Genomics and Genetic Manipulation of Protozoan Parasites Affecting Farm Animals

16

Carlos E. Suarez, Heba F. Alzan, and Brian M. Cooke

Abstract

In this chapter we present a brief but state-of-the-art account of the genomics and current gene manipulation methods that can be used to improve our understanding of the genetics and the biology of an arbitrary group of 17 protozoan parasites responsible for diseases that affect animals worldwide, including babesiosis, toxoplasmosis, theileriosis, cryptosporidiosis, eimeriosis, trypanosomiasis, and trichomoniasis. Complete genomes are available for all parasites discussed, except for *Besnoitia*, an apicomplexan parasite responsible for dermatitis and other disorders with high infection rates, but low mortality. Dramatic differences in genome sizes are evident among the group of parasites under study, consistent with the distinct dependency of parasitic lifestyle for each organism. In addition, linear regression analysis correlating the ratios of the number of genes per genome and genome size among all the selected protozoan parasites suggests a strong association between these two parameters, in alignment with the notion that smaller protozoan genomes are generally more compact than larger genomes. A brief description of the methods for genome manipulations, including transient and stable transfections and gene editing methods, is provided. These methods, required to understand gene

C.E. Suarez (✉)
Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, USA

Animal Disease Research Unit, ARS-USDA, Pullman, WA, USA
e-mail: ces@vetmed.wsu.edu

H.F. Alzan
Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, USA

Parasitology and Animal Diseases Department, National Research Center, Cairo, Egypt

B.M. Cooke
Department of Microbiology, Monash University, Clayton, VIC, Australia

413

function and for improving control measures, have been successfully developed so far in most parasites selected. Rapid progress of genomic and gene manipulation techniques will likely result in the constant emergence of novel integrated methods for the interrogation and modification of genomes, leading to our better understanding of parasite lifestyle and, ultimately, to the rational design of improved methods for the control of animal infectious diseases.

16.1   Introduction

Protozoan parasites remain as important infectious agents affecting animal and human health globally. Their often complex life cycles and relationship with their hosts present numerous challenges to our ability to develop improved methods of control. Thus, new, combined research strategies are urgently required to hasten the rate of discoveries. In this chapter, we focus on the application of genomics and genetic manipulation techniques as tools to improve our understanding of the biology of a selected group of typically neglected protozoan parasites of veterinary and medical importance (Table 16.1).

Ideally, and in addition to other measures, new and improved vaccines and novel drugs are needed to prevent or treat most of the burden of disease caused by this group of diverse protozoan pathogens. In particular, developing improved control using rational approaches requires an advanced level of understanding of the parasites' biology, their interactions with their hosts, and particularly for vaccine development and the mechanisms of protective immunity. Recent significant advances in our understanding of the biology of most protozoan parasites affecting farm animals began with the arrival of the "omics" era, including genomics, proteomics, transcriptomics, metabolomics, lipidomics, as well as other existing, or future, "omics."

The emergence of genomics, perhaps the initial "omics," and the provision of the first complete and annotated organism genomes permitted the identification of numerous species-specific genes, but it was quickly realized that genomic approaches alone were insufficient to understand gene function. The value and utility of genomic data however is greatly increased when complemented with additional approaches such as transcriptomics, proteomics, lipidomics, and metabolomics. Thus, the "omics" field is extremely dynamic, and it can be expected that progress will be accelerated as novel computer-aided strategies of data management and analysis are able to integrate the massive incoming data arriving from all of these diverse research fields. The simplified scheme shown in Fig. 16.1 depicts a model of analysis involving several different "omics" strategies. For example, genome annotation and re-annotation, an area related to genomics, usually depends on proteomic as well as transcriptomic data and so on. Additionally, distinct epigenetic markers in identical genomes can influence gene transcription and hence affect everything occurring downstream in the natural flow of information in a cell.

Metabolomics aims at integrating the flow of pathways and metabolites involved in cell function at a certain moment of the life of the cell, and lipidomics and

Table 16.1 Full genome, and "omic" studies currently performed on a selected group of protozoan parasites of veterinary importance

Organism	Genome size (Mbp)	Total gene number	"Omic" studies				References
			EST	T	P	M	
Babesia bovis	8.18	3781	+	+	+	–	de Vries et al. (2006), Al-Khedery and Allred (2006), Lau et al. (2007), Pedroni et al. (2013), Johnson et al. (2017), Mesplet et al. (2011), Rachinsky et al. (2008) and Laughery et al. (2009)
Babesia bigemina	13.84	5136	–	+	+	–	Suarez et al. (2003), Vichido et al. (2008) and Shompole et al. (1994)
Babesia microti	6.41	3560	–	+	+	–	Silva et al. (2016a, b)
Theileria parva	8.35	4167	+	+	+	–	Savadye (1999), Bishop et al. (2005), Kishima et al. (1995), Nene et al. (2004) and Sugimoto et al. (1989a, b, 1992)
Theileria orientalis	9.01	4058	+[a]	–	+	–	Hayashida et al. (2012)
Theileria equi	11.67	5397	+[a]	–	+	–	Silva et al. (2013)
Theileria annulata	8.36	3845	+[a]	+	+	–	Oura et al. (2001, 2006) and Witschi et al. (2013)
Toxoplasma gondii	61.90–64.52	8563–8920	+	+	+	+	Manger et al. (1998), Cleary et al. (2002), Li et al. (2003), Radke et al. (2005), Xia et al. (2008) and Zhou et al. (2015, 2016)
Neospora caninum	59.10	7266	+	+	+	+	Bruno et al. (2004), Reid et al. (2012), Lee et al. (2003, 2004, 2005), Shin et al. (2005) and Elsheikha et al. (2014)
Eimeria spp.	43.67–72.24	6037–10254	+	+	+	–	Aarthi et al. (2011), Min et al. (2005), Miska et al. (2008), Schwarz et al. (2010), Jiang et al. (2005), Lal et al. (2009) and Oakes et al. (2013)
Sarcocystis spp.	124.41–130.2	7089–7174	+	+[b]	+	–	Howe (2001) and Howe et al. (2005)
Cryptosporidium spp.	8.50–9.25	3886–3981	+	+	+	+	Strong and Nelson (2000), Mauzy et al. (2012), Zhang et al. (2012), Siddiki (2013), Siddiki and Wastling (2009), Snelling et al. (2007), Hublin et al. (2013) and Ng et al. (2012)

(continued)

Table 16.1 (continued)

Organism	Genome size (Mbp)	Total gene number	"Omic" studies				References
			EST	T	P	M	
Trypanosoma spp.	20.93–83.51	3397–12,050	+	+	+	+	Agüero et al. (2004), Cerqueira et al. (2005), El-Sayed et al. (1995), Verdun et al. (1998), Archer et al. (2011), Greif et al. (2013), Kolev et al. (2010), Li et al. (2016), Atwood et al. (2005), dos Santos et al. (2015), Paba et al. (2004), Panigrahi et al. (2009), Urbaniak et al. (2012), Barrett et al. (2010) and Creek et al. (2013, 2015)
Leishmania spp.	29.03–35.21	8135–9101	+	+	+	+	Gentil et al. (2007), Levick et al. (1996), Cantacessi et al. (2015), Dillon et al. (2015), Saxena et al. (2007), de Jesus et al. (2014), Drummelsmith et al. (2003), Kumar et al. (2014), Menezes et al. (2013), Singh and Dubey (2016), Veras and Bezerra de Menezes (2016), Arjmand et al. (2016), Kovarova (2016), Scheltema et al. (2010) and Westrop et al. (2015)
Acanthamoeba spp.	42.02–99.59		+	–	+	–	Kong et al. (2001), Caumo et al. (2014) and Gawryluk et al. (2014)
Tritrichomonas spp.-*Trichomonas vaginalis*	176.41	99009	+	+	+	–	Huang et al. (2009, 2012), Gould et al. (2013), Woehle et al. (2014), de Miguel et al. (2010), Benchimol et al. (2017) and Oyhenart and Breccia (2014)
Besnoitia spp.	–	–	–	–	–	–	

EST, expressed sequence tag library; **T**, transcriptome; **P**, proteome; **M**, metabolome

[a]EST data is sourced from: http://piroplasmadb.org/piro/app/record/dataset/DS_6889a51dab
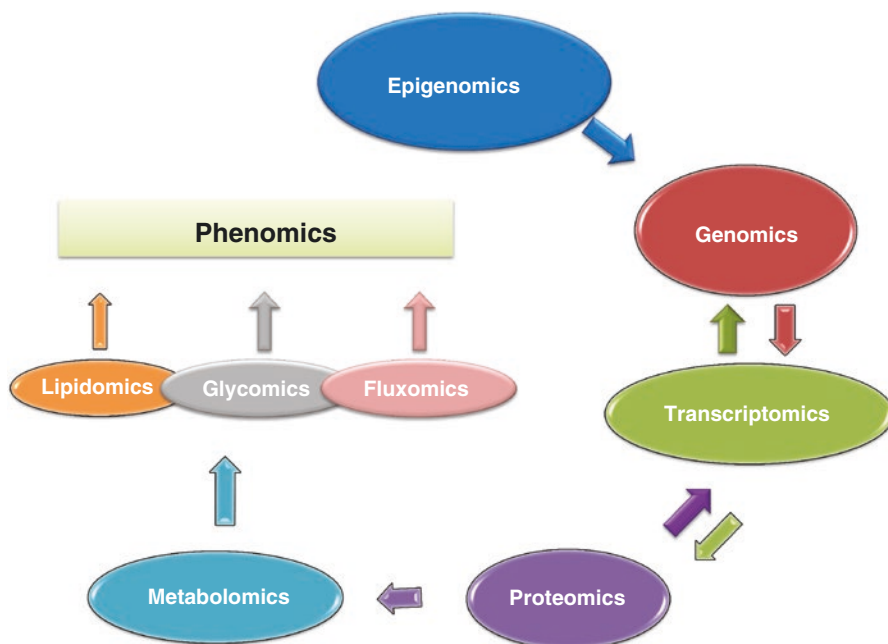
[b]Sourced from: http://www.ebi.ac.uk/

Fig. 16.1 Interrelationships among commonly used "omics" methods applied to the functional characterization of protozoan parasites

glycomics involve the study of pathways and networks of cellular lipids and sugars in biological systems, respectively, while fluxomics is aimed at determining the rates of metabolic reactions in biological systems. Finally, phenomics studies the set of physical and biochemical traits of a given organism as they respond to mutations and environmental changes. In fact, the relationships among the different "omics" are highly dynamic, and information may flow in any direction between them. In general, the combined use of these techniques used in "cross-sectional" studies can provide useful snapshots that may deliver insights into a parasite lifestyle, status, and survival strategies. In any case, comparing integrated "omics" profiles of different stages of a parasite cycle can provide useful information on the lifestyle of any unicellular organism, as is the case here for protozoa. Additionally, comparison of virulent and attenuated parasite strain/lines using "omics" approaches may also provide revealing insights into regulatory and metabolic networks and can be useful for the identification of virulence factors.

