

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Publications from USDA-ARS / UNL Faculty

U.S. Department of Agriculture: Agricultural  
Research Service, Lincoln, Nebraska

---

2009

## From parasite genomes to one healthy world: Are we having fun yet?

Dante Zarlenga

USDA, ARS, ANRI, dante.zarlenga@ars.usda.gov

Louis C. Gasbarre

*Bovine Functional Genomics Laboratory*

Follow this and additional works at: <https://digitalcommons.unl.edu/usdaarsfacpub>

---

Zarlenga, Dante and Gasbarre, Louis C., "From parasite genomes to one healthy world: Are we having fun yet?" (2009). *Publications from USDA-ARS / UNL Faculty*. 2271.

<https://digitalcommons.unl.edu/usdaarsfacpub/2271>

This Article is brought to you for free and open access by the U.S. Department of Agriculture: Agricultural Research Service, Lincoln, Nebraska at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications from USDA-ARS / UNL Faculty by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



ELSEVIER

Contents lists available at ScienceDirect

# Veterinary Parasitology

journal homepage: [www.elsevier.com/locate/vetpar](http://www.elsevier.com/locate/vetpar)

## From parasite genomes to one healthy world: Are we having fun yet?

Dante S. Zarlenga<sup>a,\*</sup>, Louis C. Gasbarre<sup>b</sup><sup>a</sup> USDA, ARS, ANRI Animal Parasitic Diseases Laboratory, Bldg. 1180, BARC-East, Beltsville, MD 20705, USA<sup>b</sup> Bovine Functional Genomics Laboratory, Beltsville, MD 20705, USA

## ARTICLE INFO

## Keywords:

Parasite  
Genome  
Drug resistance  
Phylogenomics  
Host genetics  
Drug development

## ABSTRACT

In 1990, the Human Genome Sequencing Project was established. This laid the ground work for an explosion of sequence data that has since followed. As a result of this effort, the first complete genome of an animal, *Caenorhabditis elegans* was published in 1998. The sequence of *Drosophila melanogaster* was made available in March, 2000 and in the following year, working drafts of the human genome were generated with the completed sequence (92%) being released in 2003. Recent advancements and next-generation technologies have made sequencing common place and have infiltrated every aspect of biological research, including parasitology. To date, sequencing of 32 apicomplexa and 24 nematode genomes are either in progress or near completion, and over 600k nematode EST and 200k apicomplexa EST submissions fill the databases. However, the winds have shifted and efforts are now refocusing on how best to store, mine and apply these data to problem solving. Herein we tend not to summarize existing X-omics datasets or present new technological advances that promise future benefits. Rather, the information to follow condenses up-to-date-applications of existing technologies to problem solving as it relates to parasite research. Advancements in non-parasite systems are also presented with the proviso that applications to parasite research are in the making.

Published by Elsevier B.V.

### 1. Introduction

David Smith, the former Director of the US Department of Energy's Human Genome Program once commented:

*"Some years ago, the central idea or dogma in molecular biology research was that information in DNA directs RNA, and RNA directs proteins. Today, I think there is a new paradigm to guide us: Sequence implies structure, and structure implies function..If you really understand the relationship between sequence and function, you can begin to design sequences for particular purposes. . ."* (Evolution of a Vision: Genome Project Origins, available on the Internet [http://www.ornl.gov/sci/techresources/Human\\_Genome/publicat/97pr/evolve.html](http://www.ornl.gov/sci/techresources/Human_Genome/publicat/97pr/evolve.html)).

Such an insightful statement made in 1995 was predicated upon huge amounts of DNA sequence data becoming available, the creation of high-throughput systems for studying proteins, and the development of computer algorithms to assimilate this data to problem solving. With the advent of next-generation technologies, sequencing efforts once envisioned as time consuming and costly are now coming to fruition. As such, databases are swelling at startling rates which in turn have improved our abilities to address more complicated questions.

Some applications of large scale sequencing have straightforward outcomes such as developing gene-based diagnostic tests; however, less direct benefits with equally important downstream applications involve studies on evolution, the host–pathogen interface, and marker development such as parasite drug resistance and host susceptibility to infection. Understanding population variation and how diversity can impact the “one-size-fits-all” mentality in designing 21st century diagnostics, drugs and immunotherapies is also receiving more

\* Corresponding author. Tel.: +1 301 504 8754; fax: +1 301 504 8979.  
E-mail address: [Dante.Zarlenga@ars.usda.gov](mailto:Dante.Zarlenga@ars.usda.gov) (D.S. Zarlenga).

attention as more genome and transcriptome sequences become available. The importance of genetic variation in human research became evident, when in January 2008, an international research consortium announced the “1000 Genomes Project” (<http://www.1000genomes.org/files/1000Genomes-NewsRelease.pdf>). The goal of this endeavor was to sequence the genomes of at least a thousand people from diverse biological and geographical lineages then create a detailed picture of genetic variation within the human population. With the existence of a human scaffold sequence on to which the new data could be mapped, the feasibility of such a project became possible because of high through-put sequencing technologies. Unfortunately, such benefits are not easily adapted to parasite research inasmuch as the level of genetic variation expected between 1000 individual human samples is likely partitioned into any one parasite isolate or population. Also, the lack of sequence congruence among different parasite genera is sufficiently large to defy the concept of a single model organism for use in comparative genomics (Mitrevic et al., 2005). Thus the challenges facing those working on parasites as it relates to genetic variation are immense. Understanding this variability and its relationship to phenotype predictability will form a foundation for unilateral intervention strategies.

The following review looks at progress in applying parasite and host gene sequences to problem solving rather than summarizing sequence data and technologies that are currently available. Key research areas involve drug testing and design, evolution and population genetics, host–parasite relationships and epigenetics, and host resistance to parasites. Though applications of genomics to problem solving in parasitology have been slow to advance in areas other than DNA testing and diagnosis (to be discussed in other presentations in this series), they will in time become a requisite to understanding and creating a healthy world that has parasites as its nemesis.

## 2. Drug development

### 2.1. Integrating genomics and drug validation

In classic reductionist approaches to drug development, it is rarely evident at the outset which biological pathways will be lethal to a given pathogen if interrupted. Drug development nonetheless requires selecting an appropriate target molecule, designing drugs to impede its activity, and then using high through-put screening to validate that activity. Other, more random approaches, involve searching for functional activity or predefined endpoints from extracts of natural products, identifying the specific components affective against the pathogen, then synthesizing the compound and subsequent analogues “ad nauseam” to optimize the deleterious effects. In a perfect world, using phenotype-based screens in the validation process is optimal; however, these are not easily obtainable when working with parasites. Inevitably, animal models are required which mimic the pathogen in its natural host.

Among the first applications of whole-genome and next generation sequencing to drug validation, was defining the

mechanism of action of the new drug R207910 for treating *Mycobacterium tuberculosis* (Andries et al., 2005); the first tuberculosis-specific drug candidate developed in over 40 years. In this regard, drug-resistant mutants of *M. tuberculosis* and *M. smegmatis* were produced, subjected to massive parallel sequencing, and their genomes compared to the sequences of susceptible parental strains. Point mutations in the mutant strains identified consistent changes in the *atpE* gene in nearly all resistant mutants indicating that R207910 affects the proton pump of ATP synthase and that this protein was the most probable target for drug interaction.

