Review Article

Proteomes and transcriptomes of Apicomplexa - where's the message?

JM Wastling*, D Xia*, A Sohal†, M Chaussepied‡, A Pain†, G Langsley‡

Addresses: * Department of Pre-clinical Veterinary Science, Faculty of Veterinary Science, University of Liverpool, Liverpool, L69 7ZJ, UK. †Pathogen Genomics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA. ‡ Laboratoire de Biologie Cellulaire Comparative des Apicomplexes, Département de Maladies Infectieuses, Institut Cochin, Inserm, U567, CNRS, UMR 8104, Faculté de Médecine Paris V – Hôpital Cochin, 27, rue du Faubourg Saint-Jacques, 75014 Paris, France.

Abstract

The Apicomplexa now have some of the most comprehensive and integrated proteome datasets of all pathogenic organisms. Proteomic coverage is now at a level where it can be used to predict the potential biological involvement of proteins in these parasites, without having to defer to measurement of mRNA levels. Transcriptomic data for the Apicomplexa (microarrays, EST collections and MPSS) are also abundant, enabling us to investigate the extent to which global mRNA levels correlate with proteomic data. Here, we present a proteomic and transcriptomic perspective of gene expression in key apicomplexan parasites including *Plasmodium spp.*, *Toxoplasma gondii*, *Cryptosporidium parvum*, *Neospora caninum* and *Theileria spp*. and discuss the alternative views of gene expression that they provide. Although proteomic studies do not detect every gene for which transcripts are seen, many examples of readily detected proteins, whose corresponding genes display little or no detectable transcription are seen across the Apicomplexa. These examples are not easily explained by the “guilt by association”, or “stock and go hypotheses” of gene transcription. With the advent of ultra-high-throughput sequencing technologies there will be a quantum shift in transcriptional analysis which, combined with improving quantitative proteome data sets, will provide a core component of a systems-wide approach to studying the Apicomplexa.

Introduction

The last five years has seen proteomics become established as an integral component of the functional genomics repertoire. This growth, which has resulted from fundamental technical advances in mass spectrometry and bioinformatics, has been accompanied by the emergence of numerous large-scale proteomic experiments with substantial amounts of protein expression data being deposited into increasingly sophisticated on-line proteome resources. Protozoan parasites have not been left-behind in this rush for a proteomic perspective on gene expression; on the contrary, the Apicomplexa, for example, now have some of the most comprehensive and integrated proteomic datasets of all pathogenic organisms. This continuing appetite for proteomic data follows the recognition that examining the proteome has the potential to reveal far more about putative function than can be accounted for by transcriptional data alone. Furthermore, there has been little slow-down in the pace of technological advances in both mass spectrometry and the increasing sophistication of the bioinformatic resources that underpinned the emergence of proteomics little over a decade ago. Importantly, these advances have resulted in a significant increase in the depth and breadth of proteomics coverage that is realistically achievable in an experiment. Whereas a few years ago whole-cell (or so-called “global”) proteome surveys could do little more than sample just a small top-slice of the most abundant proteins, deep-mining of the proteome is now becoming increasingly feasible and with it the ability to monitor simultaneously the expression of thousands of proteins in a biological system.

Studies on apicomplexan parasites have been especially prominent promoting a proteomic understanding of gene expression in lower eukaryotes with large-scale proteomic surveys of *Plasmodium falicparum* (for example (Florens, L. et al., 2002; Lasonder, E. et al., 2002), *Cryptosporidium parvum* (Sanderson, S. J. et al., 2008; Snelling, W. J. et al., 2007) and *Toxoplasma gondii* (Xia, D. et al., 2008) being undertaken. Apicomplexan proteomics has also benefited from a range of advances such as improved sub-fractionation of complex protein mixtures prior to analysis (Nirmalan, N. et al., 2007), separation and analysis of apicomplexan sub-proteomes (Bradley, P. J. et al., 2005; Hu, K. et al., 2006; Zhou, X. W. et al., 2005) and a strong genome bioinformatic resource populated with increasingly accurate gene models (Bahl, A. et al., 2003; Gajria, B. et al., 2008; Heiges, M. et al., 2006). Proteomic studies have not only provided valuable corroborative evidence for predicted gene models by verifying the existence of thousands of hitherto hypothetical proteins, but have provided sufficient depth of coverage to begin to query the relationship between data acquired from transcriptional surveys, such as those from EST and microarray analysis, and actual protein expression. Such comparative surveys combining datasets from ESTs, microarray expression and proteomics have already raised fascinating questions pertaining to the link between transcription and translation in the Apicomplexa.

Despite magnitude advances in the accuracy and sensitivity of mass spectrometry, proteomics still suffers from the disadvantage that, unlike DNA, proteins cannot be amplified to increase the sensitivity of detection. The debate therefore remains on whether current proteomic technologies can provide sufficient depth and breadth of coverage to describe fully global gene expression. However, at a time when technological gaps in proteomics seems to be rapidly closing, questions over the relative biological meaning of proteomic and transcriptomic datasets are timely and especially pertinent to apicomplexan biology. In this paper we review advances in proteomic and transcriptional studies in the Apicomplexa, which have enabled us for the first time to examine the relationship between transcription and translation across this important group of parasites and that highlight some fascinating, if not yet fully understood, discrepancies between these types of data.

Although still imperfect, proteomics does after all provide first hand data on the functional products of gene expression – proteins and hence their putative function. Some argue that we should even look routinely to proteomics, rather than transcriptional patterns, to give us a more meaningful picture of the biological functions of genes. It is perhaps a sign of the breathtaking speed of advance in genomic analysis in the post-genomic era that transcriptional analysis is now seen by some as an “old technology” compared to its younger cousin, proteomics. Here, we argue that a combination of proteomics and transcriptional analysis provides the better perspective on gene expression, but these technologies are still in their infancy and we still have much to learn about the intimate and complex relationship between the two in the Apicomplexa.

A global proteomic perspective of the Apicomplexa

Recent global proteomic studies of apicomplexan parasites have massively increased the amount of protein expression data available for these parasites. In order to maximise the depth of coverage obtained in these analyses a combination of specialised separation and mass spectrometry approaches have been adopted. Thus, a typical experiment may involve gel-based analysis of parasite protein (one- or two-dimensional gel electrophoresis) followed by mass spectrometry of trypsin digested bands or spots. In addition, the parasite will also be analysed by whole shotgun proteome analysis, commonly known as "MudPIT". Whereas gel-based analysis reveals potentially more detailed protein data in the form of semi-quantitation and some post-translational information, shotgun analysis involves the separation of digested peptides in liquid phase, thus avoiding some of the common problems associated with gel separation of hydrophobic proteins, or proteins with extreme mass/pI. These approaches have enabled up to nearly 50% of the predicted proteome to be resolved on a proteomics platform. A summary of some of the whole-proteome projects in the Apicomplexa is presented in Table 1 and include those for *P. falciparum* (Florens, L. et al., 2002; Lasonder, E. et al., 2002) in which four different life cycle stages were identified using MudPIT and 1-DE gel LC-MS/MS. Comprehensive proteomic approaches have also been used to analyze the proteome of *P. berghei* and *P. yoelii* (Hall, N. et al., 2005; Khan, S. M. et al., 2005; Tarun, A. S. et al., 2008). Thus, proteomic analysis of *Plasmodium* has resulted in one of the most comprehensive datasets for any micro-organism, with data detailed proteomic coverage of up to 5 stages of the complex life cycle of *Plasmodium* species. These studies have been aimed at addressing important biological questions such as determining the functional characterisation of previously unknown cellular pathways (e.g. kinase pathways that regulate sex-specific functions in *Plasmodium* described by Khan et al. (Khan, S. M. et al., 2005). Doolan et al. studied combined genome and proteome data to identify a large number of sporozoite antigens that are expressed highly in sporozoites and showed high interferon-gamma response in the PBMCs of human volunteers, thus providing a list of novel candidates that could be tested as vaccine candidates (Doolan, D. L. et al., 2003). In a study which combined the transcriptome and proteome of *P. berghei*, evidence was obtained to demonstrate the developmental stage-specific translational control of mRNA transcripts and gave rise to the “stock and go” hypothesis (Hall, N. et al., 2005). Patra and co-workers undertook a study on the ookinete/zygote proteome of *P. gallinaceum*, the results of which represent a detailed proteomic view of *Plasmodium*-mosquito midgut interactions, fundamental to the development of a novel transmission blocking vaccine in malaria (Patra, K. P. et al., 2008).