16.2 Genomics and Beyond

Genomics greatly facilitated the development of methods of genetic manipulation of these protozoan parasites, besides helping to provide a general blueprint of the biology of the organisms. Important progress has been so far achieved in genome-wide,

transcriptomic, and proteomic analysis and genetic manipulation on most of the protozoan parasites listed in Table 16.1 (1–94). However, a few have received relatively little attention, as in the case of *Besnoitia* where, for example, "omics" analysis and, predictably, gene manipulations remain unavailable. Genome size varies largely among this arbitrary selected group of highly diverse protozoan parasites (Fig. 16.2, Table 16.1), differing by a factor of more than 25 times between the smallest (*B. microti*, 6.4 Mbp) and the largest (*Tritrichomonas* spp., ~176.4 Mbp). In general, there appears to be an association between the size of an organism's genome and its dependency on intracellular parasite lifestyles (Sundberg and Pulkkinen 2015) with gene reduction being generally more drastic for obligate intracellular organisms that depend almost entirely on their host for survival. Finding out the possible associations between lifestyle and genome size and the evolutionary significance of genome size differences among this diverse collection of protozoan parasites would be of great interest. However, a better understanding of parasitic lifestyle also requires comprehensive, integrative, and comparative molecular, functional, metabolic studies, as well as improved knowledge of the parasite-host relationships. Regardless,
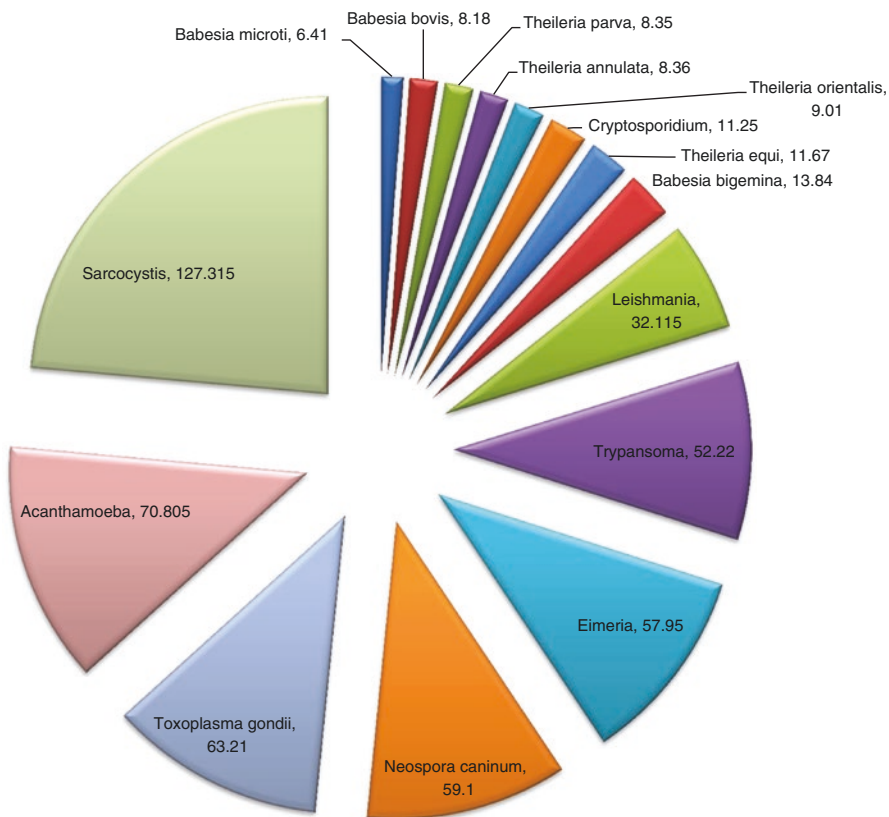

Fig. 16.2 Schematic comparison of the genome sizes of selected protozoan parasites

genomic comparative studies performed among related apicomplexans so far have resulted in a better, albeit somewhat limited, understanding of their biology (Blake 2015; Lv et al. 2015).

For instance, basic cellular and genomic research performed on *Toxoplasma gondii*, which is widely considered as a "model" apicomplexan parasite (Kim and Weiss 2004), is generally applicable to other related apicomplexans mainly because several key mechanisms, such as apical subcellular organelle formation and function, apicoplast and mitochondria function, signaling, gliding motility, intracellular molecule trafficking, cell invasion, etc., are overall well conserved among most of them (Kim and Weiss 2004; Ngô et al. 2004).

Linear regression analysis correlating the ratios of the number of genes per genome (represented as gene density) and genome size among all the protozoan parasites in Table 16.1, with the exception of *Tritrichomonas*, gives a significant linear negative correlation, with an r^2 coefficient of 0.85, suggesting a strong association between these two parameters (Fig. 16.3). This information is consistent with the notion that smaller protozoan genomes are generally more compact than larger genomes, containing more genes per megabase of DNA, likely as a result of having overall similar average gene sizes, but less repeated/redundant regions and less and/or shorter introns and noncoding intergenic regions. *Tritrichomonas* was not included in these comparisons as it has a highly atypical large and vastly
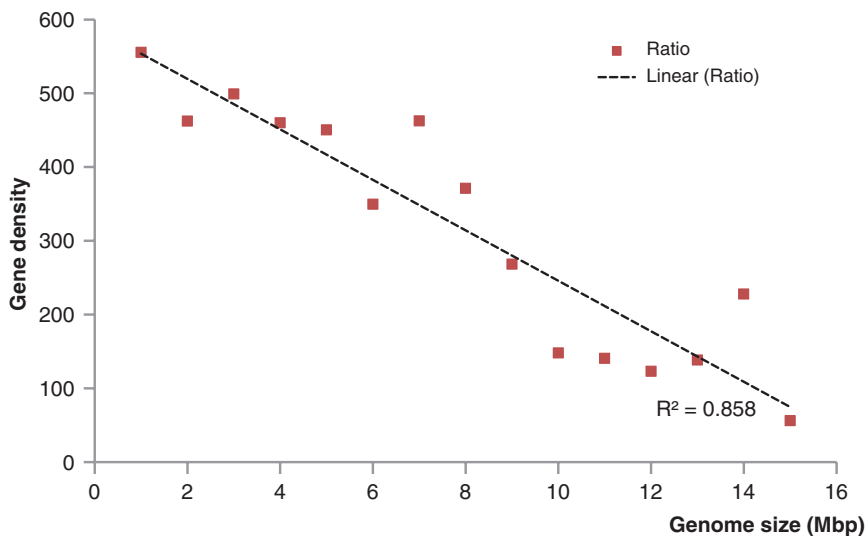

Fig. 16.3 Representation of the relationship between gene density and genome size for selected protozoan parasites. Gene density (genes/Mbp) was calculated by dividing total gene number by genome size (in Mpb) for each parasite species. The "X" axis is organized in ascending order, according to genome size: *1. Babesia microti, 2. Babesia bovis, 3. Theileria parva, 4. Theileria annulata, 5. Theileria orientalis, 6. Cryptosporidium* spp., *7. Theileria equi, 8. Babesia bigemina, 9. Leishmania* spp., *10. Trypanosoma* spp., *11. Eimeria* spp., *12. Neospora caninum, 13. Toxoplasma gondii, 14. Acanthamoeba* spp., *15. Sarcocystis* spp

repetitive genome. As shown in Fig. 16.2 and Table 16.1, coccidian parasites, such as *Toxoplasma* and *Neospora*, have a significantly larger genome size compared to the compacted genomes of the piroplasmid (*Babesia* and *Theileria*). An exception to this pattern is *Cryptosporidium* spp., a protozoan parasite with a genome size and gene density comparable to piroplasmid parasites. Interestingly, *Cryptosporidium* parasites lack a plastid, and no genes of plastid origins were identified, suggesting the early loss of the symbiotic apicoplast by these parasites (Abrahamsen et al. 2004; Xu et al. 2004), as well as the loss of the mitochondrial genome. Genome analysis on *Cryptosporidium* also revealed extremely streamlined metabolic pathways and an absence of many cellular structures and metabolic pathways found in other apicomplexans (Bouzid et al. 2013). Remarkably, *Plasmodium* parasites have a genome size that is intermediate between these two clades (~23 Mbp). Both piroplasmid and *Plasmodium* spp. share the existence of intraerythrocytic stages, but the *Plasmodium* sporozoites inoculated by the mosquito vectors are only capable of invading liver cells, and thus these parasites, in addition to an intrahepatic stage, have many other significant life cycle differences during the arthropod life stages. Interesting differences among piroplasma include the unique ability of *Babesia* parasites to transmit via transovarial mechanisms and the ability of *Theileria*, but not *Babesia* sporozoites, to invade and transform leucocytes of the mammalian host (Lau 2009). Clearly, each of these parasites faces distinct adaptive challenges, including their different strategies for causing persistent infections, requiring a unique and specific genome composition. Other remarkable differences among coccidian and piroplasmid parasites include the ability of some coccidians, such as *Toxoplasma*, to invade multiple distinct cell types in their vertebrate hosts and to form cysts. Once more, all of these phenotypic differences may account for the unique requirements in the number and quality of genes that can sustain the distinct parasitic lifestyles with different levels of complexity occurring in each species. However, and consistent with a more conservative value in the number of genes found for each of these organisms, genome size is also related to the sizes of the noncoding and intergenic areas in their genomes, as reflected in the ratios shown in Fig. 16.3. Again, a special case is the large genome of the highly related *Trichomonas* and *Tritrichomonas* parasites. *Tritrichomonas* are members of the eukaryotic supergroup Excavata, a group of free-living organisms that may or may not have a parasitic lifestyle. Their name is derived from the existence of an "excavated" ventral feeding groove. They are anaerobic parasites that lack classical mitochondria but instead contain specialized organelles, called hydrogenosomes, which are responsible for anaerobic metabolism. Consistent with its extracellular living status, *Trichomonas vaginalis* possesses a large genome, which is largely comprised by repeats and transposable elements (Carlton et al. 2007) and is shared by other related *Tritrichomonas* parasites of veterinary importance. A draft genome sequence of *T. foetus* showed that 72% of the open reading frames (ORFs) were found to be similar to those of *Trichomonas vaginalis* (Benchimol et al. 2017). In both parasites, the superabundance of repeats resulted in a highly fragmented sequence, preventing an investigation of genome architecture. The other 28% remaining ORFs have

no significant results with any other genome. The assembled genome of *T. foetus*, together with the functional annotation, is available at http://www.labinfo.lncc.br/index.php/tritrichomonas_foetus. Other study, using homology analysis, suggested that massive expansions might have occurred in the *T. foetus* genome in a similar way it was also predicted for *Trichomonas vaginalis*, while conservation assessment showed that duplications have been acquired after differentiation of the two species (Oyhenart and Breccia 2014). The authors of the former study concluded upon comparing the two genomes that gene duplications might be common among these parasitic protozoans (Oyhenart and Breccia 2014). In view of these findings, we included the genome of *T. vaginalis*, a human pathogen, together with *T. foetus* in Table 16.1. The high level of similarities among the genomes of *Trichomonas* and *Tritrichomonas* might simplify gene functional analysis using one of these organisms as a model.