Although the small genome size ( $4 \times 10^6$  bp) of *M. tuberculosis* lends itself to parallel sequencing, a similar approach was used to validate a new class of anthelmintics. Amino-acetonitrile derivatives (AAD) were recently shown to be effective alternatives to treating for parasitic nematodes, even those resistant to most commercially available drugs (Kaminsky et al., 2008). In this approach, the authors screened over 600 compounds for anthelmintic activity and performed preliminary tests for efficacy in a rodent model infected with *Haemonchus contortus*, followed by secondary studies in large animals. Positive results warranted a comprehensive program to identify the mechanisms of action; however, an investigation of this sort would have been costly if performed in the natural host. It was determined under drug treatment, that the free-living nematode *Caenorhabditis elegans* exhibited marked effects on several key physiological parameters similar to those observed in *H. contortus*, suggesting that AAD functional activity could be modeled in *C. elegans*; a benefit of the genome sequencing era. As with the discovery of R207910 for treating *M. tuberculosis*, drug-resistant mutants of *C. elegans* were generated and mapped by genetic recombination. Large scale sequencing identified *acr-23* as the target gene which encodes a nicotinic acetylcholine receptor- $\alpha$  (nAChR- $\alpha$ ) subunit coinciding with the nematode-specific DEG-3 group of nAChR, and which is distinct from the nAChR subunits of the levamisole receptor. Information from the *Haemonchus* genome project permitted transferring this technology to similar experiments with *H. contortus* AAD mutants. Results showed that the *Haemonchus* homologue (*des-2*) of *acr-23* was indeed part of the same nematode-specific DEG-3 group targeted by AAD in *C. elegans*.

Genomic, transcriptomic and proteomic applications to the discovery of drugs and vaccines in apicomplexa have been slow to evolve whether applied to human pathogens like *Plasmodium* or to zoonotic organisms like *Toxoplasma* and *Cryptosporidium*. Clearly, a solid foundation of information exists in that genome sequences, EST (expressed sequence tag) data and comparative studies are currently available from a number of apicomplexa (e.g. Abrahamsen et al., 2004; Carlton et al., 2008; Lau, 2009; Striepen and Kissinger, 2004) along with a complete complement of mining programs and websites. However, unlike comparative studies in nematodes which have used *C. elegans* as a model for discovery (e.g. Dieterich et al., 2008; McCarter, 2004; Mitrevic et al., 2004, 2005; Yin et al., 2008), drug discovery with apicomplexa is currently relegated to mining databases followed by transcriptional

analyses (Wu et al., 2003). This is due in part to the completion of many protozoan genome sequences only recently, the lack of biological, biochemical and genetic congruence among the important parasite groups, the inability of most of these organisms to survive in continuous culture, and the lack of an *in vivo* model system. However, the use of one protozoan, *Toxoplasma gondii*, seems to be breaking through this barrier at least in targeted circumstances (Boothroyd et al., 1995; Roos et al., 1999; Striepen and Kissinger, 2004; Kim and Weiss, 2004). Experimental and genetic manipulations of *T. gondii* are well suited to laboratory studies. In one example, *T. gondii* was used to validate a class of drugs (compound 1) for treating *Eimeria*. By generating genetic mutants similar to the validation of AADs, compound 1 was shown to specifically inhibit cyclic GMP-dependent protein kinase (Donald et al., 2002) in both *T. gondii* and *Eimeria*. *Toxoplasma gondii* was also used to study chemotherapeutic targets in *Cryptosporidium parvum* using genetic complementation and genomic insertion (Striepen et al., 2002). The authors demonstrated the ability to shuttle cloned *C. parvum* genomic DNA between bacterial plasmids and *T. gondii* and in so doing identified the inosine 5-monophosphate dehydrogenase homologue from *C. parvum*. This gene is involved in purine salvage that is essential for parasite development and therefore a likely target for drug intervention. Thus, analogous to *C. elegans*, the accessibility of *T. gondii* as an experimental system allows its use as a surrogate to ask many questions that cannot be easily handled in homologous systems (Kim and Weiss, 2004; Roos et al., 1999). In like manner, the utility of this model is targeted rather than holistic.

## 2.2. Identifying targets for drug or immune intervention

While a genomics approach was used to validate AADs and R207910 for treating gastrointestinal parasitic nematodes and *M. tuberculosis*, respectively, these applications were nonetheless downstream and subject to inefficiencies in the current paradigm for anthelmintic drug discovery. The most favorable drug targets are those proteins which are essential for the parasite, sufficiently conserved in sequence and function among focus organisms, and either do not exist or differ appreciably from host physiology. It is also important that the organisms do not have alternative pathways by which they can circumvent drug intervention. Pre-screening against these and other factors has yet to reach mainstream modern drug discovery. Thus, methods that systematically exploit sequence information are needed to identify putative target proteins earlier rather than later in the drug development process. This has become evident in human antiparasite drug discovery, where less than 1 in 50 of the “superior” drug candidates make it through clinical trials and to the market (Brown and Superti-Furga, 2003); the balance fail either in target identification or chemistry.

McCarter (2004) presented an example of how *C. elegans* and comparative genomics can be used to prescreen or “filter” EST data for candidate pathogen (nematode) protein targets. The first screening parameters limited target sequences to only those conserved among

Phylum members and lacked recognizable homologues among higher Metazoa. These “filters” alone reduced putative protein targets from 20,000 to 1200. After invoking additional filter criteria such as RNA interference (RNAi) phenotype, prioritization, and functional classification, the final list included among others, 3 proteins not present in mammals; (1) chitin synthase *chi-1*, which is predominantly associated with eggshell development; (2) fatty acid/retinol-binding proteins *far-1* and *far-3*, which scavenge host fatty acids and retinoids, and/or compromise the host immune response, and; (3) *fat-2*, a  $\Delta$ -12 fatty acid desaturase, which is essential at numerous steps in the cascade of nematode development. Prior studies (Stadler et al., 1994) had shown that naturally occurring fatty acids possessed nematocidal activity. Years later, Williams et al. (2005) demonstrated that the C-12 position in fatty acids with structures similar to oleic and linoleic acid was a key functional site. While the work by McCarter (2004) took place long after the first identification of nematocidal activity in fatty acids, it nonetheless validated genomic filtering as a viable approach to identifying drug targets.

A somewhat different approach was used to find putative drug targets in *Brugia malayi* (Kumar et al., 2007), where gene selection began not with ESTs but with RNAi orthologues in *C. elegans* that generate non-wild type phenotypes. The investigators reasoned that if these genes were necessary for *C. elegans* growth and development, orthologues would be similarly indispensable in *B. malayi*. Approximately 3000 *B. malayi* genes had RNAi orthologues in *C. elegans*; however, most of these shared commonality with the human dataset and were removed from consideration, leaving 589 first-pass candidates. Given that 589 genes comprised far too large a dataset to assay, additional ranking was done according to an algorithm defined by Hasan et al. (2006) where gene selection was rewarded for; (1) homology with *C. elegans* genes; (2) growth-essential characters defined by RNAi in *C. elegans*; (3) stage specific expression with “all stages” being most heavily weighted; (4) the presence of druggable protein domains (sequences that favor drug binding based on known drug targets) and; (5) expressability in soluble form in a bacterial expression system. Conversely, sequences were heavily penalized for having host (i.e. human) orthologues. Predictably, the top 40 genes using this mechanism of pre-screening included the chitin synthase and the family of fatty acid desaturases identified by McCarter (2004) and elsewhere (Williams et al., 2005), along with 10 other proteins of known or previously proposed targets. The molecular-target approach has been proposed for protozoan diseases as well (Frearson et al., 2007) where target validation and druggability are considered the most important parameters of a successful drug-discovery program.

Some mining tactics have become more focused and predicated upon conserved genes associated with metabolic functions in the nematode intestine; a key area for nutrient digestion and absorption. In this regard, Yin et al. (2008) examined 77% and 63% of the total genes available from the *Ascaris sum* and *H. contortus* databases, respectively. When compared to the *C. elegans* database, a subset

of 241 intestine-based genes (IntFam-241) was generated. These genes were conserved among all species and were defined as putative targets for effective parasite control. Caffrey et al. (2009) were interested in identifying alternative drug targets to treat *Schistosoma mansoni*. In this approach, the authors compared and contrasted the putative proteomes of *S. mansoni* with both *C. elegans* and the fruitfly *D. melanogaster* to identify orthologues for which disruption would yield deleterious phenotypes in *S. mansoni*. They used distinct selection criteria or workflows to generate 2 protein subsets and from among these selected genes that appeared in both subsets. Of the 13,283 predicted gene products of *S. mansoni*, 35 orthologous proteins were identified with druggable characteristics among which 3D structural information is available on 18 of these protein families.

