] for leaf rust haustoria were obtained through GeLC-MS, whereas here we used a gel-free proteomics approach. The proteome reported here includes 142 of the 264 proteins reported by Song et al. (54%) [18]. Improved haustoria purity is the likeliest reason why the overlap was not greater, with the present study reporting fewer nonhaustorial proteins. The total number of predicted genes for *P. triticina* Race 1 is 14 878 thus the proteome reported here covers approximately 10% of it, but is likely still incomplete.

Forty-eight credible proteins were returned from the *Aegilops tauschii* database (Supporting Information Table 2), indicating that the haustoria preparations were still contaminated with some wheat proteins, albeit at a low level. All databases were queried simultaneously to reduce the number of returns expected to result from homology between evolutionarily conserved proteins present in both the *Puccinia* and *Ae. tauschii* databases. However, photosynthesis proteins and enzymes clearly originated from contaminating plant tissue.

We found haustoria to be a recalcitrant tissue, extractable with SDS but resistant to extraction with acetone and TCA, and difficult to break by grinding. Little protein was liberated and haustoria appeared histologically intact after extraction with either acetone or acetone/TCA, but not after extraction with heated phenol (unpublished observations). To estimate the abundance of proteins within each biological replicate, the dNSAF was calculated for each protein [23]. This value is based on unique spectral counts normalized through protein length and can be calculated by Scaffold 4.0. The highest and lowest dNSAF values were 10^{-1} and 10^{-4} in both biological replicates, indicating that the haustoria proteome reported here covers at least three orders of magnitude. Abundant proteins included ribosomal proteins, prohibitin, cytochrome c oxidase, ATP synthase components, proton pump components, and metabolic enzymes (see Section 3.3) and proteins of unknown function. This is consistent with the function of haustoria, which is to obtain nutrients from the plant

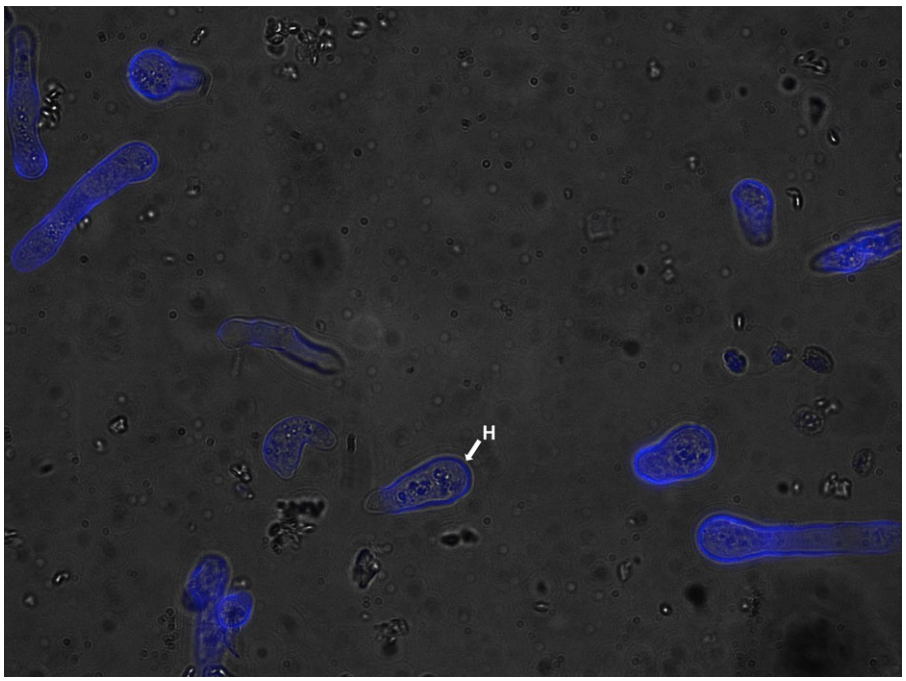

Figure 1. The monoclonal antibody binds specifically to haustoria. Bright field micrograph of a crude preparation of haustoria, contaminated with fungal and plant debris, overlayed with a fluorescence image. Haustoria (H) were decorated with monoclonal and fluorescence-labeled conjugate as described in the text, and fluoresce blue.

host, to provide energy and building blocks for the growing fungal tissue and finally to provide proteins critical for overcoming plant defense responses (Section 3.4) [9]. We used Blast2Go to group proteins by ontology. This annotation method assigned putative molecular functions, cellular locations, and biological processes to 55% of the haustoria proteome (Supporting Information Fig. 2). Ontology groupings also showed that the biggest grouping was for enzymes involved in metabolic processes, as can be expected for haustoria.