Genomic and genetic manipulation studies performed on the "model apicomplexan" *Toxoplasma gondii* and *Babesia* sp. parasites, discussed below, exemplify the potential of these techniques toward improved parasite control. It is expected that the application of these approaches in other still poorly researched protozoa of veterinary importance will enhance our understanding of the biology of these parasites and their relationships with their hosts. It can be predicted that this new knowledge will translate into improved control of important yet neglected diseases with a high public health and economic impact globally. Certainly, more dramatic advances are expected to occur with the massive application of "omics" and vaccinology approaches in the near future.

16.3   Genomic Resources for Protozoan Parasites

Complete genomes of most protozoans of veterinary, medical, and zoonotic importance are currently available (Table 16.1). Furthermore, comparison between genome sequences among apicomplexan and other protozoans is now greatly facilitated using the Eukaryotic Pathogen Database Resource, EuPathDB (http://eupathdb.org/eupathdb/). This database provides access to the full genomes of *Babesia* spp. and *Theileria* spp. organized into the PiroplasmaDB (http://piroplasmadb.org/piro/show-Application.do), *Acanthamoeba* spp. genomes organized in the AmoebaDB (http://amoebadb.org/amoeba/showApplication.do), *Cryptosporidium* spp. at the CryptoDB (http://cryptodb.org/cryptodb/showApplication.do), and Coccidian genomes, including *Toxoplasma gondii*, *Neospora caninum*, *Sarcocystis neurona*, and *Eimeria* sp. at the ToxoDB (http://toxodb.org/toxo/showApplication.do). The *Trichomonas vaginalis* genome sequence is at the TrichDB (http://trichdb.org/trichdb/showApplication.do); and *Leishmania* spp. and *Trypanosoma* spp. can be found at the TriTrypDB (http://tritrypdb.org/tritrypdb/showApplication.do).

Importantly, the information on the EuPathDB is easily available and comprehensive and not limited to genome sequences. The site also provides easy access to analytical genomic tools such as Blast and available EST, microarray, RNA-seq, and proteomics data for these organisms, among other useful information.

16.4   Genetic Manipulation of Protozoan Parasites of Veterinary and Zoonotic Importance

Genetic manipulation techniques are important tools that allow access to multiple research approaches, including identification of virulence factors, subunit vaccine components, and parasite transmission factors. Importantly, genetically manipulated parasites themselves can be potentially used for vaccine development since targeted knockout of genes encoding known virulence factors might result in the production of genetically defined attenuated parasites. Another application of interest is the development of novel vaccine delivery platforms by manipulating attenuated parasites to express foreign genes coding for exogenous or stage-specific endogenous protective antigens. Also, genetically manipulated parasites used in vaccines can be easily distinguished from their wild-type counterparts, facilitating the discrimination among vaccinated and naturally infected animals. In addition to classic gene manipulation using transfection or gene editing techniques, RNA interference (RNAi) methods are also tools for gene function characterization (Meissner et al. 2007). However, *Trypanosoma* spp. and *Leishmania* spp. parasites, as well as most apicomplexan parasites (including *Babesia* and *Plasmodium*), lack the enzymes required for this pathway, and RNAi is not generally regarded so far as a useful method of gene analysis for these parasites. The mechanisms leading to the loss of the RNAi genes in these organisms, with no recognizable traces of their past presence, are unknown, although chromosomal rearrangements may have contributed to their disappearance (Kolev et al. 2011). Interestingly, the genome sequence of *T. gondii* revealed the existence of Dicer, AGO, and RdRp homologues (Braun et al. 2010) that appear to have plant/fungal (Dicer/RdRp) and metazoan (AGO) signatures. Initial reports thus suggested that *T. gondii* is the only apicomplexan with a functional RNAi pathway (Kolev et al. 2011). However, reported experimental results on the activity of this pathway were not reproducible, and the occurrence of this mechanism in this parasite has been put into doubt, highlighting the need for more research. A recent report also described the use of RNAi techniques to inhibit *B. bovis* in vitro growth upon the targeting of three distinct genes (AbouLaila et al. 2016), but the possible mechanisms involved remain uncertain given the absence of canonical RNAi genes in *B. bovis*. In other protozoan parasites, such as *Trichomonas vaginalis*, the presence of a Dicer-like gene and two Argonaute genes suggests the existence of the RNAi pathway (Carlton et al. 2007). Identification of these components raises the possibility of using RNAi technology to manipulate *T. vaginalis* gene expression.

The most widely used genetic manipulation methods include classic transfections based on the insertion of exogenous DNA using homologous recombination mechanisms (de Koning-Ward et al. 2000) and more recently CRISP/CAS9 (Lander 2016; Wright et al. 2016) and other gene editing methods such as TALENs and zinc-finger nucleases, based on programmable nucleases (Ma and Liu 2015). Such methods have been extensively used for the genetic manipulation of apicomplexan and other protozoa as will be described below.

16.4.1 Classic Transfection Methods

In transfection, DNA (or RNA) molecules may be introduced either as extrachromosomal replicating episomes or inserted into chromosomes by homologous recombination. Stably transfected lines can then be used for multiple applications including the study of gene function and by creating parasite lines that either overexpress or lack specific genes of interest. However, a limitation of reverse genetic approaches for functional gene analysis is that essential genes may be impossible to knock out, since this will result in nonviable parasites. These limitations can now be overcome at least partially by using inducible promoter strategies, including the use of tetracycline-inducible promoters. By choice of appropriate 5′ and 3′ flanking regions in the transfecting plasmid DNA endogenous chromosome, genes can be targeted and deleted. Additionally, the method can be also used to create transfected parasites that may function as vaccine delivery systems. The study of genetically transformed parasite lines can provide important clues about gene function during the parasite life cycle.

Perhaps mirroring their importance as human pathogens, *Leishmania* and other trypanosomatid parasites were first targeted for genetic transformation using transfection methods (Bellofatto and Cross 1989; Cruz and Beverley 1990; Laban and Wirth 1989; Lee and Van der Ploeg 1990; Ten Asbroek et al. 1990). The first report of the genetic modification of an apicomplexan parasite was the description of a transient transfection method for *Toxoplasma gondii* (Soldati and Boothroyd 1993). This led shortly thereafter to the development of a method for the stable transfection of this organism (Kim et al. 1993). Transient transfection is regarded as a useful approach for finding appropriate electroporation settings and to identify and test the function and efficacy of regulatory elements (promoters and termination signals) mediating gene expression and regulation. Based in part on these findings, transient and stable transfection techniques were later also applied to some species of *Plasmodium* parasites (Goonewardene et al. 1993; Van Dijk et al. 1995).

Transient transfection methods are useful for characterizing and defining promoters and other regulatory factors and later became essential components of advanced gene engineering and editing techniques, such as those based on the CRISPR/Cas9. Briefly, transient transfection techniques (Fig. 16.4) are designed to introduce and express foreign DNA, usually in the form of a plasmid, into a nucleated cell in a non-stable manner. Thus, in transient transfection, the introduced plasmid nucleic acid does not integrate into the genome of the target cells, and the transfected genes will not be replicated. After being developed in *T. gondii* (Soldati and Boothroyd 1993), transient transfection methods were applied on *Babesia bovis* (Suarez and McElwain 2008, 2010; Suarez et al. 2004, 2006, 2007), *Eimeria mitis* (Qin et al. 2014), *Sarcocystis neurona* (Gaji et al. 2006), *Theileria parva* (De Goeyse et al. 2015), and *T. annulata* (Adamson et al. 2001). This approach proved to be useful for the definition of promoters in *B. bovis* and later in *B. bigemina* (Silva et al. 2016a, b) and settled the basis for the development of stable transfections systems for *T. gondii* (Kim et al. 1993), *Sarcocystis neurona* (Gaji et al. 2006), *Acanthamoeba castellanii* (Peng et al. 2005), and *B. bovis* (Suarez and McElwain
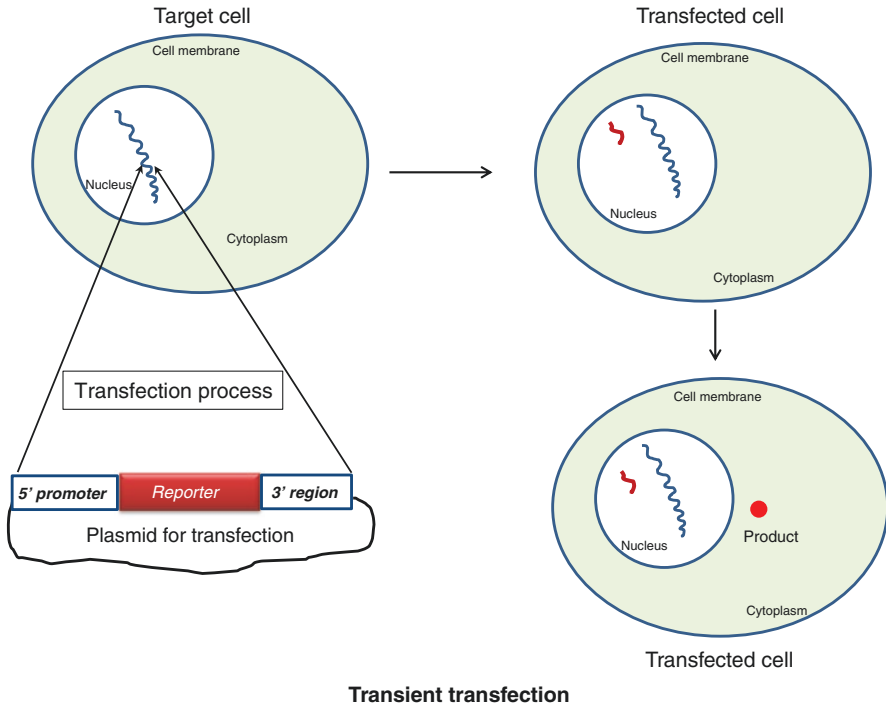
Fig. 16.4 Schematic representation of the principles and elements involved in transient transfection methods. A transfection plasmid (**a**) containing a reporter gene (*red box*) under the control of promoter and termination regions is transferred into the nucleus of a target parasite (**b**) using physical or chemical methods (transfection process). The plasmid DNA transferred into the target cell (in *red*) is not integrated stably into the genome (**c**), but it can be processed by the transcription and translation machinery of the cell to generate a product (*red dot*) (**d**) that can be quantified (e.g., by measuring luciferase activity)

2009, 2010). Transient transfection plasmids typically include a reporter gene or a gene that needs to be expressed transiently (such as required for current CRISPR/Cas9 methods), placed under the transcriptional control of a promoter and transcription and translation regions located at the 5′ and 3′ ends, respectively (Fig. 16.4). An appropriate amount of the transient transfection plasmid then is introduced into the target cells using distinct methods. These include those relying on physical treatments such as electroporation, nucleofection, biolistic delivery (gene gun), or microinjection and those relying on chemical entities, such as liposomes (Kepczynski and Róg 2016).