Pharmacogenomics is a tactical diversion from the “one-size-fits-all” mentality in drug development and is currently being evaluated in human disease research. The common dogma of pharmacogenomics involves an appreciation for the relationship between genetic polymorphism and effective drug treatment, where predetermined knowledge of the mode of action can guide researchers towards those drug(s) most likely to prove efficacious. A second, less conventional component of pharmacogenomics is developing “designer” drugs directed at a subset of organisms responsible for the greatest disease, rather than targeting proteins that are conserved among the broadest group of organisms. To date, this approach has garnered interest only in abrogating *Plasmodium* species. The use of AT-specific DNA-reactive drugs first used by Cristofanilli et al. (1998) to treat metastatic breast cancer, were secondarily evaluated by Yanow et al. (2006) and Woynarowski et al. (2007) to target “AT-rich islands” in *P. falciparum*. In particular, 2 drugs bizelesin and adozelesin that alkylate 3' ends of adenines within these AT-rich islands were tested given that *P. falciparum* consists substantially (80%) of AT bases. Promising results in killing the parasite, coupled with disappointing results on host toxicity, have prompted continuing efforts to develop less toxic designer drugs predicated upon a holistic understanding of the genetic makeup of the pathogen (Yanow et al., 2008). The important component of this research is that this class of drugs targets nucleic acid primary structure rather than proteins. As we become better at predicting and validating secondary structure among nucleic acids, drugs causing damage to these molecules maybe less subject to parasite resistance and may find their way to the market place.

### 2.3. Limitations to parasite models in drug discovery

It is clear that *C. elegans* has become the poster child for nematode research. Its use as an “incubator” for studying pathways among nematodes and even among higher Metazoa is well established (Grant, 1992); however, using *C. elegans* to explore parasitism has been slow to unfold, exposing the current limitations to such a model. Classic functional genomics began with ground breaking work using double-stranded RNA (dsRNA) to interfere (RNAi) with translation through premature mRNA degradation. The technology was pioneered by Fire et al. (1998) and first

applied to study genome-wide regulation in *C. elegans* (Kamath and Ahringer, 2003; Ashrafi et al., 2003; Simmer et al., 2003). This work was predicated on the fully characterized genome and transcriptome of *C. elegans*, which provided fewer obstacles for culturing and evaluating phenotypic changes. In addition, numerous methods had been developed to introduce dsRNA into the nematode. It was assumed that this technology could be easily transferred to other nematode genera; however, RNAi studies on parasitic nematodes such as *Ostertagia* (Visser et al., 2006), *Haemonchus* (Geldhof et al., 2006) and *Heligmosomoides* (Lendner et al., 2008) among others, have met with sporadic and equivocal results, so much so that some have questioned whether the RNAi pathways are present in modified form or even exist in many parasitic nematodes (Knox et al., 2007; Viney and Thompson, 2008). Within currently available nematode EST databases, 35–70% of the genes in any one species overlap with that of *C. elegans*, the balance are conserved among higher taxa or to a large extent unique to specific taxa. Thus, in order to perform functional studies on the broad group of sequences that are unique or shared among limited taxa and not with *C. elegans*, other “model” systems will need to be developed that better represent subclasses of worms. In the absence of RNAi machinery comparable to that of *C. elegans*, it may become necessary to genetically engineer other nematodes to render them more amenable to RNAi studies (Viney and Thompson, 2008). Still, finding ways for high-throughput RNAi screening in parasitic nematodes will provide challenges unlike those observed in free-living organisms.

A second caveat to using models and molecular-targeting to mine nematode genomes or transcriptomes lies in applying universal constants to screen or filter datasets among disparate species. Namely, if a given subset of genes must be orthologous to those present in *C. elegans* as grounds for mining nematode sequences, such filtering criteria will inevitably produce undesired and predetermined commonality among the species subsets as the databases mature. Case in point, completely different approaches used to mine the *C. elegans* (McCarter, 2004) and *B. malayi* (Kumar et al., 2007) transcriptomes for drug targets generated substantial overlap in the “common genes” because of overlapping filtering criteria, undermining the intended independence of such searches.

### 3. Phylogenomics and “the tree of life”

In concert with the increasing repertoire of completed genome sequences, the field of phylogenetics has expanded to reconstructing evolutionary histories based upon their whole genomes, now coined “Phylogenomics” (Eisen, 1998; O'Brien and Stanyon, 1999; Eisen and Fraser, 2003). As with conventional molecular systematics, inference of evolutionary trees using genomic data is predicated on robust sequencing efforts, followed by careful assignment and assembly of homologous and/or orthologous genes (usually > 100). From here, phylogenomics can be further subdivided into sequence-based and non-sequence-based analyses.

Sequence-based phylogenomics relies on the data being concatenated into a supermatrix and analyzed as a single

entity. Alternatively, the individual genes can be analyzed separately to account for independent evolutionary histories, than combined into supertrees. These sequence-based approaches are similar to those commonly used today for individual genes but are applied to much larger and more diverse datasets.

To optimize the use of whole genomes for phylogenetic reconstruction beyond the level of primary sequence, non-sequence-based trees can also be inferred by comparing whole-genome features such as shared gene content (the fraction of shared orthologues between 2 genomes) (Snel et al., 1999), conservation of gene order (Korbel et al., 2002), “DNA strings” (the distribution and frequency of short-oligonucleotide or peptide combinations) (Delsuc et al., 2005; Philippe et al., 2005a), and intron positions (Roy and Gilbert, 2005). These methodological approaches do not rely on multiple-sequence alignments to infer phylogeny but look at the character make-up of the whole genome. One advantage to these methods is that the large character-state space reduces the risk of homoplasy by convergence and reversal (Philippe et al., 2005a) and therefore may result in more reliable phylogenies. Furthermore, sampling (stochastic) error, which is a caveat of phylogenetic trees based upon a few genes, should eventually be minimized as the datasets continue to increase (Rokas et al., 2003).

### 3.1. Parasite evolution

With parasites, phylogenomics is in its infancy owing to the paucity of completed genome sequences. Nonetheless, this technology has been used to evaluate seven apicomplexa using 268 single-copy genes (71,830 aligned amino acid sites) suitable for phylogenetic inference (Kuo et al., 2008). The data were analyzed using predicted protein sequences, which evolve slowly enough to retain phylogenetic signal in lineages that may have begun diverging from one another 700–900 MYA (Douzery et al., 2004). The authors demonstrated that the topologies of their trees were consistent with those generated from ultrastructural and developmental characters. Incongruence was common among the 48 individual trees used to construct the supertree, and only 19% of the 268 single-copy genes examined offered strong support for the apicomplexan species tree; the next 7–10% of the genes differed in the placement of *C. parvum*. The authors concluded that limiting the analysis to a small subset of sequences conserved among all apicomplexan lineages may have inadvertently excluded more distinct, yet phylogenetically informative, sequences. This has been claimed as a key factor in generating misleading tree topologies (Jeffroy et al., 2006; Philippe et al., 2005a). A subset of genes (19% or 51 genes) was nonetheless generated as a collection of candidate sequences for future phylogenetic analyses.

Phylogenomics was also used to examine lateral gene transfer (LGT) in apicomplexa, predominantly *C. parvum* (Huang et al., 2004a,b; Striepen et al., 2004). Results are consistent with plant-like genes present in apicomplexan parasites, such as *C. parvum*, that lack a plastid organelle i.e. the apicoplast. At least 31 genes of putative plastid/endosymbiont ( $n = 7$ ) or prokaryotic ( $n = 24$ ) origin have

been identified (Huang et al., 2004a) and suggest that *Cryptosporidium* may have arisen from a plastid-containing lineage, and thereafter lost its apicoplast. These studies explain in part why commonly used drugs against apicomplexa are unsuccessful in treating cryptosporidiosis. Lateral gene transfer from high-throughput sequencing has also been proposed for root-knot nematodes (Scholl et al., 2003) where rhizobia genes have been identified in *Meloidogyne* spp. using screening filters for phylogenetic incongruency. In general, identifying LGT can be quite difficult because of the inability to differentiate it from other evolutionary events, such as strong selection, unusual rates of change, and gene loss (Eisen and Fraser, 2003).