3.3 Haustoria provide energy and nutrients to the pathogen

A key role of haustoria is to provide nutrients for the invading host, which has no other means to obtain them [9]. Hexose

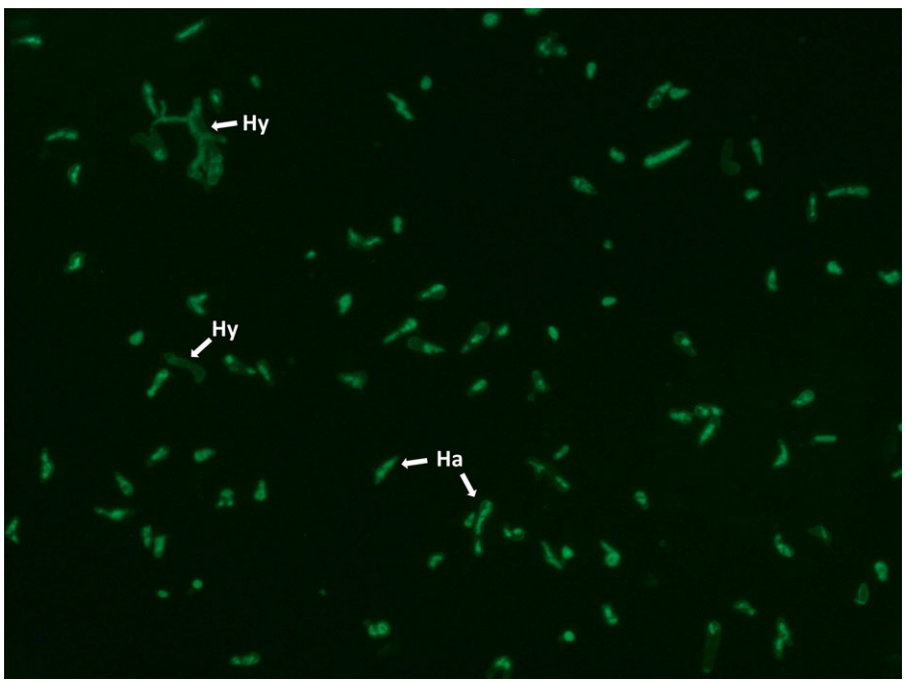

Figure 2. Immunoprecipitated haustoria (Ha) stained with calcofluor white. Chloroplasts, which would autofluoresce red under these conditions, are absent. Hyphae (Hy) are visibly septate and are a minor contaminant.

sugars in particular are taken up exclusively by haustoria via proton gradients formed by H^+ATPases [9], and several of these putative transporter subunits were observed. Furthermore, all of the enzymes of the TCA cycle, all but one enzyme of glycolysis (hexokinase was not detected), and three enzymes of the pentose phosphate pathway (G6P dehydrogenase, 6-P-glucuronate dehydrogenase, and transaldolase) were seen in the proteome. In addition 4 ATP synthase subunits, numerous NADH reductase subunits and cytochrome c oxidase subunits were detected, indicating that the haustoria were likely metabolically active, taking up sugars from the host cell and oxidizing them as an energy source to drive fungal growth, and ultimately sporulation. Since mannitol has been reported to be a major carbon-storage sugar for rust fungi and spores [9, 28], it was surprising that neither NADP-dependent mannitol dehydrogenase (PTTG 06111), nor its homolog, hypothetical protein PTTG 06246, were observed. The former is homologous to Mad1 and was localized to haustoria in *Uromyces fabae* by Vögele et al. [28]. They proposed that Mad1 converts fructose to mannitol for use both as a storage compound and a ROS scavenger in *U. fabae*.