Physical methods create reversible "holes" in the cell membranes to insert the nucleic acids, whereas chemical methods are based on the use of transfection reagents, sometimes in the form of cationic lipids that allow membrane fusion and intracellular/intranuclear delivery of the foreign DNA into the target cells. More recently, nanoparticles and other polymers have been applied to this end. In general, the transiently transfected plasmids are designed so that they do not integrate into

the genome but remain as episomes in the target cell, where the gene of interest is expressed for a limited period of time. However, promoter strength studies and comparisons using transient transfection approaches are relative and limited, and so the data should be analyzed strictly in the context in which these experiments are performed. This is so because this approach is based on measuring promoter activity by promoter regions which are cloned in transiently transfected plasmids. This experimental approach would preclude estimating the possible regulatory role and contributions of distantly located or "trans" enhancers, the potential competition for transcription factors among the native and the plasmid-cloned promoters, as well as the possible contributions to promoter activity that depends on other regulatory elements such as epigenetic factors.

The stable transfection techniques are essentially based on the ability of the parasites to insert genetic material in their DNAs using homologous recombination mechanisms, in a fashion allowing expression of the transfected genes. Again, as for transient transfection, these techniques rely either on the use of liposomes such as Lipofectamine or on the application of a controlled electrical pulse, such as in electroporation, or later in nucleofection. These procedures allow the incorporation of exogenous DNA, usually provided in the form of a circular or linearized plasmid, into the nuclear compartment of a eukaryotic cell. Therefore, basic steps involved include (1) identification of a suitable genetic marker to select for transgenic parasites ("selectable marker"); (2) preparation of a transfection plasmid vector containing, at a minimum, a selectable marker gene under the control of a suitable promoter, and 5'and 3'regions to target integration of the construct into the genome; (3) a liposome or electroporation/nucleofection protocol which does not greatly compromise the viability of the target cells; and (4) a method for selection of transfected parasites. A schematic representation of a typical stable transfection vector is shown in Fig. 16.5.

Specific integration of the transfected gene(s) at the intended site into the genome of the target parasite depends upon the operation of homologous recombination mechanisms. However, this requirement may affect the efficiency of the integration process. In fact, the efficiency of the integration process in protozoan parasites is highly variable among species, and depends heavily on the available DNA repair mechanisms operating in each cell, as found for *Toxoplasma* parasites. However, the efficiency of exogenous gene integration can also be affected by the particular DNA base composition of the target cells, as in the case of the A + T-rich genome of *Plasmodium* parasites. Interestingly, *Toxoplasma* parasites are also difficult to engineer using classic transfection technologies because they have the ability to randomly insert the foreign DNA in sites different from the targeted. This occurs, at least in part, due to the action of the NHEJ repair mechanisms based on the activity of a gene encoding the KU80 protein. This limitation has been recently addressed by preparing a genetically transformed *Toxoplasma gondii* line lacking the *KU80* gene, which makes it amenable to gene targeting using homologous recombination mechanisms (Huynh and Carruthers 2009). In contrast, transfection work performed in *Babesia bovis* suggests that this technique is efficient, at least in terms of targeting, when applied to this organism (Suarez et al. 2015). Yet, differences in the gene
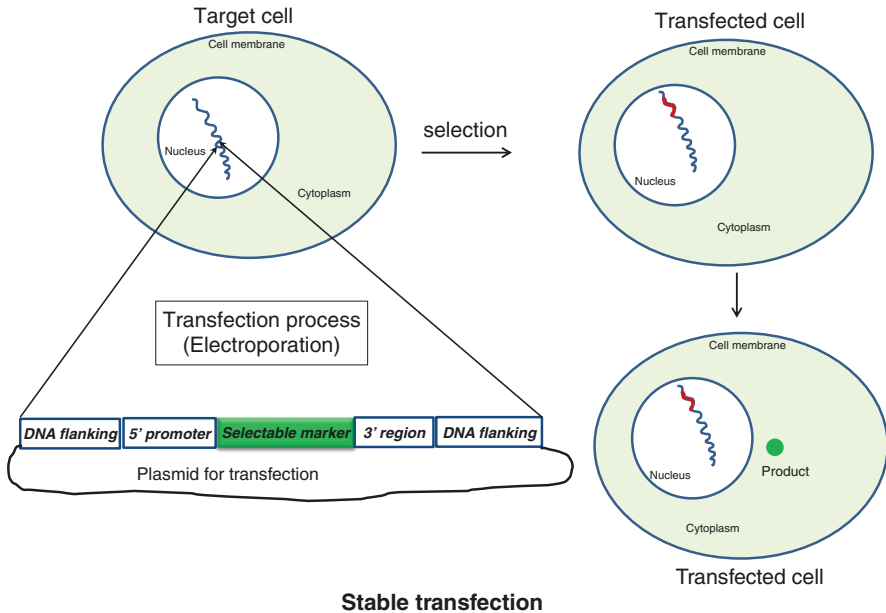
Fig. 16.5 Schematic representation of the principles and elements involved in stable transfection methods. A transfection plasmid containing a selectable drug resistance gene (*green box*) (such as bsd, dhfr, pyrimethamine, etc.) with a gene coding a fluorescent marker (such as GFP) under the control of promoter and termination regions, in addition to the 5′ and 3′ flanking regions required for homologous recombination, is transferred into the nucleus of a target parasite using physical or chemical methods. The incorporated plasmid DNA (in *red*) is integrated stably into the genome of the target cell by homologous recombination and processed by the transcription and translation machinery of the cell to generate a product (i.e., GFP-BSD) (*green dot*). Drug selection is performed to eliminate non-transfected parasites and to obtain a cell line of transfected parasites

repertoires and gene structure for other proteins involved in gene repair mechanisms among *B. bovis* and *T. gondii* together with differential regulation of their expression might also help explain the differences observed among the mechanism of gene repair operating in these two organisms. In contrast to *Plasmodium* and *Toxoplasma*, *B. bovis* appears to be quite amenable for stable transfection and, consistently, is able to specifically and efficiently integrate foreign genes. Thus, stable transfection techniques for *B. bovis* allowed highly specific KO and KO reversion experiments that are needed to study gene function (Asada et al. 2012a, b, 2015; Suarez et al. 2015). Recent progress in *Babesia* transfection technology includes the demonstration of a method for functional gene analysis by generating gene KO followed by gene function recovery (Asada et al. 2015) and the demonstration of cross-species promoter function (Silva et al. 2016a, b). This later study describes the ability of a *B. bovis* *ef-1α* promoter to function efficiently in *B. bigemina*. This observation suggests that common regulatory signals should exist, allowing the control of promoter functions among these two parasites.

16.4.2 Gene Editing Using Programmable Nucleases

Targeted genetic editing methods that allow precise modifications in a genome were more recently developed. These methods offer great potential for the manipulation of the genomes of *Toxoplasma*, *Plasmodium*, and other protozoan parasites, where transfection methods based solely on homologous recombination typically demonstrate very low efficiency. A key factor dramatically increasing efficiency of programmable nucleases is their ability to generate blunt double-strand breaks (DSB) in the target DNA of interest. The DSB results in the intervention of repairing systems of the cells, such as error-prone nonhomologous end joining (NHEJ) mechanisms, which can repair the break without the presence of donor homologous DNA. Alternatively, the breaks can be repaired by homology-directed repair (HDR) mechanisms in the presence of homologous donor double- or single-strand DNA, leading to the insertion of exogenous genetic material. The two mechanisms of DNA repair are exemplified in Fig. 16.5. The existence of these alternative pathways also suggests the possibility of using different gene manipulation strategies. Thus, introduction of simple mutations resulting in gene inactivation or disruption can be generated by double break followed by NHEJ. This repair mechanism can generate either insertions or deletions (indels) in the target gene resulting in frameshifts that disrupt the continuity of the open reading frame, usually leading to the knockout of the gene. If the objective is the insertion of foreign genes, such as reporter genes, it may then be necessary to add donor plasmid DNA containing the gene which is intended to be inserted with the addition of homologous flanking regions, to facilitate accurate targeting. In this case, the insertion of the foreign gene will likely be mediated by HDR repair mechanisms. Importantly, new discoveries on the mechanisms of DNA repair in apicomplexan parasites revealed the participation of certain proteins such as rad51 and KU80. As discussed earlier, targeted mutation of the *KU80* gene resulted in a *Toxoplasma gondii* mutant line that is more efficient for gene targeting, since it favors the KU80-indpendent HDR mechanism of repair and prevents random incorporation of transfected genes, which was commonly found in this parasite. This cell line is thus ideally suited for gene function analysis in *Toxoplasma* using homologous recombination KO approaches (Huynh and Carruthers 2009; Smolarz et al. 2014).

The specific design of gene editing experiments depends on the programming nuclease method of choice. The programmable gene editing methods currently available include the use of engineered proteins such as zinc-finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs) or RNA-guided engineered nucleases (RGEN). However, despite perceived improved target specificity of TALEN methods, the RGEN methods have several advantages over the other two, including their simple design, versatility, and cost. Briefly, the ZFN attach cutting domains derived from the prokaryote *Flavobacterium okeanokoites* to proteins called zinc fingers that can be customized to recognize certain three-base-pair DNA codes. On the other hand, TALENs fuse the same cutting domains to different proteins called TAL effectors. Both ZFN and TALENs require two cutting domains in order

to cleave double-stranded DNA. Excellent reviews on the use of ZFN and TALEN approaches for gene editing are available elsewhere (Ma and Liu 2015).