Defining hierarchical relationships among nematodes, arthropods and vertebrates is one of the most puzzling issues in animal systematics and evolutionary biology. The prevailing hypotheses are comprised of Coelomata or Ecdysozoa clades. The Coelomata hypothesis suggests that arthropods are more closely related to chordates than to nematodes, whereas the Ecdysozoa hypothesis contends that arthropods are genetically more closely linked to nematodes than to chordates. Varying the choice of gene(s) and protein(s) has been unable to resolve this issue because of problems using small datasets to evaluate deep evolutionary events. In 2004, Wolf et al. (2004) used phylogenomics to analyze over 500 protein sequences by generating a supermatrix from concatenated sequences and a supertree from optimized individual trees. They further analyzed their data using gene-content (presence/absence), domain co-occurrence, and indels, all of which quantitatively showed convergence on coelomate topology. In that same year, Dopazo et al. (2004) used gene-content to examine 25,000 amino acid sequences corresponding to >3000 genes. Similar to Wolf et al. (2004), their data unequivocally rejected the Ecdysozoa clade. Copley et al. (2004) examined 1712 orthologous genes and 2906 protein domains to study this same problem. Initially, their results strongly supported the Coelomata grouping as well; however, they did not account for substantial character loss inherent in *C. elegans*, which was used as the model nematode in this analysis. Once considered, Copley et al. (2004) found their data was more in line with the Ecdysozoa hypothesis. The following year, Dopazo and Dopazo (2005) found that contrary to their original analysis (Dopazo et al., 2004), a transition from Coelomata to Ecdysozoa clades occurred if fast-evolving sequences in *C. elegans* were removed from consideration. Dopazo and Dopazo (2005) further showed that gradual adjustments toward equal evolutionary rates between nematode and arthropod sequences shifted support from Coelomata to the Ecdysozoa hypothesis. This work was corroborated by Philippe et al. (2005b). These data demonstrated that long branch attraction biases can affect positioning of taxa on phylogenetic trees even when whole genome data was used.

### 3.2. Phylogenomics; panacea or puzzle?

Although large datasets presumably minimize sampling errors, systemic errors or compositional biases must still be accounted for in tree reconstruction (Jeffroy et al.,

2006). As example, whole genome data analyzed using maximum likelihood and parsimony methods (Rokas et al., 2003) resulted in a tree with 100% bootstrap support. When this same data was reanalyzed using minimum evolution, an alternative tree was generated that also received 100% bootstrap support (Phillips et al., 2004). Simply recoding the nucleotides as purines and pyrimidines (RY-coding) reinforced the original tree proposed by Rokas et al. (2003). Thus, caveats to these methodologies involve using probabilistic methods for analysis and phylogenetically misleading data (Jeffroy et al., 2006; Philippe et al., 2005a) where uncertainty in how best to model the evolution of genomes may limit the applicability of available reconstruction methods. Furthermore, bootstrap replicability merely testifies that a given method consistently reaches the same result when applied to a given dataset but does not guarantee the accuracy of that result to reflect the genome's evolutionary history. Inadequate-taxon sampling can also result in robust bootstrap values which do not reflect the true tree. One test for correct topology is the ability to generate congruent trees with both taxon-poor and taxon-rich sampling (Philippe et al., 2005a).

Clearly, using whole genome information in phylogenetic reconstruction is not yet the panacea originally envisioned in solving the tree of life. Problems still exist in how best to utilize the datasets and how to deal with biases that can “sway” the topology of a tree. However, tree topologies based upon whole genome sampling relative to conventional trees predicated upon one or a few genes are less affected by rare genomic changes such as unrecognized horizontal gene transfer, indels, misalignment, and to an extent, even missing data (Philippe et al., 2004). As we learn more about working with large datasets, defining reliable characters and the bases for inconsistencies, there is the hope that phylogenies will coalesce into the single true topology.

#### 4. Genomes, transcriptomes and population genetics

##### 4.1. Single nucleotide polymorphism and genetic variation

Twenty-five years ago, population genetics was relegated to comparing morphological and biological data among isolates. Today, scientists can use mass spectrometry and associated methodologies to perform 250,000 assays a day for single nucleotide polymorphisms (SNPs) and at an error rate of less than one in 10,000. This is particularly important when using large datasets to associate disease states with one or more common alleles. Population studies involving sequence variation at the level of whole genomes are scant among parasites primarily because reference sequences are required for comparative work and less than a handful of whole genome sequences are available. Hillier et al. (2008) used massive parallel, short read (35 bp) sequencing (Solexa) on *C. elegans* to reveal genome-wide variation in the form of single nucleotide polymorphisms (SNP) that can be applicable to evolutionary studies or correlating genetic markers to phenotypic characters. The availability of the whole genome of *Plasmodium falciparum* has led to

comparative studies on natural selection (Neafsey et al., 2008), barcoding for isolates identification (Daniels et al., 2008), and worldwide genetic diversity (Volkman et al., 2007). Unfortunately, genome-wide SNP studies with parasitic nematodes have been far less fruitful given the high genetic variation that exists among populations and the debatable need to evaluate the entire genome for phenotypic markers. As such, scientists have begun using genome subsets for population studies. Bekal et al. (2008) identified SNPs among 2 highly inbred biotypes of the soybean cyst nematode, *Heterodera glycines* that differed dramatically in their levels of virulence. This approach used incomplete genome sequence data for SNP identification by focusing on a single BAC clone as the reference sequence. This was followed by massive parallel sequencing of the virulent and avirulent strains using 454 microbead DNA technologies, and then comparing both datasets to the BAC sequence. Results demonstrated that 84% of the identified SNPs were correct. Van der Veer and de Vries (2004) tackled the problem of genetic variation by focusing on the mitochondrial genome of *Cooperia onchopora* which has only 525 bp of non-coding DNA. Unfortunately, their results demonstrated a rate of 1 polymorphic site every 32 bp. This far exceeds the average rate observed in the human nuclear genome (The International SNP Map Working Group, 2001) but does not vary substantially from that observed among nematode genomic sequences. Such variation renders identification of informative SNPs problematic even with sufficient taxon sampling.

As a result of the large genetic variability in non-coding DNA, studies using SNP markers have been relegated to individual genes or open reading frames i.e. ESTs. Phenotypic changes such as benzimidazole resistance have been associated with point mutations in the tubulin gene (Kwa et al., 1994; Von Samson-Himmelstjerna et al., 2007). Others have generated putative SNP markers assimilating Ivermectin resistance with P-glycoprotein (Bourguinat et al., 2008) and  $\beta$ -tubulin (Bourguinat et al., 2007), and for differentiating among *Cyclospora* and *Eimeria* species (Orlandi et al., 2003). Simões et al. (2007) used *S. mansoni* as both the target and model species to evaluate whole transcriptomes. They developed an automated pipeline consisting of 16,000 SNPs from 107k public ESTs sequences to detect SNPs in silico using high quality data and alignment parameters. Nonetheless, it is clear that until programs are developed to help weed through genetic “noise”, most SNP work will be relegated to less variable transcriptomes (Von Samson-Himmelstjerna and Blackhall, 2005; Gilleard, 2006). Efforts are being advanced in other fields to develop programs that can evaluate SNPs in highly pooled, genetically diverse DNA samples (Homer et al., 2008a,b). Targeting genes encoding highly expressed proteins may provide additional benefit. When Drummond et al. (2005) studied the rate of protein sequence evolution, they found that proteins evolve at rates inversely associated with their levels of expression and not in concert with protein function. Thus the most highly expressed proteins are believed to evolve more slowly and may provide some stability in SNP analyses if targeted. Others contend that functional constraints on protein evolution should not be neglected (Wolf et al., 2008).