Large scale protein expression profiling projects have also been carried out on the tachyzoite stage of *T. gondii* (Cohen, A. M. et al., 2002; Xia, D. et al., 2008) and similar approaches have been applied in investigating proteome of *C. parvum* sporozoites (Sanderson, S. J. et al., 2008; Snelling, W. J. et al., 2007). These studies have identified between approximately 30-40% of the “total” proteome. Further unpublished proteome data for *Toxoplasma* and *Cryptosporidium* are available via ApiDB, most notably a substantial additional data set for the *Toxoplasma* and *Cryptosporidium* proteome (unpublished, http://toro.aecom.yu.edu/biodefense/). More recently, proteome profiling of *N. caninum* has also been carried out (Wastling, unpublished data) and peptide evidence has so far been obtained for 660 of the predicted gene models in the current gene prediction set (www.genenedb.org), although this number is anticipated to increase substantially in the near future.

Sub-proteomes of the Apicomplexa

Apicomplexan sub-proteomes have been investigated in some detail, with analysis of the apical invasive organelles leading the field. Bradley and co-workers pioneered the proteomic investigation of apicomplexan rhoptry organelles, identifying many novel components of the rhoptry and rhoptry neck of *T. gondii* (Bradley, P. J. et al., 2005), whilst other key proteins released during host-cell invasion by tachyzoites have also been characterised using 2-DE and MudPIT (Zhou, X. W. et al., 2005; Zhou, X. W. et al., 2004). Rhoptry-enriched fractions have also been investigated in *Plasmodium* merozoites (Sam-Yellowe, T. Y. et al., 2004). The fractionated surface protein of parasite-infected erythrocytes of *P. falciparum* (Florens, L. et al., 2004) and the enriched cytoskeleton components of *T. gondii* (Hu, K. et al., 2006) and the cytoskeletal and membrane fractions of both *T. gondii* and *C. parvum* have also been examined (unpublished, http://toro.aecom.yu.edu/biodefense/).

Gene finding and curation in apicomplexans - a proteomic perspective:

Except in a very small number of cases where protein sequence is generated by *de novo* protein sequencing, the quality of proteomics identifications is entirely dependent on the sophistication of the gene models against which mass spectrometry data is searched against. Without accurately predicted gene models, proteomics experiments produce only a partial view of the proteome with considerable uncertainty surrounding the nature and number of proteins that may have been identified in any one experiment. Conversely, MS-generated peptide sequence data can be used in reverse logic as a powerful tool not only to provide confirmation, or correction of predicted protein-coding genes, but also to elucidate splicing patterns and as a key input to train gene finding algorithms (Choudhary, J. S. et al., 2001; Fermin, D. et al., 2006; Foissac, S. and Schiex, T., 2005; Tanner, S. et al., 2007; Sanderson, S. J. et al., 2008; Xia, D. et al., 2008). Many of the large-scale proteomics surveys of the Apicomplexa have focussed on genomes that have relatively accurate gene models such as *Cryptosporidium*, *Toxoplasma* and of course *Plasmodium*. In each of these examples proteomics has proved to be a powerful tool in corroborating thousands of hypothetical gene models. Moreover, in cases where several conflicting gene models exist for a particular region of DNA, MS-generated peptide data has been able to identify the most probable interpretation of gene structure and in some cases suggested completely alternative gene models. In one of the first genome scale proteomic survey studies in *P. falciparum*, a large number of good quality 'orphan' peptides (i.e. peptides that did not match to any existing predicted gene in *P. falciparum* during the time of publication in 2002) were used to curate manually gene boundaries and also to add missing exons in a number of genes (Florens, L. et al., 2002).

Apicomplexan proteomic database resources

Most apicomplexan proteomic datasets are now fully integrated into their respective publicly accessible online genome repositories (see Table 1). In a model developed first for *Cryptosporidium* (www.cryptodb.org) and *Toxoplasma* (www.toxodb.org), mass spectrometry data are now deposited in a standardised way in ApiDB. Thus, MS data can be interrogated in a variety of ways; for example by individual experiment; by sub-proteome; or by "alternative gene model" if variant gene annotations are suspected. One of the most informative ways of visualising proteomics data is a Genome Browser mode (GBrowse), where MS/MS peptide data are shown aligned against predicted gene structure as shown, for example, for the putative nicotinate phosphoribosyltransferasein gene (25.m01815) of *T. gondii* (Figure 1a). In this and other examples, peptide data can be seen alongside other forms of expression data such as EST analysis. A brief examination of a number of genes for which multiple forms of expression data are displayed (peptide and mRNA transcript) shows that whilst there is often broad agreement between gene expression indicators, discrepancies are also common. For example, Figure 1b illustrates the GBrowse view for a putative *Toxoplasma* oxidoreductase (37.m00770) which shows clearly that whilst substantial peptide evidence exists for this gene covering all four of the predicted exons, no corresponding EST data is present. Interestingly, this gene also shows microarrary transcript expression levels below the 25 percentile, indicating little or no transcript could be detected by microarray. Any potential biological role performed by these proteins would escape the "guilt by association" criteria that is based on inferring potential biological function from mRNA levels (Le Roch, K. G. et al., 2003).

Since all forms of expression data, including proteomics are now integrated into the same database, it is possible systematically to examine such correlations on a genome-wide scale in a way that would have been impossible in the past. The remainder of this review builds on these resources to examine some fundamental questions regarding the nature of proteomic and transcriptomic data in the Apicomplexa.

Merging transcriptional and proteomic data in the Apicomplexa

Extensive stage-specific transcriptional data have been acquired for apicomplexan parasites with the implicit assumption that transcriptional changes will reflect protein changes and that this will in turn enable key functions of proteins to be determined, for example those that play a role in stage-specific adaptations; this concept underlies the "guilt by association" hypothesis. dbEST (Boguski, M. S. et al., 1993) and ApiDB (Aurrecoechea, C. et al., 2007) host the largest collection of expressed sequence tags (EST) for the Apicomplexa. Serial analysis of gene expression (SAGE) projects have also been carried out for both *P. falciparum* and *T. gondii* (Gunasekera,

A. M. et al., 2003; Gunasekera, A. M. et al., 2007; Gunasekera, A. M. et al., 2004; Patankar, S. et al., 2001; Radke, J. R. et al., 2005). Microarray expression data are also available for *P. falciparum, P. berghei* and *T. gondii* (Ben, Mamoun C. et al., 2001; Bozdech, Z. et al., 2003; Hall, N. et al., 2005; Kidgell, C. et al., 2006; LaCount, D. J. et al., 2005).

At this time microarray data are missing for *Theileria* parasites, so to gain insights into parasite gene expression profiles a collection of ESTs from different *T. annulata* life cycle stages were sequenced, the majoritory of which (circa 10k) came from parasite infected macrophages (Pain, A. et al., 2005) and in the case of *T. parva*-infected lymphocytes an alternative powerful technique was used called Massively Parallel Signature Sequencing (MPSS) (Bishop, R. et al., 2005). MPSS is a PCR-based technique that gives sort (20bp) sequence tags of very high coverage and generates both sense and anti-sense data for a given gene. Importantly, since more than a million *T. parva* transcripts were sequenced, the number of times a transcript from the same gene was sequenced it generated a score (or a signature) that is an indication of the level of transcription of that gene. For *T. parva* MPSS scores ranged from 4 to 52 thousand per million, indicating a wide-range in gene transcription and more surprisingly, signatures could be detected for greater than 80% of genes (Bishop, R. et al., 2005). This suggests that at a given life cycle stage (schizont infected lymphocytes) the vast majority of *Theileria* genes are being transcribed, albeit at variable levels. Unfortunately, this wealth of MPSS data for *Theileria* is not backed up by proteomic data. Nonetheless, Bishop and colleagues noted that for 7 known schizont antigens the MPSS scores for the corresponding genes varied 1000-fold again underlining that protein and mRNA levels do not necessarily correlate.

Clearly, proteomic data for *Theileria* and its comparison with the MPSS data set would allow one to see how often abundant message translates into abundant protein.