The most widely used RGEN method is based on the CRISPR/Cas9 system. Deeper coverage on the discovery and function of the CRISPR/Cas9 system was described elsewhere (Lander 2016; Wright et al. 2016). Briefly, this system is divided into three types based on the Cas proteins involved. Only the simpler type II system is used for gene editing and is essentially based in just a single effector Cas9 protein, although other putative effectors can now also be used. The principle of the method is illustrated in Fig. 16.6 and its applications for gene editing in Fig. 16.7. Briefly, the acronym CRISPR is derived from "clustered regularly interspaced short palindromic repeats," which together with the Cas ("CRISPR-associated" proteins) endonucleases, such as Cas9, are part of an adaptive immune system against phages of bacteria and archaebacteria (Wright et al. 2016).

This bacterial immune system provides RNA-mediated immunity against viruses and plasmids based on copying and specifically cleaving exogenous genetic materials. It was soon realized that, upon providing the necessary components to target cells, this system could be also manipulated to edit DNA in virtually any cell. Together, CRISPR and Cas9 are able to target and cut almost any DNA in vivo, and together with transfection techniques, they quickly became an important asset as efficient and specific tools for gene editing.
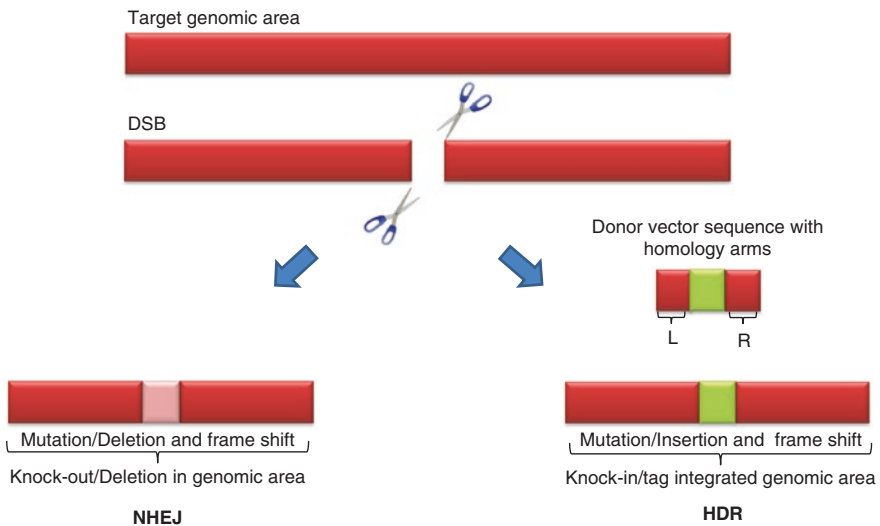

Fig. 16.6 Principles and elements involved in current gene editing methods. A targeted genome area is specifically cleaved with a double-strand break (DSB) by a nuclease and can be repaired using two different mechanisms: "nonhomologous end joining" (NHEJ) or "homology direct repair" (HDR) of the targeted cell. As a result, mutations, such as insertion/deletions (*pink box*), are caused, resulting in the inactivation of the target gene (NHEJ) or, in the presence of a donor sequence with homologous arms, the stable incorporation of new genetic material (*green box*) in the integrated genomic locus (HDR). L and R: left and right flanking homology arms

The most commonly used CRISPR/Cas9 systems are adapted from *Streptococcus pyogenes*. A CRISPR-Cas9 system specifically cleaves a DNA sequence through a two-stage recognition process, as depicted in Fig. 16.6. Initially, as more detailed below, a Cas9-sgRNA complex will be able to attach stably to a DNA sequence only if an appropriate, short (often only a few base pairs) protospacer adjacent motif (PAM) is located in close proximity (Fig. 16.7). Therefore, an important advantage of the CRISPR/Cas9 type II system is its simplicity, since only three components are required to achieve site-specific DNA recognition and cleavage. These include a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) which are required in order to guide the Cas9 enzyme to its target sequence (Fig. 16.7). These two elements (crRNA and tracrRNA) are usually combined into a single synthetic guide RNA (sgRNA). According to experimental design and specific gene targeting, the sgRNAs can be designed to include the specific base sequence that matches the target gene of interest. In that way, the complex can redirect the Cas9 enzyme to almost any preferred sequence. The *S. pyogenes* Cas9 endonuclease, which should bear nuclear localization signals (NLS), preferably requires an NGG PAM (with "N" representing any nucleobase followed by two guanine or "G" nucleobases). However, NAG and NGA PAM motifs can also sometimes be recognized. The 20 bp long sequence in the guide RNA then recognizes the homologous DNA target sequence by Watson-Crick base pairing. If a complete target sequence is confirmed, allosteric activation of the Cas9's two
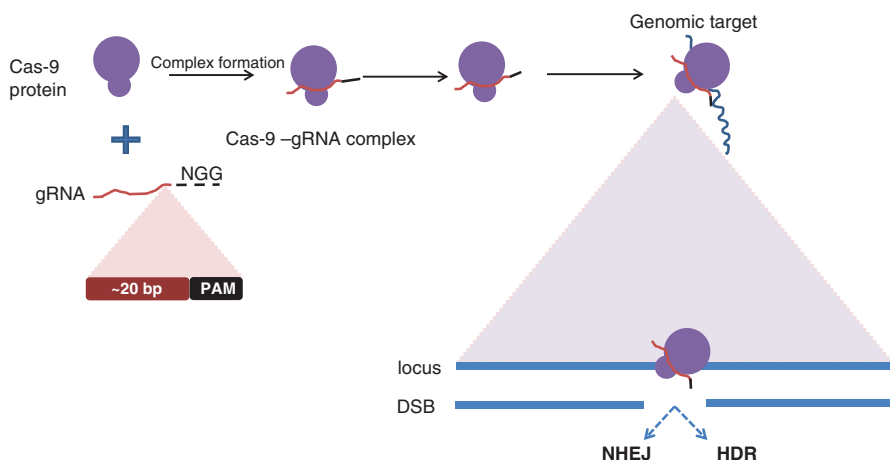

Fig. 16.7 Basic elements involved in gene editing methods based on CRISPR/Cas9. The 20 nucleotide guide RNA (gRNA) (represented in *red*) and PAM (*black box*) complexed with Cas9 is targeted to a specific sequence in the genome using Watson and Crick complementary base pairing. The complex locks into the targeted locus where it generates a double-strand break (DSB) caused by Cas9, which can be repaired using the "nonhomologous end joining" (NHEJ), or "homology direct repair" (HDR) mechanisms of the target cell. As a result, the targeted gene could be mutated or a new sequence can be integrated stably into the genome of the target cell. Gene-edited cells can later be selected using positive or negative selection procedures

nuclease domains, RuvC and HNH, will result in dual cleavage and, accordingly, a complete double-strand break in the target sequence. Clearly, the specificity of any CRISPR-Cas9 system depends heavily on the proper design of the guide RNA. This can be done sometimes using algorithms that minimize the likelihood of off-target effects. In other words, a CRISPR/Cas9 gene editing experiment requires the design of a 20 nucleotide guiding RNA (sgRNA) that can hybridize specifically with a sequence in the target gene. The sequences coding guiding RNA, the gene coding for Cas9 (including NLS), and donor DNA need to be provided to the target cells for expression in the form of plasmid DNA. Co-expression of these DNAs can be achieved following single vector or multiple vector strategies. Thus, for example, a single vector can include the genetic information necessary for the co-expression of Cas9, sgRNA, and donor DNA. This can be achieved using transient transfection of a properly engineered transfection plasmid having each gene under the transcriptional control of distinct promoters that need to be functional in the target parasite. This is now facilitated in *B. bovis* by the discovery that at least one heterologous *B. bigemina* promoter is also active in this parasite (Silva et al. 2016a, b). Alternative strategies include the delivery of in vitro transcribed sgRNA, as was the case for *T. cruzi* (Peng et al. 2015). Gene editing based on CRISPR/Cas9 has been used successfully for genetic analysis of several apicomplexan parasites of veterinary importance (Cui and Yu 2016), including *Cryptosporidium parvum* (Vinayak et al. 2015) and *Toxoplasma gondii* (Shen et al. 2014). In contrast to *T. gondii*, the low level of nonhomologous or random integration of exogenous transfected genes in *B. bovis* suggests that this parasite uses mainly HDR rather than NHEJ repair mechanisms. In addition, CRISPR/Cas9 approaches have also been used for genetic modification of trypanosomatid parasites such as *Trypanosoma cruzi*, *T. brucei*, and *Leishmania* spp. (Zhang and Matlashewski 2015). The versatility of *T. gondii* as a model apicomplexan was also employed for further development of a CRISPR/Cas9-based genome-wide genetic screen toward the identification of *T. gondii* essential genes during infection of human fibroblasts (Sidik et al. 2016). This approach allowed the description of an apicomplexan-conserved invasion factor termed claudin-like apicomplexan microneme protein (CLAMP). This novel approach has potential to be applied to other apicomplexan parasites.

Trypanosomatid parasites present the additional challenge of possessing a diploid genome; thus, deletion of an entire gene requires at least two distinct selection markers (Lander 2016). The use of these systems greatly accelerates our knowledge of the genetics of these parasites and the development of new vaccines. A website to guide the design of CRISPR tools in protozoan pathogens is a useful resource that is currently freely available (http://grna.ctegd.uga.edu/batch.html). Finally, different strategies for the selection of edited parasites, with or without the use of selectable markers, are also available (Mogollon et al. 2016).

Despite possible off-target cleavage and other potential limitations, gene editing procedures can be used for understanding gene function, generation of mutated attenuated parasites, or as a tool for the development of novel vaccines and therapeutics, thus improving the control of parasites of veterinary interest.

Conclusions

Combination of current "omics" and gene manipulation methods can improve dramatically our understanding of the genetics and the biology of protozoan parasites of veterinary and enzootic relevance. However, the rapid pace of progress of biotechnology, "omics," and other molecular and computer-aided tools will likely result in the constant emergence of novel integrated methods for the interrogation and modification of genomes, leading to a better understanding of parasite lifestyle, and, ultimately, to the rational design of improved methods for the control of animal infectious diseases.

References

Aarthi S, Raj GD, Raman M, Blake D, Subramaniam C, Tomley F. Expressed sequence tags from *Eimeria brunetti*—preliminary analysis and functional annotation. Parasitol Res. 2011; 108(4):1059–62.

AbouLaila M, Yokoyama N, Igarashi I. RNA interference (rnai) for some genes from *Babesia bovis*. RJAB. 2016;2:81–92. ISSN.2356:9433.