#### 4.2. Microsatellites, genetic variation and anthropogenic effects

Marker types such as microsatellites (MSAT) are also being studied to circumvent the ambiguity associated with defining reliable SNPs. Microsatellites or Simple Sequence Repeats are short repeat sequences usually 2–8 base pairs in length most often found in non-coding DNA. These sequences are presumed not to undergo selection and are repeated in tandem > 8 times. Consequently, variation in the length of the repeats can be used to score these allelic loci. As with SNPs, many MSAT studies involve one or a few markers (Casulli et al., 2009; Grillo et al., 2006; Morlais et al., 2001) where identification is predicated upon extensive and specific cloning protocols followed by validation (Johnson et al., 2006). Webster et al. (2008) used large scale cloning to identify MSAT markers that address the disparity in benzimidazole resistance among nematodes infecting avian and livestock hosts. They postulated that the absence of resistance alleles in avian nematodes may result from alternative mechanisms in this system, or that selection pressures may not be the same between avian and livestock hosts.

As databases continue to expand, it will become more commonplace to mine genomes and EST for MSATs, thereby eliminating the extensive cloning steps (Rodrigues et al., 2002). Programs such as TRF (Benson, 1999), TROLL (Castelo et al., 2002), STRING (Parisi et al., 2003), Mreps (Kolpakov et al., 2003), and TRAP (Sobreira et al., 2006) have been used successfully for this purpose. The advantages of having genome sequence data can be seen in the recent work by Rosenthal et al. (2008) who scanned the near complete genome of *Trichinella spiralis* for MSAT markers. They identified and validated 9 targets that could be reproducibly amplified in all species and genotypes of *Trichinella* known to date. To understand the population structure of this genus, the authors evaluated hundreds of isolates spanning 28 countries and 4 continents using these MSAT markers. They concluded that other than for specific regions in Central Asia, *T. spiralis* found in pigs worldwide is basically the same population. They concluded that this species was likely dispersed by European colonists in only the past several hundred years as a consequence of host translocation and modern agricultural practices.

Similar techniques have been used in population studies of apicomplexa. Mining the *Plasmodium vivax* genome, Karunaweera et al. (2007) identified 14 MSATs that were used to study worldwide population diversity (Karunaweera et al., 2008). When they examined 164 isolates from 4 continents, they found 150 unique, multilocus haplotypes characteristic of extensive genetic diversity, with significant multilocus linkage disequilibrium in local populations; however, they saw no significant geographic clustering. They concluded that human infections followed by host relocation and parasite relapses, which are common in *P. vivax* infections, likely resulted in dissemination of strains across broad geographical localities and that this occurred very recently. They also suggested that the modest diversity among some local populations could lend itself to mapping loci associated

with virulence or drug resistance. Programs are being developed to accurately assess MSAT allele frequencies in pooled samples or in the case of parasites, to deal with the genetic variation that exists among individuals in a single population (Schnack et al., 2004; Silva et al., 2006).

Using a whole genome “fingerprinting technique” i.e. amplified fragment length polymorphism, which is independent of sequence data, Nejsun et al. (2005a) demonstrated a similar but more local effect in *Ascaris suum*. They found that *A. suum* populations among farms in Denmark were essentially homogeneous and indicative of free gene flow among geographical localities, or common recent origins similar to that observed with *T. spiralis*. The underlying concern is that once established, anthelmintic resistance is likely to spread rapidly among the Danish pig population. This is particularly intriguing in light of preliminary epidemiological evidence suggesting that Danish patients can acquire *Ascaris* infections from domestic pigs, raising concerns over *A. suum* as a zoonoses (Nejsun et al., 2005b).

#### 4.3. Indels as genetic markers; a new solution to an old problem?

New approaches are being developed to address high amounts of genetic variability by targeting small insertions and deletions (indels) rather than SNPs or microsatellites (Nairz et al., 2007). Indels are normally far less frequent than are SNPs. Studies in *C. elegans* found that indels constitute between 25% and 28% of all polymorphisms (Wicks et al., 2001; Swan et al., 2002) and that most of these are in the order of 1–2 bp. Simple analysis on Elchrom gels or capillary sequencers make this approach amenable to high-throughput screening if the application reaches mainstream (Nairz et al., 2007). Recently, a large scale analysis of 214,000 polypeptides from 32 nematode species identified insertions and deletions unique to >1000 families of nematode proteins (Wang et al., 2009). Among these, the authors identified three sizable deletions that are specific to parasitic nematodes and as such maybe targets for drug design.

### 5. Genomics and host–parasite relationships

#### 5.1. Genomics vs. epigenomics

With the discovery in China of the primitive bird Archaeopteryx, an intermediary link between avian evolution and dinosaurs was substantially confirmed (Asara et al., 2007; Schweitzer et al., 2007). Surprisingly, using a gene known to induce tooth formation, Harris et al. (2006) demonstrated that they could regenerate archosaurian teeth in chick embryos and alter the oral/aboral boundary of the jaw. Using similar methods, embryonic chickens produced tails with 16-vertebrae more representative of reptiles than the four to eight vertebrae present in spines of normal developing chickens (<http://www.dailymail.co.uk/sciencetech/article-1026340/>). Such captivating findings and particularly in the embryonic stages demonstrate that seemingly “lost” genes persist in the chicken genome but are subjected to changes in

developmental control. This raises intriguing questions linking epigenetics, or heritable non-genetic factors, to the way an organism's genes behave, where environmental forces (and not merely gene content) affect the phenotype of an organism. It further tempers what we can glean from whole genome sequences only. David Allis of Rockefeller University, New York, once commented;

*"The Human Genome Project... provided the blueprint for life, but the epigenome will tell us how this whole thing gets executed" (Bradbury, 2003).*

Epigenetics is the study of developmental regulation at the cell level where mechanisms such as histone modification, methylation, and gene silencing via small RNAs, regulate chromatin structure and cause genetically homogeneous cells to be phenotypically heterogeneous. Some have even suggested that histones and other regulators of chromatin structure may be viable targets for new antiparasite drugs (Rawat et al., 2007). Many epigenetic events are incorporated early on in embryogenesis though alterations are not limited to this period of development. Evidence for conservation of epigenetic pathways among Metazoa has been demonstrated in the nematode *C. elegans* (Lee et al., 2004, 2007; Simonet et al., 2007) and in the Protozoa *T. gondii* (Gissot et al., 2007; Gissot and Kim, 2008); however, modifications are not necessarily conserved among parasites in a population. Seemingly innocuous sequence variation between individual parasites can cause disparity in the way the genomic DNA and chromatin are recognized and modified. In the plant model, *Arabidopsis*, epigenomic studies showed how changes in the local sequence composition can affect DNA methylation and, in turn, the transcript abundance of hundreds of genes (Lister et al., 2008). Results further showed that modification was not inconsequential in that nearly 6% of all cytosines in the *Arabidopsis* genome were methylcytosine. Extrapolating this to parasites one can surmise that the manner in which a parasite responds to drugs or immunotherapy will be dictated not only by its genetic framework but by epigenetic factors as well. The relationship between epigenetics and drug resistance in parasites, however, has received little if any attention; yet, antiparasite drug treatments are known to affect transcriptome (Natalang et al., 2008; Silveira et al., 2007) and proteome (Choi et al., 2003; Prieto et al., 2008) profiles in the target organisms. Differences in gene expression need not be relegated to whole organisms. Recent work looking at proteome responses of "genetically identical" individual cancer cells following drug treatment clearly showed changes in expression levels and localization of proteins associated with the drug mechanism (Cohen et al., 2008). Among the 1000 proteins examined, a well-defined subset was differentially expressed in association with the outcome of the cell i.e. survival or death.