There have been a small number of studies designed to obtain a simultaneous system-wide view of transcript and protein expression capable of testing the relationship between transcription and the proteome in *Plasmodium* (Hall, N. et al., 2005; Tarun, A. S. et al., 2008). Overall these studies have revealed a relatively weak correlation between mRNA and protein expression, with many genes being uniquely detected either by transcriptome or proteome. Similar discrepancies have been noted in a recent proteomic study of *Toxoplasma* (Xia, D. et al., 2008). In this study 2252 proteins were identified from the tachyzoite stage of the parasite using a multiplatform proteome approach. When these data are compared to genes that have transcriptional evidence from the same life-stage, 626 genes are detected solely by EST evidence and 1131 solely by microarray expression evidence (despite the 68% genome coverage by ESTs and nearly 99% microarray coverage). Significantly, peptide evidence for 72 tachyzoite genes was obtained from proteomics for which no transcripts were observed either by EST, or by microarray (Figure 2). This latter observation is particularly fascinating which argues against the common misconception that proteomics is relatively insensitive compared with transcriptional analysis. The presence of proteome evidence in the absence of detectable mRNA transcripts has also been noted in mammalian examples, where large numbers of proteins without transcriptional evidence were detected by proteomics in Hela cells (Cox, J. and Mann, M., 2007).

Given the abundance of good quality transcriptional and translational data across the Apicomplexa we decided to test systematically two related hypotheses concerning the relationship between proteins and their mRNA message: (1) that discrepancies between proteomic and transcriptional datasets occur frequently across the Apicomplexa (2) that orthologs of proteins that show conflicting transcriptional and proteomics profiles behave in the same way across the Apicomplexa i.e. we hoped to identify apicomplexan-wide groups of proteins which behaved aberrantly with respect to gene transcription and translation. To do this, EST and microarray data (where available) were first compared to their respective proteomics datasets for four species of Apicomplexa including *T. gondii* tachyzoites, *C. parvum* sporozoites, *P. falciparum* (all life-stages) and *N. caninum* tachyzoites in order to identify sub-sets of proteins for which transcriptional evidence was apparently missing (Figure 3). All the genes identified by major proteome projects listed in ApiDB were included in the analysis and comparative EST libraries and microarray expression data were used (no microarray data were available for *Neospora* or *Cryptosporidium*). Each column represents the total number of proteins identified by proteomics, with the red portion indicating proteins without any EST evidence and the green proportion showing proteins without either EST, or microarray data (where suitable microarray data are available). These data show clearly that a significant number of genes could be detected by proteomics for which neither EST, nor microarray evidence existed (103 for *Plasmodium* and 72 for *Toxoplasma*).

We reasoned that if the discrepancy between proteome and transcriptome is caused by a biological phenomenon that is conserved across apicomplexan parasites, the orthologs of “proteome only” proteins should have a similar expression pattern in the closely related species, i.e. have proteome evidence, but no transcript evidence. To test this we examined proteome and transcriptome expression signatures for *P. falciparum*, *T. gondii* and *N. caninum* (we did not include *C. parvum* because of its relatively poor EST coverage). First, the identities were obtained for every gene for which any form of proteome, EST or microarray expression data were available (in the case of *Plasmodium*, data were included from all life-stages). The criteria for inclusion were any gene that has (i) peptide evidence (ii) an EST hit (iii) ≥25% microarray expression. Next, proteins were sorted into the following categories (a) transcript present but no protein detected (b) protein detected but no EST evidence *and* no transcript detected by microarray ≥25% threshold (c) protein detected but no EST evidence. We then determined which proteins from each species were shared between each category using an orthology table derived from a one:many OrthoMCL analysis. Figure 4(a) shows that of the genes which lacked proteome data, but for which transcripts were present, significant numbers had orthologs in other species, with 313 being common between all three species. This is perhaps an unsurprising result, since it is known that certain types of proteins may be under-represented in proteomic analysis due to their physiochemical composition, low levels of expression or high rates of turn-over and degradation. Further analysis of these orthologous genes would be merited to determine why their corresponding peptide evidence is apparently missing.

Performing the same analysis in reverse reveals that out of the genes for which protein evidence occurs in the absence of detectable EST and microarray transcripts (356 across all species), only a handful are shared as orthologs (Figure 4b), although when the analysis is performed with EST data alone (Figure 4c) a larger number of proteins are shared, including two orthologs seen across all three species. In general however, these data appear to disprove our second hypothesis that a shared biological phenomenon might account for these apparently contradictory expression patterns across the phylum.

From the analysis performed above, there is no apparent underlying rule that dominates the discrepancy between proteome and transcriptome across apicomplexan parasites, except perhaps for a very small number of genes. There are some interesting candidates in the comparison (59.m00090, coatomer protein gamma 2-subunit) which consistently produces convincing peptide evidence (e.g. 37 peptides and 53 spectra in *T. gondii*), but is without transcript evidence at the EST level in *T. gondii*, *N. caninum* and *C. parvum,* with only a single EST seen in a *P. falicparum* blood-stage EST library. The ortholog of this gene in *T. parva* also appears in the lower than 25 percentile MPSS expression analysis (Bishop, R. et al., 2005) and interestingly an orthology search in *Saccharomyces cerevisiae* (YNL287W) also reveals a gene for which no EST evidence has been found, although it is detected by proteomics (The Global Proteome Machine Database) (Craig, R. et al., 2004). It is not known why the coatomer protein, a Golgi-coat associated protein, appears so reluctant to reveal itself at the transcript level across not just the Apicomplexa, but other eukaryotes.

Despite their discrepancies, it is clear that both transcriptomes and proteomes continue to provide experimental evidence for gene expression following the central dogma of Gene-Transcription-Translation. Apparent contradictions between the datasets for a specific set of genes may still be accounted for by genuine biological phenomena such as post-transcriptional control mechanisms as those described by

Hall and colleagues (Hall, N. et al., 2005), who combined genome-scale transcriptome and proteome data for several life cycle stages of *P. berghei* and observed evidence for post-transcriptional gene silencing through translational repression of messenger RNA during sexual development of the parasite. A further explanation may be the "stock and go hypothesis" in *Plasmodium* (Mair, G. R. et al., 2006), where translational repression of messenger RNAs (mRNAs) may play an important role in sexual differentiation and gametogenesis.

Proteomics and transcriptomics at the host-cell interface

It would be remiss to end a review on gene expression in the Apicomplexa without acknowledging the intimate relationship between parasite and host-cell gene expression. A considerable number of studies have been undertaken to describe global host-cell gene expression changes associated with the infection of Apicomplexa and other intracellular protozoa, but these are dominated by transcriptional rather than proteomic experiments (summarised in Table 2). It is immediately clear that even comparisons between various microarray studies are difficult, because of the considerable experimental variables introduced into each study, including infection time-course (Blader, I. J. et al., 2001; Jensen, K. et al., 2008; Knight, B. C. et al., 2006; Okomo-Adhiambo, M. et al., 2006; Vaena de, Avalos S. et al., 2002), parasite strain (Knight, B. C. et al., 2006), and host cell type (Chaussabel, D. et al., 2003; Jensen, K. et al., 2008). Notably, the importance of the experimental system chosen and especially the host cell type is critical. For example, infection of macrophages and dendritic cells with various pathogens will elicit quite distinct transcriptional responses (Chaussabel, D. et al., 2003) illustrating not only a pathogen-specific response, but also a cell-type specific response. For technical reasons, the microarrays are often not made from the host cell type that is naturally infected and this complicates further interpretations regarding disease. When comparing different analyses the precise genetic background of the relevant natural host cell type also has to be taken into consideration, as *T. annulata*-infected macrophages from two different breeds of cow (resistant and susceptible to disease) show changes in their expression profiles when infected with the same genetically cloned parasite (Jensen, K. et al., 2008).

The modulation of the host-cell proteome by *T. gondii* has been examined in depth by quantitative two-dimensional electrophoresis (Nelson, M. M. et al., 2008) providing an opportunity to compare directly proteomic data with transcriptional data from an identically designed experiment (Blader, I. J. et al., 2001). In this analysis only a weak relationship was observed between host-cell transcriptional data and host proteome data at the individual gene level (Nelson, M. M. et al., 2008). Significantly however, despite differences in detail, both transcriptomic and proteomic analyses came to similar overall conclusions regarding the modulation of key host-cell pathways by *Toxoplasma*. This perhaps illustrates an important overriding principle when dealing with transcript and protein expression data: that they are complementary data which, although linked intimately, are capable of providing a different, rather than conflicting perspective on the same problem.