Abrahamsen MS, Templeton TJ, Enomoto S, Abrahante JE, Zhu G, Lancto CA, et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. Science. 2004;304(5669):441–5.

Adamson R, Lyons K, Sharrard M, Kinnaird J, Swan D, Graham S, et al. Transient transfection of *Theileria annulata*. Mol Biochem Parasitol. 2001;114(1):53–61.

Agüero F, Abdellah KB, Tekiel V, Sánchez DO, González A. Generation and analysis of expressed sequence tags from *Trypanosoma cruzi* trypomastigote and amastigote cDNA libraries. Mol Biochem Parasitol. 2004;136(2):221–5.

Al-Khedery B, Allred DR. Antigenic variation in Babesia bovis occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription. Mol Microbiol. 2006;59(2):402–14.

Archer SK, Inchaustegui D, Queiroz R, Clayton C. The cell cycle regulated transcriptome of *Trypanosoma brucei*. PLoS One. 2011;6(3):e18425.

Arjmand M, Madrakian A, Khalili G, Najafi A, Zamani Z, Akbari Z. Metabolomics-based study of logarithmic and stationary phases of promastigotes in *Leishmania* major by 1H NMR spectroscopy. Iran Biomed J. 2016;20(2):77.

Asada M, Tanaka M, Goto Y, Yokoyama N, Inoue N, Kawazu S-i. Stable expression of green fluorescent protein and targeted disruption of thioredoxin peroxidase-1 gene in *Babesia bovis* with the WR99210/dhfr selection system. Mol Biochem Parasitol. 2012a;181(2):162–70.

Asada M, Goto Y, Yahata K, Yokoyama N, Kawai S, Inoue N, et al. Gliding motility of *Babesia bovis* merozoites visualized by time-lapse video microscopy. PLoS One. 2012b;7(4): e35227.

Asada M, Yahata K, Hakimi H, Yokoyama N, Igarashi I, Kaneko O, et al. Transfection of *Babesia bovis* by double selection with WR99210 and blasticidin-S and its application for functional analysis of thioredoxin peroxidase-1. PLoS One. 2015;10(5):e0125993.

Atwood J, Weatherly D, Minning T, Bundy B, Cavola C, Opperdoes F, et al. The *Trypanosoma cruzi* proteome. Science. 2005;309(5733):473–6.

Barrett MP, Bakker BM, Breitling R. Metabolomic systems biology of trypanosomes. Parasitology. 2010;137(09):1285–90.

Bellofatto V, Cross GA. Expression of a bacterial gene in a trypanosomatid protozoan. Science. 1989;244(4909):1167–70.

Benchimol M, de Almeida LGP, Vasconcelos AT, et al. Draft genome sequence *of Tritrichomonas foetus* strain K. Genome Announc. 2017;5(16):e00195-17. https://doi.org/10.1128/genomeA. 00195-17.

Bishop R, Shah T, Pelle R, Hoyle D, Pearson T, Haines L, et al. 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. 2005;33(17):5503–11.

Blake DP. *Eimeria* genomics: where are we now and where are we going? Vet Parasitol. 2015;212(1):68–74.

Bouzid M, Hunter PR, Chalmers RM, Tyler KM. *Cryptosporidium* pathogenicity and virulence. Clin Microbiol Rev. 2013;26(1):115–34.

Braun L, Cannella D, Ortet P, Barakat M, Sautel CF, Kieffer S, et al. A complex small RNA repertoire is generated by a plant/fungal-like machinery and effected by a metazoan-like Argonaute in the single-cell human parasite *Toxoplasma gondii*. PLoS Pathog. 2010;6(5):e1000920.

Bruno S, Duschak VG, Ledesma B, Ferella M, Andersson B, Guarnera EA, et al. Identification and characterization of serine proteinase inhibitors from *Neospora caninum*. Mol Biochem Parasitol. 2004;136(1):101–7.

Cantacessi C, Dantas-Torres F, Nolan MJ, Otranto D. The past, present, and future of *Leishmania* genomics and transcriptomics. Trends Parasitol. 2015;31(3):100–8.

Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science. 2007;315(5809):207–12.

Caumo KS, Monteiro KM, Ott TR, Maschio VJ, Wagner G, Ferreira HB, et al. Proteomic profiling of the infective trophozoite stage of *Acanthamoeba polyphaga*. Acta Trop. 2014;140:166–72.

Cerqueira GC, DaRocha WD, Campos PC, Zouain CS, Teixeira SM. Analysis of expressed sequence tags from *Trypanosoma cruzi* amastigotes. Mem Inst Oswaldo Cruz. 2005;100(4):385–9.

Cleary MD, Singh U, Blader IJ, Brewer JL, Boothroyd JC. *Toxoplasma gondii* asexual development: identification of developmentally regulated genes and distinct patterns of gene expression. Eukaryot Cell. 2002;1(3):329–40.

Creek DJ, Nijagal B, Kim D-H, Rojas F, Matthews KR, Barrett MP. Metabolomics guides rational development of a simplified cell culture medium for drug screening against *Trypanosoma brucei*. Antimicrob Agents Chemother. 2013;57(6):2768–79.

Creek DJ, Mazet M, Achcar F, Anderson J, Kim D-H, Kamour R, et al. Probing the metabolic network in bloodstream-form *Trypanosoma brucei* using untargeted metabolomics with stable isotope labelled glucose. PLoS Pathog. 2015;11(3):e1004689.

Cruz A, Beverley SM. Gene replacement in parasitic protozoa. Nature. 1990;348(6297):171.

Cui Y, Yu L. Application of the CRISPR/Cas9 gene editing technique to research on functional genomes of parasites. Parasitol Int. 2016;65(6):641–4.

De Goeyse I, Jansen F, Madder M, Hayashida K, Berkvens D, Dobbelaere D, et al. Transfection of live, tick derived sporozoites of the protozoan Apicomplexan parasite *Theileria parva*. Vet Parasitol. 2015;208(3):238–41.

de Jesus JB, Mesquita-Rodrigues C, Cuervo P. Proteomics advances in the study of *Leishmania* parasites and leishmaniasis. Proteins and proteomics of *Leishmania* and *Trypanosoma*. Berlin: Springer; 2014. p. 323–49.

de Koning-Ward TF, Janse CJ, Waters AP. The development of genetic tools for dissecting the biology of malaria parasites. Annu Rev Microbiol. 2000;54(1):157–85.

de Miguel N, Lustig G, Twu O, Chattopadhyay A, Wohlschlegel JA, Johnson PJ. Proteome analysis of the surface of *Trichomonas vaginalis* reveals novel proteins and strain-dependent differential expression. Mol Cell Proteomics. 2010;9(7):1554–66.

de Vries E, Corton C, Harris B, Cornelissen AW, Berriman M. Expressed sequence tag (EST) analysis of the erythrocytic stages of *Babesia bovis*. Vet Parasitol. 2006;138(1):61–74.

Dillon LA, Okrah K, Hughitt VK, Suresh R, Li Y, Fernandes MC, et al. Transcriptomic profiling of gene expression and RNA processing during *Leishmania* major differentiation. Nucleic Acids Res. 2015;43:6799–813.

dos Santos ADCM, Kalume DE, Camargo R, Gómez-Mendoza DP, Correa JR, Charneau S, et al. Unveiling the *Trypanosoma cruzi* nuclear proteome. PLoS One. 2015;10(9):e0138667.

Drummelsmith J, Brochu V, Girard I, Messier N, Ouellette M. Proteome mapping of the protozoan parasite *Leishmania* and application to the study of drug targets and resistance mechanisms. Mol Cell Proteomics. 2003;2(3):146–55.

El-Sayed NM, Alarcon CM, Beck JC, Sheffield VC, Donelson JE. cDNA expressed sequence tags of *Trypanosoma brucei* rhodesiense provide new insights into the biology of the parasite. Mol Biochem Parasitol. 1995;73(1–2):75–90.

Elsheikha HM, Alkurashi M, Kong K, Zhu X-Q. Metabolic footprinting of extracellular metabolites of brain endothelium infected with *Neospora caninum in vitro*. BMC Res Notes. 2014;7(1):406.

Gaji RY, Zhang D, Breathnach CC, Vaishnava S, Striepen B, Howe DK. Molecular genetic transfection of the coccidian parasite *Sarcocystis neurona*. Mol Biochem Parasitol. 2006;150(1):1–9.

Gawryluk RM, Chisholm KA, Pinto DM, Gray MW. Compositional complexity of the mitochondrial proteome of a unicellular eukaryote (*Acanthamoeba castellanii*, supergroup Amoebozoa) rivals that of animals, fungi, and plants. J Proteome. 2014;109:400–16.

Gentil LG, Lasakosvitsch F, Silveira JFD, Santos MRMD, Barbiéri CL. Analysis and chromosomal mapping of *Leishmania* (*Leishmania*) amazonensis amastigote expressed sequence tags. Mem Inst Oswaldo Cruz. 2007;102(6):707–11.

Goonewardene R, Daily J, Kaslow D, Sullivan TJ, Duffy P, Carter R, et al. Transfection of the malaria parasite and expression of firefly luciferase. Proc Natl Acad Sci. 1993;90(11):5234–6.

Gould SB, Woehle C, Kusdian G, Landan G, Tachezy J, Zimorski V, et al. Deep sequencing of *Trichomonas vaginalis* during the early infection of vaginal epithelial cells and amoeboid transition. Int J Parasitol. 2013;43(9):707–19.

Greif G, De Leon MP, Lamolle G, Rodriguez M, Piñeyro D, Tavares-Marques LM, et al. Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*. BMC Genomics. 2013;14(1):149.

Hayashida K, Hara Y, Abe T, Yamasaki C, Toyoda A, Kosuge T, et al. Comparative genome analysis of three eukaryotic parasites with differing abilities to transform leukocytes reveals key mediators of *Theileria*-induced leukocyte transformation. MBio. 2012;3:e00204-212. https://doi.org/10.1128/mBio.00204-12. PubMed: 22951932. 2012.

Howe D. Initiation of a Sarcocystis neurona expressed sequence tag (EST) sequencing project: a preliminary report. Vet Parasitol. 2001;95(2):233–9.

Howe DK, Gaji RY, Mroz-Barrett M, Gubbels M-J, Striepen B, Stamper S. Sarcocystis neurona merozoites express a family of immunogenic surface antigens that are orthologues of the *Toxoplasma gondii* surface antigens (SAGs) and SAG-related sequences. Infect Immun. 2005;73(2):1023–33.

Huang K-Y, Chien K-Y, Lin Y-C, Hsu W-M, Fong I-K, Huang P-J, et al. A proteome reference map of *Trichomonas vaginalis*. Parasitol Res. 2009;104(4):927.