For a long time, changes in the ability of pathogens to infect different hosts was attributed to variations in the host immune response only; however, the proteomes of the pathogens can change in response to the host. Substantial incongruence was observed in the soluble proteins of *Heligmosomoides polygyrus bakeri* passed through C57/B10 mice that can house the parasite for

long periods (slow responders), relative to SWR mice that eliminate the parasite within a few weeks (fast responders) (Morgan et al., 2006). Using microarrays, similar variations in transcriptome profiles were observed in *Strongyloides ratti* subjected to rats exhibiting low or high immune pressures on the parasite. Collectively, these results underscore the difficulties in identifying vaccine targets against heterogeneous parasite populations, and demonstrate the affects that changes in the intestinal milieu can have on the transcriptomes of invading organisms.

## 5.2. Genomics and the vertebrate host

Having a well-defined host genome makes available a number of avenues by which the impact of parasites on their vertebrate hosts can be minimized. The recent completion of the genome sequence of several vertebrate species provides a framework for the development of innovative disease control programs. Upon completing the human genome sequence in 2003, focus turned to sequencing other species in the phylogenetic tree and to characterizing the diversity of these genomes. This resulted in the sequencing of the whole genomes of two agricultural importance species, i.e. the chicken (2004) and the cow (2006), and two other species of veterinary importance i.e. the dog (2004), and the horse (2007). In addition, the assembly of 10 chromosomes of the pig was released in 2008. These sequence data have provided a foundation for identifying inherent structural variations that influence host-parasite interactions, as well as for characterizing metabolic and immunologic pathways involved in resistance to infection and/or parasite-induced pathology.

There is overwhelming evidence that susceptibility/resistance to parasitic infection is influenced by host genetics. The importance of genetically controlled resistance has been demonstrated in a large number of protozoan infections such as leishmaniasis, trypanosomiasis, malaria, giardiasis, etc. (for review see Blackwell, 1988), and in helminth infections involving trematodes and nematodes (Whitlock, 1955; Gasbarre and Miller, 2000; Crawford et al., 2006; Coppieters et al., 2009). Arguably, the most extensively defined host-parasite systems are those involving gastrointestinal nematodes and their mammalian hosts (for review see Gasbarre and Miller, 2000). Numerous studies have evaluated the heritability of genetically controlled resistance, and have published estimates from below 10% to as high as 80%; however, most in the field would define this phenomenon as "moderately heritable" ranging between 20% and 40%. Part of the problem with accurately defining heritability is that parasite populations do not often exhibit a normal distribution among their hosts. Instead, they tend to exhibit an "overdispersed" distribution, i.e. a negative binomial distribution (Crofton, 1971a,b). As a result, traditional statistical methods for estimating heritability fail to accurately assess this trait because they are designed for use on data from infections which exhibit normal distribution patterns. The recent development of a new data transformation procedure appears to address this issue and may result in more accurate estimations of heritability (Silva, submitted for publication).

Regardless of the broad range of values obtained to date, it is evident that variations in the host genome can have a significant impact on the course of parasitic infections. A “Genomic Toolbox” will be a valuable resource in defining these variations, and in developing and applying new parasite control programs.

There are 3 major approaches that utilize genomics to address the parasite–host interface. Each method attempts to define the genetic make-up of the vertebrate host and understand how variation in the genome can influence the ability of the host to see and react to parasitic infections. With this knowledge, one can develop programs to reduce the rate or intensity of infection, and lessen the physiological impact of the parasite on the host. In order to demonstrate these 3 approaches we will concentrate on the relatively well-studied ruminant–gastrointestinal nematode system.

#### 5.2.1. Association of chromosomal regions and resistance

Success in identifying chromosomal regions influencing a given trait requires three important components; the first requirement is a working linkage map. As established previously, individual animals respond differently to parasites, and to a significant degree this disparity in responses is the result of genetic variation among hosts. Initial efforts to identify these genetic anomalies have focused on correlating specific physiologic responses to infection (phenotype) with distinct chromosomal regions of the host (genotype). In bovine, studies of this nature became possible following publication of the first linkage map (Barendse et al., 1994) and the subsequent refinement of this map by Kappes et al. (1997). This map resulted in the placement of more than 1200 polymorphic markers over the 29 bovine chromosomes which were separated by an average distance of 2.5 cM. While these distances are large, this information allowed identification of the first bovine quantitative trait loci (QTL), i.e. distinct regions of the genome that contain structural variations coincident with a physiological trait of interest. Most of these markers were microsatellites, which offered the advantage of numerous alleles, but had the disadvantage of being physically distant from one another within the genome; an estimated 30–40,000 total markers were separated by an average distance of 70–100,000 base pairs (bp).

The second requirement is a reliable measure that accurately reflects the physiological trait being investigated. In the case of GI nematodes, the trait is resistance to infection; the measure is the fecal egg count (FEC), representing the number of eggs passed in the feces. Analyses of FEC values in cattle herds revealed several important facts. First if calves are sampled on consecutive days, the repeatability of FEC values is approximately 0.6 (Gasbarre et al., 1996). This repeatability decreases dramatically if the time between samples is extended (Stear et al., 1984). Second, FEC values decrease with the age of the animal, and change throughout the year based on grazing behaviour of the animals. Third, the sex of the animal influences FEC values where bulls have higher FEC values than cows (Gasbarre et al., 1990; Stear et al., 1990). Finally, FEC values are not normally distributed within a cattle herd but follow an “overdispersed” distribution

(Crofton, 1971a,b; Genchi et al., 1989; Gasbarre et al., 1993). In such a distribution, where the standard error of the mean exceeds the mean, most individuals have relatively low fecal FEC values, and only a small percentage of animals are responsible for the majority of the eggs released on the pastures. This results in an apparent group of susceptible animals or “non-responders” representing only 15–25% of the total population (Anderson and May, 1985; Genchi et al., 1989).

The third and final requirement for QTL identification is a population of animals where the trait can be traced through a series of generations such that the phenotype (in this case FEC) can be related to the genotype (i.e. the microsatellite markers). This resource was produced at the Agricultural Research Service at Beltsville MD by selectively breeding Angus cattle for susceptibility or resistance to infection by GI nematodes. These cattle were initially derived from the Wye Angus herd of the University of Maryland. This herd was started with 18 registered yearling heifers and one bull; 10 of the heifers had the same sire, thus creating half-sisters. No other females were ever introduced into the herd. Between 1942 and 1958, an additional 19 bulls were imported from the British Isles (Lingle and Koch, 2001). As a result, a pedigree was produced comprising at least 9 generations that could be traced back to 1 of 18 dams and 1 of 19 bulls. All calves were placed at weaning on pastures containing *Ostertagia ostertagi* and *Cooperia oncophora*. The calves are kept on these pastures for a minimum of 120 days to allow them to express immunity to the parasite exposure (Gronvold et al., 1992). During the test period all animals were monitored weekly, and at the end of the test period some animals are slaughtered for parasitological and immunological assessment. Using this protocol, over 300 animals were completely characterized for resistance/susceptibility to the parasites.

An initial whole genome scan was run using 196 markers that were polymorphic in the population. The results indicated 17 loci on 7 different chromosomes with statistical correlation to traits of interest. Unfortunately, the extreme distances between the microsatellites made further efforts extraordinarily difficult. To lessen this problem, studies were initiated using SNP rather than microsatellite markers in the genome scans. A limitation of SNPs is that they exist in only 2 allelic forms, and as such are less informative than polymorphic microsatellites. In their favor, the number of SNPs in cattle is estimated to be between 7 and 14 million, with an average frequency of about 1 per 200–400 bp. Thus, large numbers of markers can be used in whole or partial genome scans and at sufficient densities to identify informative genome segments and permit re-sequencing of suspect areas. The recent development of the commercially available Infinium BovineSNP50 BeadChip<sup>®</sup> (Illumina, San Diego, CA) has provided us the tools to address this problem. This chip was developed as a cooperative effort of a private company and a number of government and University laboratories. It uses over 50,000 SNPs spread over the bovine genome to identify markers that can be associated with traits of interest. In addition to the advantage of characterizing smaller chromosomal segments, the chip

also reduces the level of necessary information from genotype, phenotype, and pedigree to simply knowing the genotype and phenotype of the animal. Therefore, data analysis changes from a complex algorithm that traces associations through generations to a simple Poisson association of trait and marker(s) (Gibbs and Singleton, 2006). To address the problem of only 2 alleles, adjacent SNPs can be strung together to yield SNP haplotypes that cover a region thereby yielding a polymorphic marker (Gabriel et al., 2002). Similar studies have been generated elsewhere using both microsatellites and SNPs to fine-map parasite resistance in Dutch Holstein–Friesian dairy cattle. These studies involved 4000 animals and genotypes from more than 750 within-family extremes (Coppieters et al., 2009). Their results revealed two chromosome segments with strong evidence for QTL, and these corresponded with orthologous positions in sheep discovered by Crawford et al. (2006) and Beh et al. (2002).