Conclusions and outlook

It is important to acknowledge that both proteomics and transcriptomics are still relatively young technologies, representing some of the first generation of genome-wide data to follow the apicomplexan genome sequencing projects. Until recently we have been in an exploratory phase, systematically cataloguing what is expressed by apicomplexan parasites, when expression occurs (stage-specific expression) and where expression occurs (organelle proteomic). Whilst these studies have indeed been pioneering, the focus of proteomics is about to be rapidly altered and extended to the proteomics of protein modifications, drug-parasite and host-parasite interactions. In particular the emphasis will shift to more sensitive and accurate proteomic measurements, with quantitative proteomics enabling us to undertake more meaningful comparisons between transcript abundance and protein abundance. Advances in the context of transcriptional analysis are also anticipated such as the application of MPSS to other Apicomplexa over and above *Theileria*. With the advent of ultra-high-throughput sequencing technologies [e.g. Roche (454), Illumina(Solexa); ABI-SoliD], there will be a quantum shift in our ability to fine-map the transcript boundaries of the genes by directly sequencing the transcripts to a high coverage (Graveley, B. R., 2008). Recent studies using these state-of-art techniques have provided unprecedented insight into the transcription states (including alternative splice variants and a large number of previously unrecognised transcripts) in the fission yeast *S. pombe* and human at a single nucleotide resolution (Sultan, M. et al., 2008; Wilhelm, B. T. et al., 2008). Similar transcript sequencing studies are now also underway in apicomplexan parasites and thus the accuracy of gene predictions is expected to get significantly higher in the near future that in turn, will prove highly beneficial to the proteomics. As demonstrated for *T. parva*, the depth of transcript sequencing will also allow us to determine the dynamic range (i.e. signature) of a given transcript. The development of these advanced technologies and their application to other Apicomplexa are likely to reveal even more complexity in the relationship between protein and its message. They will also provide an ever more powerful tool to determine the extent of non-coding RNAs (anti-sense, micro and macro) and their eventual contribution to the success Apicomplexa have demonstrated in parasitizing such a wide range of host cells.

Acknowledgements

The authors gratefully acknowledge support form the COST 857 action “Apicomplexan biology in the post-genomic era”, which provided an invaluable forum for much of the discussion contained in this manuscript.

Reference List

Aurrecoechea, C., Heiges, M., Wang, H., Wang, Z., Fischer, S., Rhodes, P., Miller, J., Kraemer, E., Stoeckert, C. J., Jr., Roos, D. S., and Kissinger, J. C., 2007. ApiDB: integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Res. 35, D427-D430.

Bahl, A., Brunk, B., Crabtree, J., Fraunholz, M. J., Gajria, B., Grant, G. R., Ginsburg, H., Gupta, D., Kissinger, J. C., Labo, P., Li, L., Mailman, M. D., Milgram, A. J., Pearson, D. S., Roos, D. S., Schug, J., Stoeckert, C. J., Jr., and Whetzel, P., 2003. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. Nucleic Acids Res. 31, 212-215.

Ben, Mamoun C., Gluzman, I. Y., Hott, C., MacMillan, S. K., Amarakone, A. S., Anderson, D. L., Carlton, J. M., Dame, J. B., Chakrabarti, D., Martin, R. K., Brownstein, B. H., and Goldberg, D. E., 2001. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite Plasmodium falciparum revealed by microarray analysis. Mol. Microbiol. 39, 26-36.

Bishop, R., Shah, T., Pelle, R., Hoyle, D., Pearson, T., Haines, L., Brass, A., Hulme, H., Graham, S. P., Taracha, E. L., Kanga, S., Lu, C., Hass, B., Wortman, J., White, O., Gardner, M. J., Nene, V., and de Villiers, E. P., 2005. Analysis of the transcriptome of the protozoan Theileria parva using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. Nucleic Acids Res. 33, 5503-5511.

Blader, I. J., Manger, I. D., and Boothroyd, J. C., 2001. Microarray analysis reveals previously unknown changes in Toxoplasma gondii-infected human cells
1. J. Biol. Chem. 276, 24223-24231.

Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M., 1993. dbEST--database for "expressed sequence tags"
1. Nat. Genet. 4, 332-333.

Bozdech, Z., Llinas, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L., 2003. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum
1. PLoS. Biol. 1, E5-

Bradley, P. J., Ward, C., Cheng, S. J., Alexander, D. L., Coller, S., Coombs, G. H., Dunn, J. D., Ferguson, D. J., Sanderson, S. J., Wastling, J. M., and Boothroyd, J. C., 2005. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. J. Biol. Chem. 280, 34245-34258.

Chaussabel, D., Semnani, R. T., McDowell, M. A., Sacks, D., Sher, A., and Nutman, T. B., 2003. Unique gene expression profiles of human macrophages and dendritic cells to phylogenetically distinct parasites
1. Blood. 102, 672-681.

Choudhary, J. S., Blackstock, W. P., Creasy, D. M., and Cottrell, J. S., 2001. Matching peptide mass spectra to EST and genomic DNA databases. Trends Biotechnol. 19, S17-S22.

Cohen, A. M., Rumpel, K., Coombs, G. H., and Wastling, J. M., 2002. Characterisation of global protein expression by two-dimensional electrophoresis and mass spectrometry: proteomics of Toxoplasma gondii1. Int. J. Parasitol. 32, 39-51.

Cox, J. and Mann, M., 2007. Is proteomics the new genomics? Cell. 130, 395-398.

Craig, R., Cortens, J. P., and Beavis, R. C., 2004. Open source system for analyzing, validating, and storing protein identification data. J. Proteome. Res. 3, 1234-1242.

Deng, M., Lancto, C. A., and Abrahamsen, M. S., 2004. Cryptosporidium parvum regulation of human epithelial cell gene expression
1. Int. J. Parasitol. 34, 73-82.

Doolan, D. L., Southwood, S., Freilich, D. A., Sidney, J., Graber, N. L., Shatney, L., Bebris, L., Florens, L., Dobano, C., Witney, A. A., Appella, E., Hoffman, S. L., Yates, J. R., III, Carucci, D. J., and Sette, A., 2003. Identification of Plasmodium falciparum antigens by antigenic analysis of genomic and proteomic data. Proc. Natl. Acad. Sci. U. S. A. 100, 9952-9957.

Fermin, D., Allen, B. B., Blackwell, T. W., Menon, R., Adamski, M., Xu, Y., Ulintz, P., Omenn, G. S., and States, D. J., 2006. Novel gene and gene model detection using a whole genome open reading frame analysis in proteomics. Genome Biol. 7, R35-

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D. J., Yates, J. R., III, and Wub, Y., 2004. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol. Biochem. Parasitol. 135, 1-11.

Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., Wu, Y., Gardner, M. J., Holder, A. A., Sinden, R. E., Yates, J. R., and Carucci, D. J., 2002. A proteomic view of the Plasmodium falciparum life cycle. Nature. 419, 520-526.

Foissac, S. and Schiex, T., 2005. Integrating alternative splicing detection into gene prediction. BMC. Bioinformatics. 6, 25-34.

Gail, M., Gross, U., and Bohne, W., 2001. Transcriptional profile of Toxoplasma gondii-infected human fibroblasts as revealed by gene-array hybridization
1. Mol. Genet. Genomics. 265, 905-912.

Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X., Heiges, M., Iodice, J., Kissinger, J. C., Mackey, A. J., Pinney, D. F., Roos, D. S., Stoeckert, C. J., Jr., Wang, H., and Brunk, B. P., 2008. ToxoDB: an integrated Toxoplasma gondii database resource. Nucleic Acids Res. 36, D553-D556.

Graveley, B. R., 2008. Molecular biology: power sequencing. Nature. 453, 1197-1198.

Gunasekera, A. M., Myrick, A., Le, Roch K., Winzeler, E., and Wirth, D. F., 2007. Plasmodium falciparum: genome wide perturbations in transcript profiles among mixed stage cultures after chloroquine treatment
1. Exp. Parasitol. 117, 87-92.

Gunasekera, A. M., Patankar, S., Schug, J., Eisen, G., Kissinger, J., Roos, D., and Wirth, D. F., 2004. Widespread distribution of antisense transcripts in the Plasmodium falciparum genome
1. Mol. Biochem. Parasitol. 136, 35-42.

Gunasekera, A. M., Patankar, S., Schug, J., Eisen, G., and Wirth, D. F., 2003. Drug-induced alterations in gene expression of the asexual blood forms of Plasmodium falciparum
1. Mol. Microbiol. 50, 1229-1239.

Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., Harris, B., Harris, D., Churcher, C., Quail, M. A., Ormond, D., Doggett, J., Trueman, H. E., Mendoza, J., Bidwell, S. L., Rajandream, M. A., Carucci, D. J., Yates, J. R., III, Kafatos, F. C., Janse, C. J., Barrell, B., Turner, C. M., Waters, A. P., and Sinden, R. E., 2005. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses1. Science. 307, 82-86.

Heiges, M., Wang, H., Robinson, E., Aurrecoechea, C., Gao, X., Kaluskar, N., Rhodes, P., Wang, S., He, C. Z., Su, Y., Miller, J., Kraemer, E., and Kissinger, J. C., 2006. CryptoDB: a Cryptosporidium bioinformatics resource update. Nucleic Acids Res. 34, D419-D422.

Hu, K., Johnson, J., Florens, L., Fraunholz, M., Suravajjala, S., DiLullo, C., Yates, J., Roos, D. S., and Murray, J. M., 2006. Cytoskeletal components of an invasion machine--the apical complex of Toxoplasma gondii. PLoS. Pathog. 2, e13-

Jensen, K., Paxton, E., Waddington, D., Talbot, R., Darghouth, M. A., and Glass, E. J., 2008. Differences in the transcriptional responses induced by Theileria annulata infection in bovine monocytes derived from resistant and susceptible cattle breeds
1. Int. J. Parasitol. 38, 313-325.

Khan, S. M., Franke-Fayard, B., Mair, G. R., Lasonder, E., Janse, C. J., Mann, M., and Waters, A. P., 2005. Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. Cell. 121, 675-687.

Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., Johnson, J. R., Le, Roch K., Sarr, O., Ndir, O., Mboup, S., Batalov, S., Wirth, D. F., and Winzeler, E. A., 2006. A systematic map of genetic variation in Plasmodium falciparum. PLoS. Pathog. 2, e57-

Knight, B. C., Kissane, S., Falciani, F., Salmon, M., Stanford, M. R., and Wallace, G. R., 2006. Expression analysis of immune response genes of Muller cells infected with Toxoplasma gondii 1. J. Neuroimmunol. 179, 126-131.

LaCount, D. J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J. R., Schoenfeld, L. W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S., and Hughes, R. E., 2005. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature. 438, 103-107.

Lasonder, E., Ishihama, Y., Andersen, J. S., Vermunt, A. M., Pain, A., Sauerwein, R. W., Eling, W. M., Hall, N., Waters, A. P., Stunnenberg, H. G., and Mann, M., 2002. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature. 419, 537-542.

Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., de, la, V, Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A., 2003. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 301, 1503-1508.

Mair, G. R., Braks, J. A., Garver, L. S., Wiegant, J. C., Hall, N., Dirks, R. W., Khan, S. M., Dimopoulos, G., Janse, C. J., and Waters, A. P., 2006. Regulation of sexual development of Plasmodium by translational repression. Science. 313, 667-669.

Nelson, M. M., Jones, A. R., Carmen, J. C., Sinai, A. P., Burchmore, R., and Wastling, J. M., 2008. Modulation of the host cell proteome by the intracellular apicomplexan parasite Toxoplasma gondii 1. Infect. Immun. 76, 828-844.

Nirmalan, N., Flett, F., Skinner, T., Hyde, J. E., and Sims, P. F., 2007. Microscale solution isoelectric focusing as an effective strategy enabling containment of hemeoglobin-derived products for high-resolution gel-based analysis of the Plasmodium falciparum proteome. J. Proteome. Res. 6, 3780-3787.

Okomo-Adhiambo, M., Beattie, C., and Rink, A., 2006. cDNA microarray analysis of host-pathogen interactions in a porcine in vitro model for Toxoplasma gondii infection 1. Infect. Immun. 74, 4254-4265.

Pain, A., Renauld, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R. M., Cronin, A., de Villiers, E. P., Fraser, A., Fosker, N., Gardner, M., Goble, A., Griffiths-Jones, S., Harris, D. E., Katzer, F., Larke, N., Lord, A., Maser, P., McKellar, S., Mooney, P., Morton, F., Nene, V., O'Neil, S., Price, C., Quail, M. A., Rabbinowitsch, E., Rawlings, N. D., Rutter, S., Saunders, D., Seeger, K., Shah, T., Squares, R., Squares, S., Tivey, A., Walker, A. R., Woodward, J., Dobbelaere, D. A., Langsley, G., Rajandream, M. A., McKeever, D., Shiels, B., Tait, A., Barrell, B., and Hall, N., 2005. Genome of the host-cell transforming parasite Theileria annulata compared with T. parva 1. Science. 309, 131-133.

Patankar, S., Munasinghe, A., Shoaibi, A., Cummings, L. M., and Wirth, D. F., 2001. Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite 1. Mol. Biol. Cell. 12, 3114-3125.

Patra, K. P., Johnson, J. R., Cantin, G. T., Yates, J. R., III, and Vinetz, J. M., 2008. Proteomic analysis of zygote and ookinete stages of the avian malaria parasite Plasmodium gallinaceum delineates the homologous proteomes of the lethal human malaria parasite Plasmodium falciparum. Proteomics. 8, 2492-2499.

Radke, J. R., Behnke, M. S., Mackey, A. J., Radke, J. B., Roos, D. S., and White, M. W., 2005. The transcriptome of Toxoplasma gondii. BMC. Biol. 3, 26-

Sam-Yellowe, T. Y., Florens, L., Wang, T., Raine, J. D., Carucci, D. J., Sinden, R., and Yates, J. R., III, 2004. Proteome analysis of rhoptry-enriched fractions isolated from Plasmodium merozoites. J. Proteome. Res. 3, 995-1001.

Sanderson, S. J., Xia, D., Prieto, H., Yates, J., Heiges, M., Kissinger, J. C., Bromley, E., Lal, K., Sinden, R. E., Tomley, F., and Wastling, J. M., 2008. Determining the protein repertoire of Cryptosporidium parvum sporozoites. Proteomics. 8, 1398-1414.

Slater, G. S. and Birney, E., 2005. Automated generation of heuristics for biological sequence comparison 1. BMC. Bioinformatics. 6, 31-

Snelling, W. J., Lin, Q., Moore, J. E., Millar, B. C., Tosini, F., Pozio, E., Dooley, J. S., and Lowery, C. J., 2007. Proteomics analysis and protein expression during sporozoite excystation of Cryptosporidium parvum (Coccidia, Apicomplexa). Mol. Cell Proteomics. 6, 346-355.

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keeffe, S., Haas, S., Vingron, M., Lehrach, H., and Yaspo, M. L., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science. 321, 956-960.

Tanner, S., Shen, Z., Ng, J., Florea, L., Guigo, R., Briggs, S. P., and Bafna, V., 2007. Improving gene annotation using peptide mass spectrometry. Genome Res. 17, 231-239.

Tarun, A. S., Peng, X., Dumpit, R. F., Ogata, Y., Silva-Rivera, H., Camargo, N., Daly, T. M., Bergman, L. W., and Kappe, S. H., 2008. A combined transcriptome and proteome survey of malaria parasite liver stages. Proc. Natl. Acad. Sci. U. S. A. 105, 305-310.

Vaena de, Avalos S., Blader, I. J., Fisher, M., Boothroyd, J. C., and Burleigh, B. A., 2002. Immediate/early response to Trypanosoma cruzi infection involves minimal modulation of host cell transcription 1. J. Biol. Chem. 277, 639-644.

Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bahler, J., 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature. 453, 1239-1243.

Xia, D., Sanderson, S. J., Jones, A. R., Prieto, J. H., Yates, J. R., Bromley, E., Tomley, F. M., Lal, K., Sinden, R. E., Brunk, B. P., Roos, D. S., and Wastling, J. M., 2008. The proteome of Toxoplasma gondii: integration with the genome provides novel insights into gene expression and annotation. Genome Biol. 9, R116-

Zhou, X. W., Blackman, M. J., Howell, S. A., and Carruthers, V. B., 2004. Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex. Mol. Cell Proteomics. 3, 565-576.

Zhou, X. W., Kafsack, B. F., Cole, R. N., Beckett, P., Shen, R. F., and Carruthers, V. B., 2005. The opportunistic pathogen *Toxoplasma gondii* deploys a diverse legion of invasion and survival proteins. J. Biol. Chem. 280, 34233-34244.