One of the roles of the pentose phosphate pathway in haustoria is to provide reducing power (NADPH) for fatty acid biosynthesis [9], and indeed, acetyl-CoA carboxylase, which catalyzes the first committed step of fatty acid biosynthesis was also detected. In addition to energy, the growing fungus also requires amino acids and for this, aspartate amino transferase and glutamate dehydrogenase—both key enzymes of nitrogen/amino acid metabolism—were present. In addition to feeding the growing fungal biomass, haustoria are also thought to secrete proteins that alter host metabolism to favor fungal growth [9]. There is little information on how this metabolic reengineering is accomplished at the biochemical level within the host cell. The present study is restricted to the haustoria proteome, however, four sec protein subunits, including highly conserved sec 23 and 24 were detected. These have known roles in secreting folded proteins into the host cytoplasm [29], although the identity of the exported proteins themselves is not known.

The regulated generation of ROS is a hallmark of pathogen invasion [30], and enzymes such as ascorbate peroxidases, catalases, and thioredoxins are frequently reported in plant-pathogen proteomes as some of the most abundant changes in proteomes [31]. Only two enzymes involved in reactive oxygen detoxification were detected, suggesting that the inside of haustoria remain relatively free of reactive oxygen species, including H_2O_2 that can diffuse across at least some membranes [e.g. 32] and is associated with pathogen defense. Haustoria are protected by multiple layers [7] and their permeability to H_2O_2 is not known. These enzymes were a mitochondrial superoxide dismutase (PTTG 09275), likely involved in limiting O_2^- produced during aerobic respiration, a peroxiredoxin (PTTG 08121), and a heme-binding peroxidase (PTTG 07367); the last is homologous to class-I peroxidases and may also be mitochondrial.

Other proteins observed belong to the translation machinery, including elongation factors and ribosomal subunits; proteasome subunits; as well as enzymes of nucleic acid and lipid biosynthesis, again consistent with a metabolically active organ. Heat shock proteins and putative signaling proteins were also detected. Transcriptome data show similar findings in some regards, but there are also differences. Perhaps the most relevant published transcriptome for comparison is that of Garnica et al. [33] who used ConA affinity chromatography to enrich haustoria from *P. striiformis* and then performed a 454-based transcriptome analysis. They reported that the most abundant transcripts were from Energy Metabolism, Translation, Vitamin Metabolic Process, and Carbohydrate Metabolic Process however, our GO analysis did not return many proteins from the last three categories (Supporting Information Fig. 2), indicating that there are differences in the relative abundance of haustorial transcripts versus proteins in these closely related rust species.

3.4 The haustoria proteome revealed 140 effector protein candidates

A large group comprising 45% of the haustoria proteome consists of proteins of unknown function. These are very likely to include effector proteins—secreted proteins with a function in the plant-pathogen interaction, because this area of research is poorly investigated. Identifying these proteins and understanding their functions will be critical to understanding the rust-wheat interaction. Since effector proteins act in the host cell cytoplasm they must first be secreted and the synthesis and secretion of such proteins appears to be one of the primary functions of haustoria [9, 10]. This is generally assumed to be accomplished through the action of N-terminal signal or transit peptides, which are cleaved off to create mature effector proteins in the host cytoplasm. Effector protein candidates can be detected using a bioinformatics approach, however, only known sequences can be used and since no universal signalling mechanism has been found in rusts, it is likely that noncanonical sequences are also used by the pathogen. Saunders et al. [34] published an extensive bioinformatics-based effector protein screen in *P. graminis* f.sp. *tritici* (wheat stem rust) and *Melampsora larici-populina,* (poplar rust) and listed eight universal features of known fungal effector proteins. Some of these were features of genome organization, as this was a genomics-based screen, and these are not applicable to proteomics data analysis. Aside from the signal sequence most effector proteins are thought to be shorter than 300 residues in length and rich in cysteine [35]. To produce a list of candidate effector proteins from the haustoria proteome, all 1744 proteins identified were used (i.e. also those identified by a single peptide) since short proteins may be adequately represented by a single peptide. SignalP was first used to identify known signal peptides present in classically secreted proteins. TargetP was then used to refine this list by identifying proteins with transmembrane domains and

Table 1. Candidate effector proteins, (see Supporting Information Table 3 for the complete list of 140 candidates).