### 5.2.2. Identification of structural variations that influence resistance

The identification of chromosomal regions containing structural variations affecting traits of interest provides an opportunity to isolate and define specific genes associated with the phenotype. Structural variation can take on various forms, from a single base difference to large chromosomal rearrangements (Feuk et al., 2006). More traditional approaches have involved an examination of the identified area for likely gene candidates, eg. immune response genes. Once a subset of target genes were identified, variation in the genes and their regulating sequences in animals with differing phenotypes were determined by standard molecular techniques such as PCR. Unfortunately, such approaches are limited by *a priori* assumptions of the physiology of the trait. Additionally structural variations outside of the genes themselves will be missed, as will major structural anomalies such as copy number variation (CNV).

Newer genomic technologies offer a variety of methods to characterize important structural variations. The detection of large variations such as CNV can be addressed through a combination of specialized microarrays and next generation sequencing (Emanuel and Saitta, 2007). For smaller variations, a combination of QTL identification followed by resequencing of individuals with well defined phenotypes is a powerful way to accurately identify the actual structural variation that influences the trait of interest (Soria et al., 2005; Kunieda et al., 2005). The feasibility of this approach has increased dramatically with the advent of new, rapid and less costly means of sequencing. Modern short read sequencing technologies can provide several gigabases of data in less than 48 h at a fraction of the cost of the older platforms. As the speed and cost continue to decrease, the feasibility of this approach becomes limited only by our ability to deal with the enormous amounts of data.

### 5.2.3. Identification of metabolic and immunologic pathways that influence resistance and/or pathology

The final area where a “Genomic Toolbox” can be of great value is in the identification of host metabolic and

immunologic pathways induced during and after parasitic infections. Prior to genomics research, the discovery of important metabolic or immunologic pathways was similar to that used to identify structural variations, i.e. a targeted, candidate gene approach. In most cases this involved the development of gene-specific primer sets followed by differential RT-PCR of RNA samples from infected and non-infected hosts. While these procedures told us much about changes in expression profiles of a small number of targeted genes (Canals et al., 1997, 1998; Almeria et al., 2003), they were nonetheless limited in scope to only those genes or pathways deemed “important”. The advent of holistic methods to interrogate the entire transcriptome have allowed for experimental designs unencumbered by predetermined criteria and are meant to discover new or unique aspects of the parasite–host interaction. The most commonly used methodology today is the microarray. Microarray technology has been successful in the discovery of a number of interactions not previously explored. A recent report investigating the relationship between *Cooperia oncophora* and its bovine host shows the complexity and breadth of this interaction, and the need to look at pathways not previously explored (Li and Gasbarre, 2009). Mackinnon et al. (2009) recently used microarrays to evaluate gene expression patterns among parasite-resistant and susceptible sheep species infected with *H. contortus* and proposed that increased parasite resistance is associated with stronger inflammatory responses in animals infected with the parasite.

Although microarray technologies have the advantage of simultaneously evaluating changes in whole transcriptomes, it nonetheless has severe limitations including high rates of false positives and difficulties in accurate quantification (Cookson et al., 2009). The most promising approach to addressing these shortcomings is next-generation sequencing platforms that permit sequencing whole transcriptomes and at great depths (Shendure, 2008); however, this is not without its own challenges. The very power of this methodology is its own Achilles heel, i.e. massive amount of data. A single run on current sequencers generates 4–5 terabytes of data, and in the near future this figure will double. Thus, significant computer hardware and sophisticated bioinformatics support is required. As these problems are addressed, the technology will offer for the first time, an opportunity to accurately quantify all transcripts generated at specific points during the host–parasite interaction. As such we are entering an age where we can begin to clarify and understand the total complexity of the interactome.

### 5.3. Methods by which host genetics could be used to control parasitic infections

As an example of how the information derived from the use of a “Genomic Toolbox” could be used to control parasitic infections we can once again look at the heavily studied relationship between gastrointestinal nematodes and their ruminant hosts.

Among cattle breeds, adequate genetic variation is coincident with resistance to nematode infection, and the heritability of this trait is high enough that significant

genetic progress can be made by classic selection programs. The question is: “Do parasites exert enough economic pressure to warrant a selection program for this trait?” The answer is less clear and regionally dependent. For US producers in the Southeast, Northeast, and Northwest where stocking rates and parasite infections are sufficiently high, genetic improvement could have a profound affect on production efficiencies. In contrast, stockmen in areas of low grazing density may not see demonstrable improvement; however, they might still find an advantage to reducing parasites in animals based on marginal nutritional planes. In addition, as questions on land use and pesticide application become more important to consumers, any program that reduces dependence on anthelmintics will likely become more readily adopted by producers.

A more important issue relates to the ramifications of implementing breeding programs based on resistance to a single disease entity. Resistance to GI nematodes is a complex mixture of many types of immune responses and host protection against dissimilar parasites is unlikely to be congruent (Canals et al., 1997; Almeria et al., 1998; Stear et al., 2007). It is also becoming evident that distinct classes of immune responses, i.e. the so called Th1/Th2 paradigm, are counter-regulatory. Thus, animals that are superior in some responses may be inferior in others. This issue coupled with the potential loss of diversity in agricultural species makes it unlikely that major programs will be implemented to breed for helminth resistance except in areas exhibiting extremely high parasite pressure.

Amidst the controversy in breeding for helminth resistance in cattle, there are other ways that host genetics could be used to increase the efficiency of cattle and dairy production. One example is in dealing with nematodiasis which is a “quantitative” disease; namely, as parasite burdens increase so does loss. As such, very low parasitemias are likely to induce little or no economic effect in otherwise well-managed animals. With this in mind, the goal of modern parasite control programs should be to keep transmission rates low, and not eradicate the parasites, albeit unrealistic. In conjunction with the quantitative nature of the disease, the “overdispersed” character of nematode distributions must also be considered. This pattern implies that a few members of a herd are responsible for most of the parasite transmission. In an elegant analysis, Anderson and May (1985) calculated that controlling disease by treating/removing/vaccinating susceptible animals would be nearly as effective as treating the entire population. From an economic standpoint, rapid and inexpensive genetic identification of these “susceptible/non-responders” could reduce treatment costs substantially. Such an approach would lessen the selective pressures placed on parasite populations, and should slow the development of drug resistance. The key is in the development of reliable markers for susceptibility/resistance.

In a similar manner, identification of the genes involved in protection should provide producers with the information to make educated breeding decisions. The past few years have seen cattle producers move away from selecting

for single traits and become more aware of the benefits of selecting for a variety of production traits. The use of “breeding values” is a common subject among cattle producers. Providing producers with reliable genetic information on the performance of their cattle under parasite pressure will allow them to make informed breeding decisions based upon this trait. Producers in areas of high parasite pressure might choose not to use “susceptible” bulls, while those in areas with less parasite transmission could use these bulls if they provided other traits more important to their overall breeding programs.

Finally, as our ability to manipulate genomes becomes more refined, it will one day be possible to genetically engineer food producing animals to meet the specific needs of a given producer. The ability to do this with parasite resistance will first depend upon developing the means to rapidly and reliably identify the resistance phenotype of an individual animal. This will be followed by deciphering the genes, the mechanisms, and the precise regulatory processes involved.