Table 1

Summary of global proteomic studies in the Apicomplexa

Species	**Life Cycle Stage**	**Platform**	**References**	**Database resource?**	**Number of unique proteins identified**	**Estimated Proportion of Proteome**	**Transcript Expression Data?**
P.falciparum	Sporozoite, Merozoite, Trophozoite Gametocyte Trophozoite/ Schizont	1-DE Gel-LC MS/MS MudPIT	(Florens, L. et al., 2002; Florens, L. et al., 2004; Lasonder, E. et al., 2002)	ApiDB	2427	~45%	EST SAGE, Microarray
P.berghei	Gametocyte, Asexual blood stage, Ookinete	1-DE Gel LC-MS/MS MudPIT	(Hall, N. et al., 2005; Khan, S. M. et al., 2005)	ApiDB	2924	~24%	EST Microarray
P. yoelii	Liver Stage Schizont	1-DE Gel LC-MS/MS	(Tarun, A. S. et al., 2008)	None	816	~10%	Microarray
T.gondii	Tachyzoite	1-DE Gel LC-MS/MS, 2-DE Gel LC-MS/MS MudPIT	(Bradley, P. J. et al., 2005; Hu, K. et al., 2006; Xia, D. et al., 2008)	ApiDB	2457	~31%	EST SAGE Microarray
C.parvum	Oocyst/ sporozoite	1-DE Gel LC-MS/MS, 2-DE Gel LC-MS/MS MudPIT	(Sanderson, S. J. et al., 2008; Snelling, W. J. et al., 2007)	ApiDB	1322	~30%	EST
N.caninum	Tachyzoite	MudPIT	Un-published	None	660 genes	~15%	EST

Table 2

Summary of host-cell transcriptional studies in the apicomplexan infections

Parasite	Target cells	Species	Time points	Microarray	References
Theileria annulata sporozoïtes	Peripheral-blood monocytes	*Bos taurus* (S) & *B. indicus* (R)	0, 2, 72hrs	Cattle 5K Immune cDNA (ARK-Genomics)	(Jensen, K. et al., 2008)
Toxoplasma gondii tachyzoite strain TS-4	PK13, porcine kidney epithelial cell line	*Sus scrofa*	0, 1, 2, 4, 6, 24, 48, 72hrs	Porcine custom cDNA	(Okomo-Adhiambo, M. et al., 2006)
- *Toxoplasma gondii* tachyzoite strain RH	Peripheral-blood monocytes differentiated to macrophages or dendritic cells	*Homo sapiens*	0, 16hrs	HU95A (Affymetrix) probe array	(Chaussabel, D. et al., 2003)
Toxoplasma gondii tachizoites RH strain	Human foreskin fibroblasts (HFF)	*Homo sapiens*	0, 24hrs	Human cDNA array (Human Atlas Array, Clontech)	(Gail, M. et al., 2001)
Toxoplasma gondii RH strain tachizoites and Prugniaud strain cysts	Human Müller cell line (MOI-M1)	*Homo sapiens*	0, 2, 24hrs	Human apoptosis and custom probe arrays (Affymetrix)	(Knight, B. C. et al., 2006)
Toxoplasma gondii	Human foreskin fibroblasts (HFF)	*Homo sapiens*	0, 1, 2, 4, 6, 24hrs	Human custom cDNA	(Blader, I. J. et al., 2001)
Cryptosporidium parvum oocytsts	HCT-8 epithelial cell line	*Homo sapiens*	0, 24hrs	HG-U95Av2 probe array (Affymetrix)	(Deng, M. et al., 2004)

Figure 1

Visualisation of proteomic and transcriptomic expression data in ToxoDB

(a) A screenshot of the annotated *T. gondii* gene 25.m01815 (nicotinate phosphoribosyltransferase, putative) on ToxoDB Genome Browser (www.toxodb.org). Predicted gene structures of gene 25.m01815, where blue boxes represent exons, are shown on the top of the figure. EST and proteome (MS/MS peptide) evidence identified for this gene are aligned underneath the gene sequence. The relationship between proteomic (peptide) and transcriptomic (EST) data can be directly visualised. Note that peptide evidence confirms several predicted intron-exon boundaries (shown by the joins between peptides). (b) GBrowse view for a putative *Toxoplasma* oxidoreductase (37.m00770) gene which shows clearly that whilst substantial peptide evidence exists for this gene covering all four of the predicted exons, no corresponding EST data is present. Interestingly, this gene also shows microarrary transcript levels below the 25 percentile, indicating little or no transcript could be detected by microarray.

Figure 2

Genes with proteome and transcriptome evidence in *T. gondii*

Diagram illustrating the relationship between proteomics, EST and microarray gene expression data in *T. gondii* (data from (Xia, D. et al., 2008). In total 2252 non-redundant proteins were identified from *T.gondii* tachyzoites (blue circle). These were compared with genes that have tachyzoite EST evidence (green circle) and microarray expression data (orange circle), where higher than 25 expression percentile is observed. The data show that 626 genes have uniquely EST evidence, 1131 genes have uniquely microarray expression evidence, whilst 72 tachyzoite genes are uniquely identified by peptide data and have no transcript expression evidence.

Figure 3

Proteome and transcriptome comparisons across four species of Apicomplexa

The numbers of proteins identified by peptide evidence in *T. gondii* tachyzoites, *C. parvum* sporozoites, *P. falciparum* (all life-stages) and *N. caninum* tachyzoites are shown. The red portion indicates proteins without EST evidence and the green portion indicates genes without EST and microarray evidence (less than 25 expression percentile). Note that no microarray data were available for *Neospora* or *Cryptosporidium*. All the genes identified by major proteome projects listed in ApiDB are included and comparative EST libraries and microarray expression data were used in the analysis. For *N.caninum*, ESTs were downloaded from dbEST and were aligned to genes that have proteomic evidence under whole genome scaffold using software Exonerate (Slater, G. S. and Birney, E., 2005).

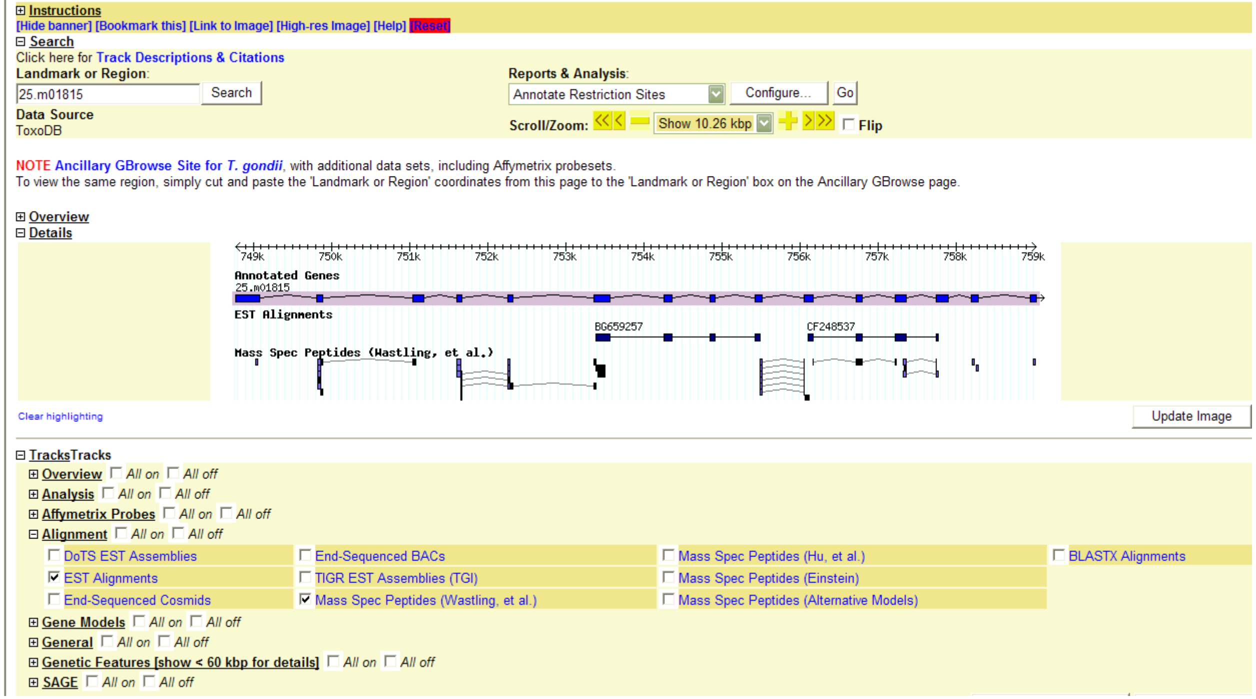
Figure 4

Genes from three Apicomplexa which exhibit discrepancies between transcriptional data and proteome data

Each circle represents the number of genes for which a discrepancy was seen between transcriptional data and proteome data for *P. falciparum*, *T. gondii* and *N. caninum* based on (a) transcript present but no protein detected (b) protein detected but no EST evidence *and* no transcript detected by microarray ≥25% threshold (c) protein detected but no EST evidence. The intersections show the numbers of orthologs (as determined by OrthoMCL) shared between the species that exhibit contradictory transcriptional and protein expression patterns.