Name	Length	#Cys	%Cys	Uniq Pep	Sequence[a)]	Mascot[b)]	X! [c)]	dNSAF[d)]	%Cov[e)]
PTTG_03405T0	189	12	6.3	1	(K)AAASGNYSTIPTEAACK(G)	95	9.29	0.0154	9
PTTG_11654T0	177	12	6.8	8	(K)VSP[oxM]QCGNIFLPLSPVDIAEMQSK(G)	67	8.36	0.0954	41
PTTG_11707T0	121	11	9.1	1	(K)CNLSVTGESST[oxM]PCPK(L)	94	8.48	0.01885	13
PTTG_27172T0	251	11	2.4	5	(R)THTIALPYSTGIQTR(S)	86	10.27	0.027095	22
PTTG_06324T0	296	10	3.4	5	(K)IPDTIATSTGPSSSNLPGDGSK(N)	108	10.51	0.02605	23
PTTG_02417T0	258	9	3.5	4	(R)SFPLSQDGVNCQFQSFAQGR(Q)	51	6.52	0.01244	33
PTTG_03653T0	276	9	3.3	16	(K)DLIWNFDNEQGTNPK(N)	111	10.64	0.583	80
PTTG_06577T0	208	9	4.3	4	(K)LIASMTWDFNNLNFDNPK(N)	82	75.8	0.0368	48
PTTG_09595T0	185	9	4.9	3	(R)NSQQLSPLSGPR(L)	66	3.96	0.012335	15
PTTG_03497T0	209	8	3.8	9	(K)CTSPTCYPGTFQPPALEDCQK(V)	86	5.11	0.1775	24
PTTG_03539T0	272	8	2.9	2	(R)VTQGPFANLEITLGPGFQNVR(R)	57	4.52	0.0079	12
PTTG_09007T0	216	8	3.7	5	(K)EHTSPVQAFDMPSQSNLAK(R)	73	7.48	0.0427	25
PTTG_25421T0	152	8	5.3	1	(R)APGDWGPACQCAPR(Y)	103	13.7	0.0219	9
PTTG_09932T0	211	7	3.3	4	(K)HTFEALNPTPAER(D)	46	4.6	0.0185	19
PTTG_27021T0	118	7	5.9	2	(R)ATALGEYDVATEYGL[oxM]R(A)	41	3.92	0.011	31
PTTG_04406T0	192	6	3.1	5	(K)SLPAGQKPTISVELSK(Q)	116	9.59	0.1525	39
PTTG_12548T0	119	6	5	1	(K)VGEKPTIDTACCATGTTPSGGAPDPK(G)	42	4.22	0.00822	22
PTTG_25167T0	229	6	2.6	3	(R)LPALL[oxM]GGPVGFDFLSR(E)	137	10.33	0.01025	14
PTTG_25283T0	194	6	3.1	2	(R)VVDGCSLADQKPLTIEEGCSTIYVTR(L)	100	5.34	0.0123	21
PTTG_05834T0	223	5	2.2	2	(R)LALY[deamN]GFFAK(A)	33	1.38	0.01165	9

a) These proteins were chosen because they (i) possess a known secretion signal, (ii) are rich in cysteine residues, and (iii) are shorter than 300 amino acids in length.
b) Sequence of the highest scoring peptide.
c) Mascot ions score $-10\log(p)$, where p is the probability that the observed match is a random event. Individual ions scores > 32 indicate identity or extensive homology ($p < 0.05$).
d) X!Tandem score.
e) Distributed normalized abundance factor is defined in Section 3.1 [23].
f) % Sequence coverage of all observed unique peptides.