Huang K-Y, Huang P-J, F-M K, Lin R, Alderete JF, Tang P. Comparative transcriptomic and proteomic analyses of *Trichomonas vaginalis* following adherence to fibronectin. Infect Immun. 2012;80(11):3900–11.

Hublin JSN, Ryan U, Trengove R, Maker G. Metabolomic profiling of faecal extracts from *Cryptosporidium parvum* infection in experimental mouse models. PLoS One. 2013;8(10):e77803.

Huynh M-H, Carruthers VB. Tagging of endogenous genes in a *Toxoplasma gondii* strain lacking Ku80. Eukaryot Cell. 2009;8(4):530–9.

Jiang L-L, Huang B, Han H-Y, Zhao Q, Dong H, Chen Z. Comparison of the proteome of the sporutated oocysts of *Eimeria tenella* diclazuril sensitive strain with diclazuril resistant strain. Chin J Biotechnol. 2005;21(3):435–9.

Johnson WC, Taus NS, Reif KE, Bohaliga GA, Kappmeyer LS, Ueti MW. Analysis of stage specific protein expression during *Babesia bovis* development within female *Rhipicephalus microplus*. J Proteome Res. 2017. https://doi.org/10.1021/acs.jproteome.6b00947.

Kepczynski M, Róg T. Functionalized lipids and surfactants for specific applications. Biochim Biophys Acta Biomembr. 2016;1858(10):2362–79.

Kim K, Weiss LM. Toxoplasma gondii: the model apicomplexan. Int J Parasitol Parasites. 2004;34(3):423–32.

Kim K, Soldati D, Boothroyd JC. Gene replacement in *Toxoplasma gondii* with chloramphenicol acetyltransferase as selectable marker. Science. 1993;262:911.

Kishima M, Dolan T, Njamunggeh R, Nkonge C, Spooner P. Humoral immune responses to *Theileria parva* in cattle as measured by two-dimensional western blotting. Parasitol Res. 1995;81(4):334–42.

Kolev NG, Franklin JB, Carmi S, Shi H, Michaeli S, Tschudi C. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog. 2010;6(9):e1001090.

Kolev NG, Tschudi C, Ullu E. RNA interference in protozoan parasites: achievements and challenges. Eukaryot Cell. 2011;10(9):1156–63.

Kong H-H, Hwang M-Y, Kim H-K, Chung D-I. Expressed sequence tags (ESTs) analysis of *Acanthamoeba healyi*. Korean J Parasitol. 2001;39(2):151–60.

Kovarova J. Unravelling metabolism of *Leishmania* by metabolomics: University of Glasgow; 2016.

Kumar A, Misra P, Sisodia B, Shasany AK, Sundar S, Dube A. Mass spectrometry-based proteomic analysis of *Leishmania donovani* soluble proteins in Indian clinical isolate. Pathog Dis. 2014;70(1):84–7.

Laban A, Wirth DF. Transfection of *Leishmania enriettii* and expression of chloramphenicol acetyltransferase gene. Proc Natl Acad Sci. 1989;86(23):9119–23.

Lal K, Bromley E, Oakes R, Prieto JH, Sanderson SJ, Kurian D, et al. Proteomic comparison of four *Eimeria tenella* life-cycle stages: Unsporulated oocyst, sporulated oocyst, sporozoite and second-generation merozoite. Proteomics. 2009;9(19):4566–76.

Lander ES. The heroes of CRISPR. Cell. 2016;164(1):18–28.

Lau AO. An overview of the *Babesia*, *Plasmodium* and *Theileria* genomes: a comparative perspective. Mol Biochem Parasitol. 2009;164(1):1–8.

Lau AO, Tibbals DL, McElwain TF. *Babesia bovis*: the development of an expression oligonucleotide microarray. Exp Parasitol. 2007;117(1):93–8.

Laughery JM, Lau AO, White SN, Howell JM, Suarez CE. *Babesia bovis*: transcriptional analysis of rRNA gene unit expression. Exp Parasitol. 2009;123(1):45–50.

Lee MG-S, Van der Ploeg LH. Homologous recombination and stable transfection in the parasitic protozoan *Trypanosoma brucei*. Science. 1990;250(4987):1583–8.

Lee EG, Kim JH, Shin YS, Shin GW, Suh MD, Kim DY, et al. Establishment of a two-dimensional electrophoresis map for *Neospora caninum* tachyzoites by proteomics. Proteomics. 2003;3(12):2339–50.

Lee E-G, Kim J-H, Shin Y-S, Shin G-W, Kim Y-H, Kim G-S, et al. Two-dimensional gel electrophoresis and immunoblot analysis of *Neospora caninum* tachyzoites. J Vet Sci. 2004;5(2):139–46.

Lee E-G, Kim J-H, Shin Y-S, Shin G-W, Kim Y-R, Palaksha K, et al. Application of proteomics for comparison of proteome of *Neospora caninum* and *Toxoplasma gondii* tachyzoites. J Chromatogr B. 2005;815(1):305–14.

Levick MP, Blackwell JM, Connor V, Coulson RM, Miles A, Smith HE, et al. An expressed sequence tag analysis of a full-length, spliced-leader cDNA library from *Leishmania* major promastigotes. Mol Biochem Parasitol. 1996;76(1–2):345–8.

Li L, Brunk BP, Kissinger JC, Pape D, Tang K, Cole RH, et al. Gene discovery in the apicomplexa as revealed by EST sequencing and assembly of a comparative gene database. Genome Res. 2003;13(3):443–54.

Li Y, Shah-Simpson S, Okrah K, Belew AT, Choi J, Caradonna KL, et al. Transcriptome remodeling in *Trypanosoma cruzi* and human cells during intracellular infection. PLoS Pathog. 2016;12(4):e1005511.

Lv Z, Wu Z, Zhang L, Ji P, Cai Y, Luo S, et al. Genome mining offers a new starting point for parasitology research. Parasitol Res. 2015;114(2):399–409.

Ma D, Liu F. Genome editing and its applications in model organisms. Genomics Proteomics Bioinformatics. 2015;13(6):336–44.

Manger ID, Hehl A, Parmley S, Sibley LD, Marra M, Hillier L, et al. Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. Infect Immun. 1998;66(4):1632–7.

Mauzy MJ, Enomoto S, Lancto CA, Abrahamsen MS, Rutherford MS. The *Cryptosporidium parvum* transcriptome during *in vitro* development. PLoS One. 2012;7(3):e31715.

Meissner M, Agop-Nersesian C, Sullivan WJ. Molecular tools for analysis of gene function in parasitic microorganisms. Appl Microbiol Biotechnol. 2007;75(5):963–75.

Menezes JPBD, Almeida TFD, Petersen ALOA, Guedes CES, Mota M, Lima JGB, et al. Proteomic analysis reveals differentially expressed proteins in macrophages infected with *Leishmania amazonensis* or *Leishmania major*. Microb Infect. 2013;15(8):579–91.

Mesplet M, Palmer GH, Pedroni MJ, Echaide I, Florin-Christensen M, Schnittger L, et al. Genome-wide analysis of peptidase content and expression in a virulent and attenuated *Babesia bovis* strain pair. Mol Biochem Parasitol. 2011;179(2):111–3.

Min W, Lillehoj HS, Ashwell CM, Van Tassell CP, Dalloul RA, Matukumalli LK, et al. Expressed sequence tag analysis of *Eimeria*-stimulated intestinal intraepithelial lymphocytes in chickens. Mol Biotechnol. 2005;30(2):143–9.

Miska K, Fetterer R, Rosenberg G. Analysis of transcripts from intracellular stages of *Eimeria acervulina* using expressed sequence tags. J Parasitol. 2008;94(2):462–6.

Mogollon CM, van Pul FJ, Imai T, Ramesar J, Chevalley-Maurel S, de Roo GM, et al. Rapid generation of marker-free *P. falciparum* fluorescent reporter lines using modified CRISPR/Cas9 constructs and selection protocol. PLoS One. 2016;11(12):e0168362.

Nene V, Lee D, Kang'a S, Skilton R, Shah T, de Villiers E, et al. Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. Insect Biochem Mol Biol. 2004;34(10):1117–28.

Ng JSY, Ryan U, Trengove RD, Maker GL. Development of an untargeted metabolomics method for the analysis of human faecal samples using *Cryptosporidium*-infected samples. Mol Biochem Parasitol. 2012;185(2):145–50.

Ngô HM, Yang M, Joiner KA. Are rhoptries in Apicomplexan parasites secretory granules or secretory lysosomal granules? Mol Microbiol. 2004;52(6):1531–41.

Oakes RD, Kurian D, Bromley E, Ward C, Lal K, Blake DP, et al. The rhoptry proteome of *Eimeria tenella* sporozoites. Int J Parasitol. 2013;43(2):181–8.

Oura C, Tait A, Shiels B. *Theileria annulata*: identification, by differential mRNA display, of modulated host and parasite gene expression in cell lines that are competent or attenuated for differentiation to the merozoite. Exp Parasitol. 2001;98(1):10–9.

Oura CA, McKellar S, Swan DG, Okan E, Shiels BR. Infection of bovine cells by the protozoan parasite *Theileria annulata* modulates expression of the ISGylation system. Cell Microbiol. 2006;8(2):276–88.

Oyhenart J, Breccia JD. Evidence for repeated gene duplications in *Tritrichomonas foetus* supported by EST analysis and comparison with the *Trichomonas vaginalis* genome. Vet Parasitol. 2014;206(3–4):267–76.

Paba J, Santana JM, Teixeira AR, Fontes W, Sousa MV, Ricart CA. Proteomic analysis of the human pathogen *Trypanosoma cruzi*. Proteomics. 2004;4(4):1052–9.

Panigrahi AK, Ogata Y, Zíková A, Anupama A, Dalley RA, Acestor N, et al. A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome. Proteomics. 2009;9(2):434–50.

Pedroni MJ, Sondgeroth KS, Gallego-Lopez GM, Echaide I, Lau AO. Comparative transcriptome analysis of geographically distinct virulent and attenuated *Babesia bovis* strains reveals similar gene expression changes through attenuation. BMC Genomics. 2013;14(1):763.

Peng Z, Omaruddin R, Bateman E. Stable transfection of *Acanthamoeba castellanii*. Biochim Biophys Acta. 2005;1743(1):93–100.

Peng D, Kurup SP, Yao PY, Minning TA, Tarleton RL. CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. MBio. 2015;6(1):e02097–14.

Qin M, Liu XY, Tang XM, Suo JX, Tao GR, Suo X. Transfection of *Eimeria mitis* with yellow fluorescent protein as reporter and the endogenous development of the transgenic parasite. PLoS One. 2014;9(12):e114188.