Figure 1

(a)



⊞ Instructions
[Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset]
⊟ Search
Click here for Track Descriptions & Citations
Landmark or Region:
25.m01815 Search
Data Source
ToxoDB
Reports & Analysis:
Annotate Restriction Sites Configure... Go
Scroll/Zoom: << < − Show 10.26 kbp + > >> ☐ Flip

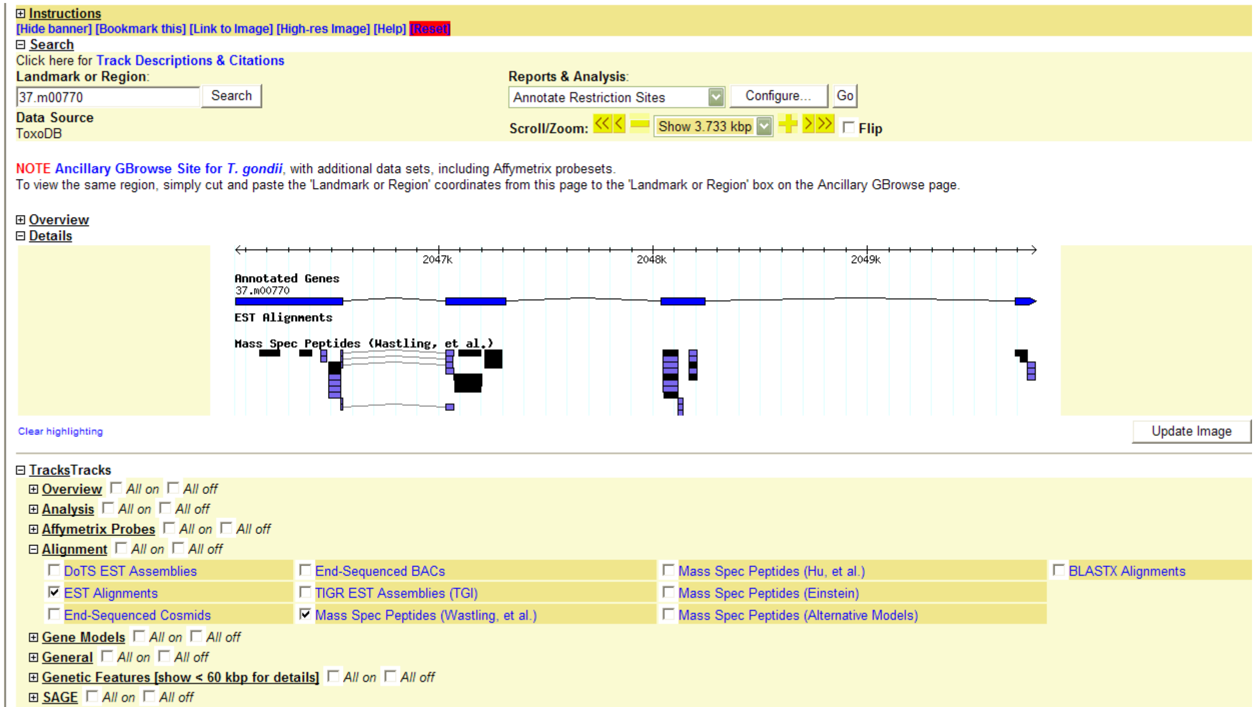
NOTE Ancillary GBrowse Site for *T. gondii*, with additional data sets, including Affymetrix probesets.
To view the same region, simply cut and paste the 'Landmark or Region' coordinates from this page to the 'Landmark or Region' box on the Ancillary GBrowse page.

⊞ Overview
⊟ Details

Clear highlighting
Update Image

⊟ TracksTracks
⊞ Overview ☐ All on ☐ All off
⊞ Analysis ☐ All on ☐ All off
⊞ Affymetrix Probes ☐ All on ☐ All off
⊟ Alignment ☐ All on ☐ All off
☐ DoTS EST Assemblies ☐ End-Sequenced BACs ☐ Mass Spec Peptides (Hu, et al.) ☐ BLASTX Alignments
☑ EST Alignments ☐ TIGR EST Assemblies (TGI) ☐ Mass Spec Peptides (Einstein)
☐ End-Sequenced Cosmids ☑ Mass Spec Peptides (Wastling, et al.) ☐ Mass Spec Peptides (Alternative Models)
⊞ Gene Models ☐ All on ☐ All off
⊞ General ☐ All on ☐ All off
⊞ Genetic Features [show < 60 kbp for details] ☐ All on ☐ All off
⊞ SAGE ☐ All on ☐ All off

(b)



⊞ Instructions
[Hide banner] [Bookmark this] [Link to Image] [High-res Image] [Help] [Reset]
⊟ Search
Click here for Track Descriptions & Citations
Landmark or Region:
37.m00770 Search
Data Source
ToxoDB
Reports & Analysis:
Annotate Restriction Sites Configure... Go
Scroll/Zoom: << < − Show 3.733 kbp + > >> ☐ Flip

NOTE Ancillary GBrowse Site for *T. gondii*, with additional data sets, including Affymetrix probesets.
To view the same region, simply cut and paste the 'Landmark or Region' coordinates from this page to the 'Landmark or Region' box on the Ancillary GBrowse page.

⊞ Overview
⊟ Details

Clear highlighting
Update Image

⊟ TracksTracks
⊞ Overview ☐ All on ☐ All off
⊞ Analysis ☐ All on ☐ All off
⊞ Affymetrix Probes ☐ All on ☐ All off
⊟ Alignment ☐ All on ☐ All off
☐ DoTS EST Assemblies ☐ End-Sequenced BACs ☐ Mass Spec Peptides (Hu, et al.) ☐ BLASTX Alignments
☑ EST Alignments ☐ TIGR EST Assemblies (TGI) ☐ Mass Spec Peptides (Einstein)
☐ End-Sequenced Cosmids ☑ Mass Spec Peptides (Wastling, et al.) ☐ Mass Spec Peptides (Alternative Models)
⊞ Gene Models ☐ All on ☐ All off
⊞ General ☐ All on ☐ All off
⊞ Genetic Features [show < 60 kbp for details] ☐ All on ☐ All off
⊞ SAGE ☐ All on ☐ All off