proteins targeted to mitochondria. These are less likely to be effectors and were removed. Although using these coarse filters may miss proteins with important roles in pathogenicity, they nevertheless present a list of 140 proteins containing predicted secretion signals, with 63 of those at 300 residues in length or shorter. These 140 proteins are listed in Supporting Information Table 3. All these candidate effector proteins are of unknown function, with only two described further than "hypothetical protein" (PTTG 07281, "glucose regulated," and PTTG 25205, "arf/Sar family"). The list includes five of the six wheat leaf rust potential candidate effector proteins described by Song et al. [18]. These candidates are enriched in cysteine residues, with PTTG 00248, PTTG 04843, and PTTG 08334 containing 13 residues each. The 177 residue protein PTTG 11654 was 6.8% cysteine (12 residues). Cysteine residues are rare, occurring approximately 84 000 times in the entire predicted genome of *P. triticina* (calculated from the Broad Institute's sequences), or at a frequency of <1%. In the 140 candidates the average frequency of C is 4.66 residues per protein. It has been suggested that these cysteines protect secreted proteins from protease degradation [12] but it is also possible that they participate in redox signaling, since there will likely be an increased level of ROS in infected cells [36], and cysteine is a prime target for oxidation. However, a rich cysteine content is not a guarantee of effector function. Table 1 presents a short-list of candidate effector proteins.

To confirm their biological roles unequivocally as effector proteins, and to demonstrate their roles in pathogenicity or virulence is a complex process requiring a low-to-medium throughput approach where each protein is examined individually [37]. It is therefore desirable first to shorten the list to candidates whose role is supported through further direct biological evidence, or indirect bioinformatics evidence. It would be most desirable either to quantify the expression of candidate proteins through time, or alternatively, to seek differences between haustorial proteomes of rust races differing in virulence genetics. We intend to do the latter by comparing the haustoria proteomes of three sequenced rust races (races 1, 9, and 161) in a semiquantitative proteomics experiment.

Calculating dNSAF values reveals an estimate of the dynamic range and the relative abundance of proteins in the sample [23]: a more robust quantitation will require more biological replicates or other quantitation techniques. Not all of the candidate effector proteins were present at low abundance. In both biological replicates the most abundant proteins of the haustoria proteome (PTTG 06270 and PTTG 03653) were predicted to be secreted, and these had dNSAF scores of 0.686 and 0.73, respectively. Ten other putative secreted proteins had dNSAF scores in the 10^{-1} range, indicating that some secreted proteins are abundant. The least abundant were in the 10^{-4} range (four proteins, see

Supporting Information Table 3); the remainder were: 10^{-3} (63 proteins) and 10^{-2} (58 proteins). In addition, PTTG0 3653 had 9 cys residues. It remains to be demonstrated whether these candidates have effector roles.

None of the nine *B. graminis* f.sp. *hordei* haustoria-specific proteins reported by Bindschedler et al. [27], and none of the three predicted effector candidates highlighted in *P. striiformis* by Cantu et al. [35] had homolog in the present dataset ($E < 10^{-3}$ by pBLAST). A published *P. striiformis* haustoria transcriptome obtained through 454 pyrosequencing revealed 437 secreted proteins, which were considered effector candidates [33]. Remarkably, only a single one of these (designated as Hcontig9953, [33]) aligned to the coding sequence of any proteins presented in our dataset (PTTG 04652, Supporting Information Fig. 3). This protein was not considered as an effector candidate in our work, because it is >100 kDa in size and the algorithm we used did not detect a secretion signal. Such a low incidence of homology among effector protein candidates, even between species of the *Puccinia* genus that colonize wheat, suggests that these proteins are structurally very diverse, or that proteome coverage is still too shallow to detect more homologs.

4 Concluding remarks

A comprehensive proteome of leaf rust haustoria, comprising 1192 proteins, is presented. Focus has so far mainly been on the transcriptome [16, 33, 35, 38] and on using bioinformatics to identify effector proteins based on their known key features [34]. Here proteome data are presented from highly purified haustoria. The advantages of investigating the proteome rather than the transcriptome are that protein abundances are measured directly—eliminating the need to correlate transcript and protein levels—and that most PTMs can be identified by MS. This research will facilitate future work on effector proteins as it reports proteins directly obtained from haustoria. It will also enable comparative proteomics between haustoria from different races of leaf rust, or between rust species, because the antibodies react to haustoria from a wide range of species. Using haustoria from races or species for which complete genome sequences are available will enable researchers to pinpoint subtle proteome differences if custom, race-specific databases are used. Research on rust is complicated by the fact that it is an obligate parasite, and obtaining comparative samples (e.g. different developmental stages) from a single race is impossible. Obtaining a comparative datasets between races will further highlight effector protein candidates, since these differ in abundance and composition between races.