Rachinsky A, Guerrero FD, Scoles GA. Proteomic profiling of *Rhipicephalus* (*Boophilus*) *microplus* midgut responses to infection with Babesia bovis. Vet Parasitol. 2008;152(3):294–313.

Radke JR, Behnke MS, Mackey AJ, Radke JB, Roos DS, White MW. The transcriptome of *Toxoplasma gondii*. BMC Biol. 2005;3(1):26.

Reid AJ, Vermont SJ, Cotton JA, Harris D, Hill-Cawthorne GA, Könen-Waisman S, et al. Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: Coccidia differing in host range and transmission strategy. PLoS Pathog. 2012;8(3):e1002567.

Savadye D. Establishment of an expressed sequence-tag database, mapping of *theileria parva* schizont cDNAs and molecular characterisation of a parasite-specific protein: University of Zimbabwe; 1999.

Saxena A, Lahav T, Holland N, Aggarwal G, Anupama A, Huang Y, et al. Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. Mol Biochem Parasitol. 2007;152(1):53–65.

Scheltema RA, Decuypere S, T'kindt R, Dujardin J-C, Coombs GH, Breitling R. The potential of metabolomics for *Leishmania* research in the post-genomics era. Parasitology. 2010;137(09):1291–302.

Schwarz RS, Fetterer RH, Rosenberg GH, Miska KB. Coccidian merozoite transcriptome analysis from *Eimeria maxima* in comparison to *Eimeria tenella* and *Eimeria acervulina*. J Parasitol. 2010;96(1):49–57.

Shen B, Brown KM, Lee TD, Sibley LD. Efficient gene disruption in diverse strains of *Toxoplasma gondii* using CRISPR/CAS9. MBio. 2014;5(3):e01114-14.

Shin Y-S, Shin G-W, Kim Y-R, Lee E-Y, Yang H-H, Palaksha K, et al. Comparison of proteome and antigenic proteome between two *Neospora caninum* isolates. Vet Parasitol. 2005;134(1):41–52.

Shompole S, McElwain TF, Jasmer DP, Hines SA, Katende J, Musoke AJ, et al. Identification of *Babesia bigemina* infected erythrocyte surface antigens containing epitopes conserved among strains. Parasite Immunol. 1994;16(3):119–27.

Siddiki A. Sporozoite proteome analysis of *Cryptosporidium parvum* by one-dimensional SDS-PAGE and liquid chromatography tandem mass spectrometry. J Vet Sci. 2013;14(2):107–14.

Siddiki A, Wastling JM. Charting the proteome of *Cryptosporidium parvum* sporozoites using sequence similarity-based BLAST searching. J Vet Sci. 2009;10(3):203–10.

Sidik SM, Huet D, Ganesan SM, Huynh MH, Wang T, Nasamu AS, Thiru P, Saeij JP, Carruthers VB, Niles JC, Lourido S. A genome-wide CRISPR screen in *Toxoplasma* identifies essential Apicomplexan genes. Cell. 2016;166(6):1423–1435.e12.

Silva MG, Graça T, Suarez CE, Knowles DP. Repertoire of *Theileria equi* immunodominant antigens bound by equine antibody. Mol Biochem Parasitol. 2013;188(2):109–15.

Silva JC, Cornillot E, McCracken C, Usmani-Brown S, Dwivedi A, Ifeonu OO, et al. Genome-wide diversity and gene expression profiling of *Babesia microti* isolates identify polymorphic genes that mediate host-pathogen interactions. Sci Rep. 2016a;6:35284.

Silva MG, Knowles DP, Suarez CE. Identification of interchangeable cross-species function of elongation factor-1 alpha promoters in *Babesia bigemina* and *Babesia bovis*. Parasit Vectors. 2016b;9(1):576.

Singh S, Dubey VK. Quantitative proteome analysis of *Leishmania donovani* under Spermidine starvation. PLoS One. 2016;11(4):e0154262.

Smolarz B, Wilczyński J, Nowakowska D. DNA repair mechanisms and *Toxoplasma gondii* infection. Arch Microbiol. 2014;196(1):1–8.

Snelling WJ, Lin Q, Moore JE, Millar BC, Tosini F, Pozio E, et al. Proteomics analysis and protein expression during sporozoite excystation of *Cryptosporidium parvum* (*Coccidia*, Apicomplexa). Mol Cell Proteomics. 2007;6(2):346–55.

Soldati D, Boothroyd JC. Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. Science. 1993;260:349.

Strong WB, Nelson RG. Preliminary profile of the *Cryptosporidium parvum* genome: an expressed sequence tag and genome survey sequence analysis. Mol Biochem Parasitol. 2000;107(1):1–32.

Suarez CE, McElwain TF. Transient transfection of purified *Babesia bovis* merozoites. Exp Parasitol. 2008;118(4):498–504.

Suarez CE, McElwain TF. Stable expression of a GFP-BSD fusion protein in *Babesia bovis* merozoites. Int J Parasitol. 2009;39(3):289–97.

Suarez CE, McElwain TF. Transfection systems for *Babesia bovis*: a review of methods for the transient and stable expression of exogenous genes. Vet Parasitol. 2010;167(2):205–15.

Suarez CE, Palmer GH, Florin-Christensen M, Hines SA, Hötzel I, McElwain TF. Organization, transcription, and expression of rhoptry associated protein genes in the *Babesia bigemina* rap-1 locus. Mol Biochem Parasitol. 2003;127(2):101–12.

Suarez CE, Palmer GH, LeRoith T, Florin-Christensen M, Crabb B, McElwain TF. Intergenic regions in the rhoptry associated protein-1 (rap-1) locus promote exogenous gene expression in *Babesia bovis*. Int J Parasitol Parasites. 2004;34(10):1177–84.

Suarez CE, Norimine J, Lacy P, McElwain TF. Characterization and gene expression of *Babesia bovis* elongation factor-1α. Int J Parasitol Parasites. 2006;36(8):965–73.

Suarez C, Lacy P, Laughery J, Gonzalez MG, McElwain T. Optimization of *Babesia bovis* transfection methods. Parassitologia. 2007;49:67.

Suarez CE, Johnson WC, Herndon DR, Laughery JM, Davis WC. Integration of a transfected gene into the genome of *Babesia bovis* occurs by legitimate homologous recombination mechanisms. Mol Biochem Parasitol. 2015;202(2):23–8.

Sugimoto C, Conrad PA, Mutharia L, Dolan T, Brown W, Goddeeris B, et al. Phenotypic characterization of *Theileria parva* schizonts by two-dimensional gel electrophoresis. Parasitol Res. 1989a;76(1):1–7.

Sugimoto C, Mutharia LM, Conrad PA, Dolan TT, Brown WC, Goddeeris BM, et al. Protein changes in bovine lymphoblastoid cells induced by infection with the intracellular parasite *Theileria parva*. Mol Biochem Parasitol. 1989b;37(2):159–69.

Sugimoto C, Mutharia L, Brown W, Pearson T, Dolan T, Conrad PA. Analysis of *Theileria parva* immunodominant schizont surface antigen by two-dimensional polyacrylamide gel electrophoresis and immunoblotting. Parasitol Res. 1992;78(1):82–5.

Sundberg L-R, Pulkkinen K. Genome size evolution in macroparasites. Int J Parasitol. 2015;45(5):285–8.

Ten Asbroek AL, Ouellette M, Borst P. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of *Trypanosoma brucei*. Nature. 1990;348:174–5.

Urbaniak MD, Guther MLS, Ferguson MA. Comparative SILAC proteomic analysis of *Trypanosoma brucei* bloodstream and procyclic lifecycle stages. PLoS One. 2012;7(5):e36619.

Van Dijk M, Waters A, Janse C. Stable transfection of malaria parasite blood stages. Science. 1995;268(5215):1358.

Veras PST, Bezerra de Menezes JP. Using proteomics to understand how *Leishmania* parasites survive inside the host and establish infection. Int J Mol Sci. 2016;17(8):1270.

Verdun RE, Di Paolo N, Urmenyi TP, Rondinelli E, Frasch AC, Sanchez DO. Gene discovery through expressed sequence tag sequencing in *Trypanosoma cruzi*. Infect Immun. 1998;66(11):5393–8.

Vichido R, Falcon A, Ramos JA, Alvarez A, Figueroa JV, Norimine J, et al. Expression analysis of heat shock protein 20 and Rhoptry-associated protein 1a in sexual stages and Kinetes of *Babesia bigemina*. Ann N Y Acad Sci. 2008;1149(1):136–40.

Vinayak S, Pawlowic MC, Sateriale A, Brooks CF, Studstill CJ, Bar-Peled Y, et al. Genetic modification of the diarrheal pathogen *Cryptosporidium parvum*. Nature. 2015;523(7561):477–80.

Westrop GD, Williams RA, Wang L, Zhang T, Watson DG, Silva AM, et al. Metabolomic analyses of *Leishmania* reveal multiple species differences and large differences in amino acid metabolism. PLoS One. 2015;10(9):e0136891.

Witschi M, Xia D, Sanderson S, Baumgartner M, Wastling J, Dobbelaere D. Proteomic analysis of the *Theileria annulata* schizont. Int J Parasitol Parasites. 2013;43(2):173–80.

Woehle C, Kusdian G, Radine C, Graur D, Landan G, Gould SB. The parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. BMC Genomics. 2014;15(1):906.

Wright AV, Nuñez JK, Doudna JA. Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell. 2016;164(1):29–44.

Xia D, Sanderson SJ, Jones AR, Prieto JH, Yates JR, Bromley E, et al. The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. Genome Biol. 2008;9(7):R116.

Xu P, Widmer G, Wang Y, Ozaki LS, Alves JM, Serrano MG, et al. The genome of *Cryptosporidium hominis*. Nature. 2004;431(7012):1107–12.

Zhang W-W, Matlashewski G. CRISPR-Cas9-mediated genome editing in *Leishmania donovani*. MBio. 2015;6(4):e00861-15.

Zhang H, Guo F, Zhou H, Zhu G. Transcriptome analysis reveals unique metabolic features in the *Cryptosporidium parvum* Oocysts associated with environmental survival and stresses. BMC Genomics. 2012;13(1):647.

Zhou C-X, Zhou D-H, Elsheikha HM, Liu G-X, Suo X, Zhu X-Q. Global metabolomic profiling of mice brains following experimental infection with the cyst-forming *Toxoplasma gondii*. PLoS One. 2015;10(10):e0139635.

Zhou C-X, Zhou D-H, Elsheikha HM, Zhao Y, Suo X, Zhu X-Q. Metabolomic profiling of mice serum during toxoplasmosis progression using liquid chromatography-mass spectrometry. Sci Rep. 2016;6:19557.