Figure 2

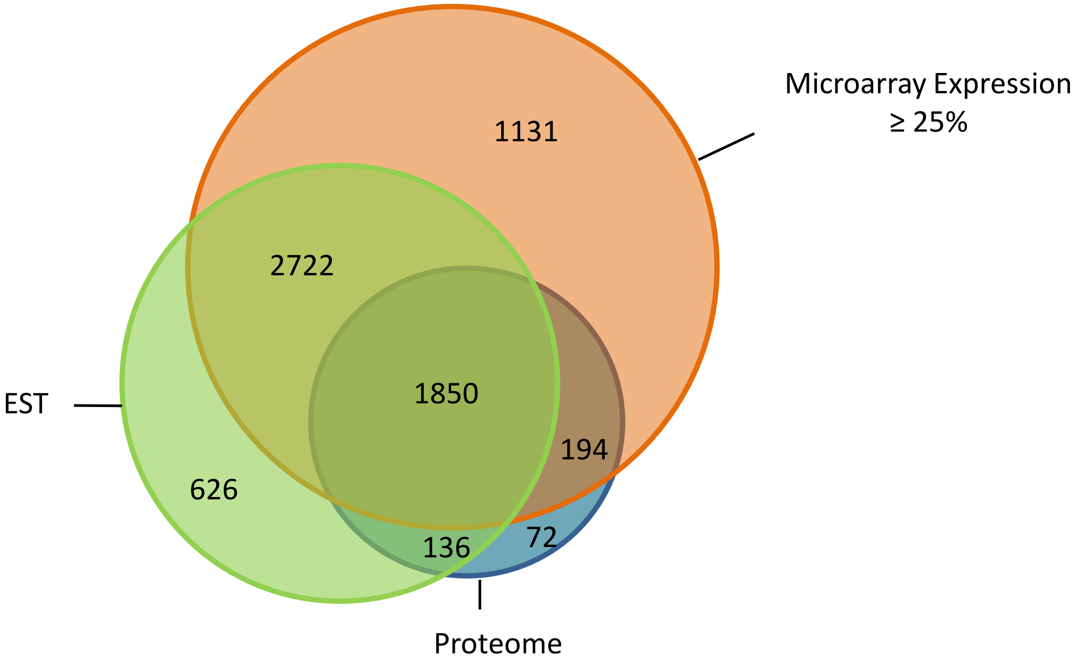

Microarray Expression
≥ 25%
1131
2722
EST
1850
194
626
136
72
Proteome

Figure 3

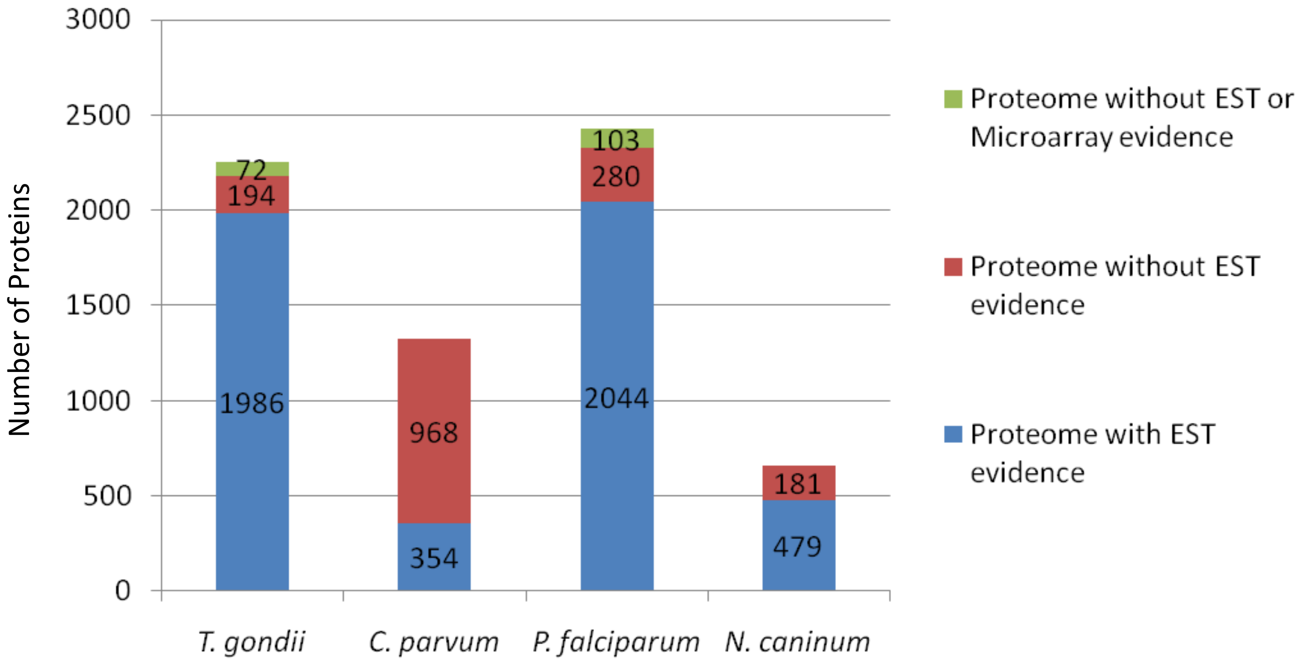
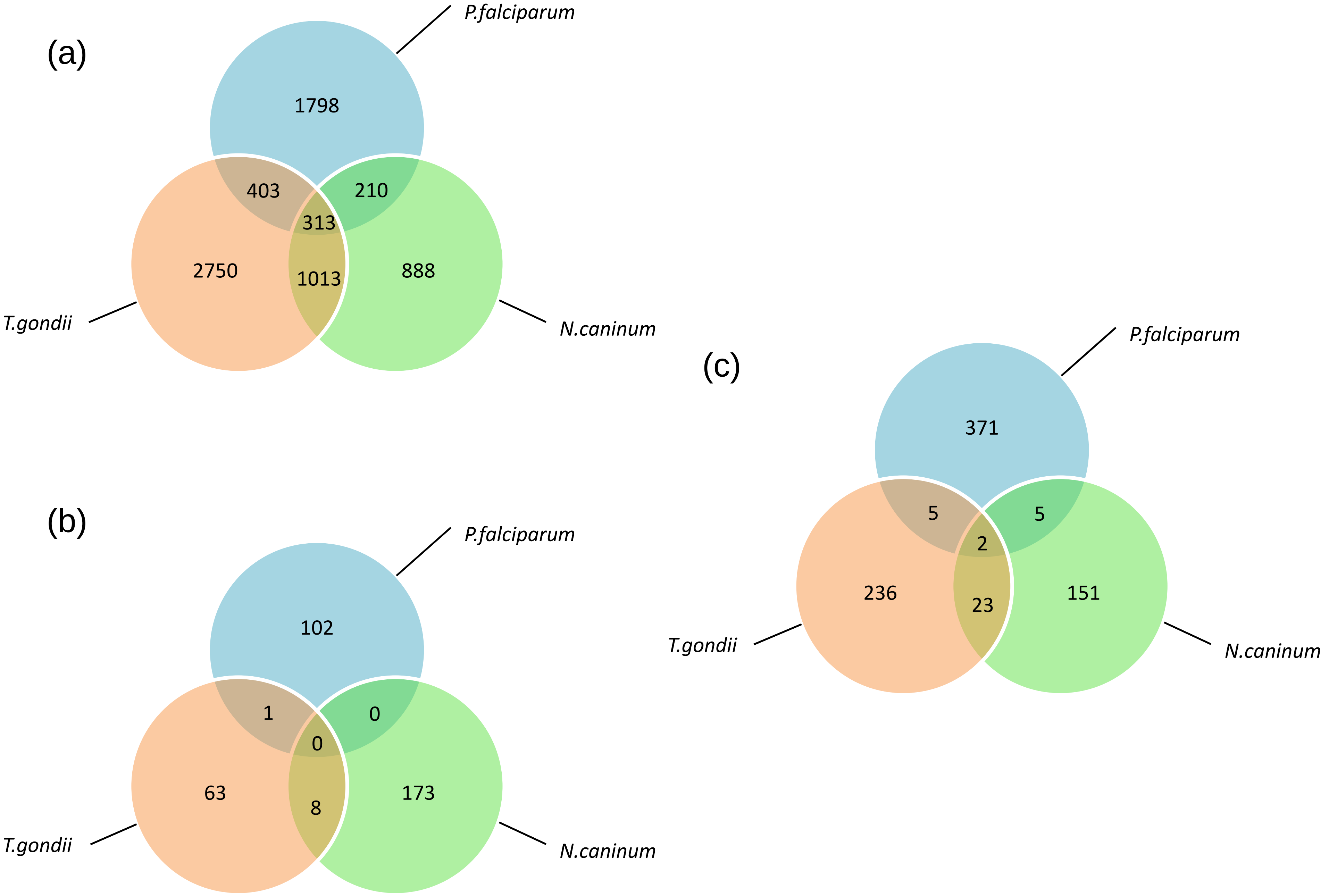

3000
2500
2000
1500
1000
500
0
Number of Proteins
72
194
1986
968
354
103
280
2044
181
479
T. gondii
C. parvum
P. falciparum
N. caninum
🟩 Proteome without EST or Microarray evidence
🟥 Proteome without EST evidence
🟦 Proteome with EST evidence

Figure 4. Proteome vs transcriptome cross apicomplexan parasites


(a)
P.falciparum
T.gondii
N.caninum
(b)
P.falciparum
T.gondii
N.caninum
(c)
P.falciparum
T.gondii
N.caninum