The authors thank J. Chong, B. McCallum, X. Wang, C. Hiebert, G. Bakkeren, and F. You. This project was funded by a grant from Agriculture and Agrifood Canada to C.R., and a grant from TUBITAK to A.G.

The authors have declared no conflict of interest.

5 References

[1] Huerta-Espino, J., Singh, R. P., Germán, S., McCallum, B. D. et al., Global status of wheat leaf rust caused by *Puccinia triticina*. *Euphytica* 2011, *179*, 143–160.

[2] Bolton, M. D., Kolmer, J. A., Garvin, D. F., Wheat leaf rust caused by *Puccinia triticina*. *Molec. Plant Pathol.* 2008, *9*, 563–575.

[3] McCallum, B., Hiebert, C., Huerta-Espino, J., Cloutier, S., Wheat leaf rust, in: Sharma, I. (Ed.), *Disease Resistance in Wheat*, Chapter 3, CABI Plant Protection Series, Cambridge, MA 2012.

[4] Flor, H. H., Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* 1971, *9*, 275–296.

[5] Jones, J. D. G., Dangl, J. L., The plant immune system. *Nature* 2006, *444*, 323–329.

[6] Petre, B., Kamoun S., How do filamentous pathogens deliver effector proteins into plant cells? *PloS Biol.* 2014, *12*, e1001801.

[7] Chong, J., Harder, D. E., Rohringer, R., Cytochemical studies on *Puccinia graminis* f. sp. *tritici* in a compatible wheat host. II. Haustorium mother cell walls at the host cell penetration site, haustorial walls, and the extrahaustorial matrix. *Can. J. Bot.* 1986, *64*, 2561–2575.

[8] Hu, G., Rijkenberg, F. H. J., Scanning electron microscopy of early infection structure formation by *Puccinia recondita* f. sp. *tritici* on and in susceptible and resistant wheat lines. *Mycol. Res.* 1998, *102*, 391–399.

[9] Vögele, R. T., Mengden, K., Rust haustoria: nutrient uptake and beyond. *New Phytol.* 2003, *159*, 93–100.

[10] Hogenhout, S. A., Van der Hoorn, R. A., Terauchi, R., Kamoun, S., Emerging concepts in effector biology of plant-associated organisms. *Mol. Plant Microbe Interact.* 2009, *22*, 115–122.

[11] Ellis, J. G., Dodds, P. N., Lawrence, G. J., The role of secreted proteins in diseases of plants caused by rust, powdery mildew and smut fungi. *Curr. Opin. Microbiol.* 2007, *10*, 326–331.

[12] Stergiopoulos, I., deWit, P. J. G. M. Fungal Effector Proteins. *Annu. Rev. Phytopathol.* 2009, *47*, 233–263.

[13] Catanzariti, A-M., Dodds, P. N., Ellis, J. G., Avirulence proteins from haustoria-forming pathogens. *FEMS Microbiol. Lett.* 2006, *269*, 181–188.

[14] Ridout, C. J., Skamnioti, P., Porritt, O., Sacristan, S., Jones, J. D. G., Brown, J. K. M., Multiple avirulence paralogues in cereal powdery mildew fungi may contribute to parasite fitness and defeat of plant resistance. *Plant Cell* 2006, *18*, 2402–2418.

[15] Garnica, D., Upadhyaya, N., Dodds, P., Rathjen, J. Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS One* 2013, *8*, e67150.

[16] Link, T., Lang, P., Scheffler, B., Duke, M. et al., The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakospora Pachyrhizi* and their candidate effector families. *Molec. Plant Pathol.* 2014, *4*, 379–393.

[17] Godfrey, D., Zhang, Z., Saalbach, G., Thordal-Christensen, H., A proteomics study of barley powdery mildew haustoria. *Proteomics* 2009, *9*, 3222–3232.

[18] Song, X., Rampitsch, C., Soltani, B., Mauthe, W. et al., Proteome analysis of wheat leaf rust fungus, *Puccinia triticina*, infection structures enriched for haustoria. *Proteomics* 2011, *11*, 944–963.

[19] Rampitsch, C., Bykova, N. V., McCallum, B., Beimcik, E., Ens, W., Analysis of the wheat and *Puccinia triticina* (leaf rust) proteomes during a susceptible host—pathogen interaction. *Proteomics* 2006, *6*, 1897–1907.

[20] Tiburzy, R., Martins, E. M. F., Reisener, H. J., Isolation of haustoria of *Puccinia graminis* f.sp. *tritici* from wheat leaves. *Exp. Mycol*. 1992, *16*, 324–328.

[21] Jia, J., Zhao, S., Kong, X., Li, Y. et al., *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 2013, *496*, 91–95.

[22] Nesvizhskii, A. I., Keller, A., Kolker, E., Aebersold, R., A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 2003, *75*, 4646–4658.

[23] Zhang, Y., Wen, Z., Washburn, M. P., FLorens, L., Refinements to lable-free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* 2010, *82*, 2272–2281.

[24] Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., Robles, M., Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005, *21*, 3674–3676.

[25] Petersen, T., Brunak, S., von Heijne, G., Nielsen H., SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 2011, *8*, 785–786.

[26] Emanuelsson, O., Nielsen, H., Brunak S., von Heijne, G., Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 2000, *300*, 1005–1016.

[27] Bindschedler, L. V., Burgis, T. V., Mills, D. J. S., Ho, J. T. C. et al., *In Planta* proteomics and proteogenomics of the biotrophic barley fungal pathogen *Blumeria graminis* f. sp. hordei. *Mol. Cell. Proteomics* 2009, *8*, 2368–2381.

[28] Vögele, R., Hahn, M., Lohaus, G., Link, T. et al., Possible roles for mannitol and mannitol dehydrogenase in the biotrophic plant pathogen *Uromyces fabae*. *Plant Physiol.* 2005, *137*, 190–198.

[29] Stephenson, K., Sec-dependent protein translocation across biological membranes: evolutionary conservation of an essential protein transport pathway (review). *Mol. Membr. Biol.* 2005, *22*, 17–28.

[30] Torres, M. A., Jones, J. D. G., Dangl, J. L. Reactive oxygen species signaling in response to pathogens. *Plant Physiol*. 2006, *141*, 373–378.

[31] Rampitsch, C., Bykova, N. V., Proteomics and plant disease: advances in combating a major threat to the global food supply. *Proteomics* 2012, *12*, 673–690.

[32] Mubarakshina, M., Ivanov, B. N., Naydov, I. A., Hilier, W. et al., Production and diffusion of chloroplastic H_2O_2 and its implication to signalling. *J. Exp. Bot.* 2010, *61*, 3577–3587.

[33] Garnica, D. P., Upadhyaya, N. M., Dodds, P. N., Rathjen, J. P. Strategies for wheat stripe rust pathogenicity identified by transcriptome sequencing. *PLoS One* 2013, *8*, e67150

[34] Saunders, D. G. O., Win, J., Cano, L. M., Szabo, L. J. et al., Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust. *PLos One* 2012, *7*, e29847.

[35] Cantu, D., Segovia, V., MacLean, D., Bayles, R. et al., Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. *BMC Genomics* 2013, *14*, 270.

[36] Baxter, A., Mittler, R., Suzuki, N. ROS as key players in plant stress signalling. *J. Exp. Bot.* 2013, *65*, 1229–1240.

[37] Duplessis, S., Joly, D. L., Dodds, P. N. Rust effectors, in: Martin, F., Kamoun, S. (Eds.), *Effectors in Plant-Microbe Interactions*, John Wiley & Sons, UK 2012, pp. 155–193.

[38] Weßling, R., Schmidt, S M., Micali, C. O., Knaust, F., Transcriptome analysis of enriched *Golovinomyces orontii* haustoria by deep 454 pyrosequencing. *Fung. Genet. Biol.* 2012, *49*, 470–482.