## 6. The beginnings

Sequencing whole genomes and transcriptomes have positioned parasitology at the very foundation of a new era of investigative research. Next-generation advancements in sequencing technologies now permit us to evaluate the entire transcriptome of a 100 Mb organism to a depth of 100× coverage within a single set of experiments and in a period of days rather than months or years. Data are being generated at such alarming rates that the predictive value of sequence analysis programs are unable to keep up with the flurry of new information (Lu et al., 2007). As this industry moves forward, we hope one day to relate DNA sequence to structure and eventually to function, and use systems biology to gain holistic views of how the pieces work together at the cellular, organism and environmental levels. These are the goals of the 21st century parasitologist. But as we advance our understanding of genomes and transcriptomes, and as databases continue to balloon with information, we must not lose sight of factors other than sequence order that can and do affect disease outcomes.

## Conflict of interest

None declared.

## Acknowledgement

We would like to thank Dr. Benjamin Rosenthal (USDA, ARS, Beltsville, MD, USA) for helpful conversation and critically reading the manuscript.

## References

- Abrahamson, M.S., Templeton, T.J., Enomoto, S., Abrahante, J.E., Zhu, G., Lancto, C.A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G.A., Xu, P., Bankier, A.T., Dear, P.H., Konfortov, B.A., Spriggs, H.F., Iyer, L., Anantharaman, V., Aravind, L., Kapur, V., 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441–445.

- Almeria, S., Canals, A., Gomez, M.T., Zarlenga, D.S., Gasbarre, L.C., 1998. Characterization of protective immune responses in local lymphoid tissues after drug attenuated infections with *Ostertagia ostertagi*. *Vet. Parasitol.* 80, 53–64.
- Almeria, S., De Marez, T., Dawson, H., Araujo, R., Dubey, J.P., Gasbarre, L.C., 2003. Cytokine gene expression in dams and fetuses after experimental *Neospora caninum* infection of heifers at 110 days of gestation. *Parasite Immunol.* 25, 383–392.
- Anderson, R.M., May, R.M., 1985. Herd immunity to helminth infection and implications for parasite control. *Nature* 315, 493–496.
- Andries, K., Verhasselt, P., Guillemont, J., Göhlmann, H.W., Neefs, J.M., Winkler, H., Van Gestel, J., Timmerman, P., Zhu, M., Lee, E., Williams, P., de Chaffoy, D., Huitric, E., Hoffner, S., Cambau, E., Truffot-Pernot, C., Lounis, N., Jarlier, V., 2005. A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307, 223–227.
- Asara, J.M., Schweitzer, M.H., Freimark, L.M., Phillips, M., Cantley, L.C., 2007. Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* 316, 280–285.
- Ashrafi, K., Chang, F.Y., Watts, J.L., Fraser, A.G., Kamath, R.S., Ahringer, J., Ruvkun, G., 2003. Genome-wide RNAi analysis of *Caenorhabditis elegans* fat regulatory genes. *Nature* 421, 268–272.
- Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Ryan, A.M., Clayton, D., Li, L., Neibergs, H.L., Zhang, N., Grosse, W.M., Weiss, J., Creighton, P., McCarthy, F., Ron, M., Teale, A.J., Fries, R., McGraw, R.A., Moore, S.S., Georges, M., Soller, M., Womack, J.E., Hetzel, D.J.S., 1994. A genetic linkage map of the bovine genome. *Nat. Genet.* 6, 227–235.
- Bekal, S., Craig, J.P., Hudson, M.E., Niblack, T.L., Domier, L.L., Lambert, K.N., 2008. Genomic DNA sequence comparison between two inbred soybean cyst nematode biotypes facilitated by massively parallel 454 micro-bead sequencing. *Mol. Genet. Genomics* 279, 535–543.
- Beh, K.J., Hulme, D.J., Callaghan, M.J., Leish, Z., Lenane, I., Windon, R.G., Maddox, J.F., 2002. A genome scan for quantitative trait loci affecting resistance to *Trichostrongylus colubriformis* in sheep. *Anim. Genet.* 33, 97–106.
- Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580.
- Blackwell, J.M., 1988. Protozoan infections. In: Wakelin, D., Blackwell, J.M. (Eds.), *Genetics of Resistance to Bacterial and Parasitic Infections*. Taylor & Francis, London, pp. 103–151.
- Boothroyd, J.C., Kim, K., Ortega, E., Sibley, L.D., Soldati, D., 1995. Toxoplasma as a paradigm for the use of genetics in the study of parasitic protozoa. In: Boothroyd, J.C., Komuniecki, R. (Eds.), *Molecular Approaches to Parasitology*. Wiley-Liss, New York, pp. 211–225.
- Bourguinat, C., Ardelli, B.F., Pion, S.D., Kamgno, J., Gardon, J., Duke, B.O., Boussinesq, M., Prichard, R.K., 2008. P-glycoprotein-like protein, a possible genetic marker for ivermectin resistance selection in *Onchocerca volvulus*. *Mol. Biochem. Parasitol.* 158, 101–111.
- Bourguinat, C., Pion, S.D., Kamgno, J., Gardon, J., Duke, B.O., Boussinesq, M., Prichard, R.K., 2007. Genetic selection of low fertile *Onchocerca volvulus* by ivermectin treatment. *PLoS Negl. Trop. Dis.* 1, e72.
- Bradbury, J., 2003. Human Epigenome Project—Up and Running. *PLoS Biol.* 1, e82.
- Brown, D., Superti-Furga, G., 2003. Rediscovering the sweet spot in drug discovery. *Drug Discov. Today* 8, 1067–1077.
- Caffrey, C.R., Rohwer, A., Oellien, F., Marhöfer, R.J., Braschi, S., Oliveira, G., McKerrow, J.H., Selzer, P.M., 2009. A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PLoS ONE* 4, e4413.
- Canals, A., Zarlenga, D.S., Almeria, S., Gasbarre, L.C., 1997. Cytokine profile induced by a primary infection with *Ostertagia ostertagi* in cattle. *Vet. Immunol. Immunopathol.* 58, 63–75.
- Canals, A., Pasquali, P., Zarlenga, D.S., Fayer, R., Almeria, S., Gasbarre, L.C., 1998. Local ileal cytokine responses in cattle during a primary infection with *Cryptosporidium parvum*. *J. Parasitol.* 84, 125–130.
- Carlton, J.M., Escalante, A.A., Neafsey, D., Volkman, S.K., 2008. Comparative evolutionary genomics of human malaria parasites. *Trends Parasitol.* 24, 545–550.
- Castelo, A.T., Martins, W., Gao, G., 2002. TROLL-Tandem Repeat Occurrence Locator. *Bioinformatics* 18, 634–636.
- Casulli, A., Bart, J.M., Knapp, J., La Rosa, G., Dusher, G., Gottstein, B., Di Cerbo, A., Manfredi, M.T., Genchi, C., Piarroux, R., Pozio, E., 2009. Multi-locus microsatellite analysis supports the hypothesis of an autochthonous focus of *Echinococcus multilocularis* in northern Italy. *Int. J. Parasitol.* 39, 837–842.
- Choi, B.K., Chitwood, D.J., Paik, Y.K., 2003. Proteomic changes during disturbance of cholesterol metabolism by azacoprostane treatment in *Caenorhabditis elegans*. *Mol. Cell. Proteomics* 2, 1086–1095.
- Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., Cohen, L., Danon, T., Perzov, N., Alon, U., 2008. Dynamic proteomics of individual cancer cells in response to a drug. *Science* 322, 1511–1516.
- Cookson, W., Liang, L., Abeasis, G., Moffatt, M., Lathrop, M., 2009. Mapping complex disease traits with global gene expression. *Nat. Rev. Gen.* 10, 184–193.
- Copley, R.R., Aloy, P., Russell, R.B., Telford, M.J., 2004. Systematic searches for molecular synapomorphies in model metazoan genomes give some support for Ecdysozoa after accounting for the idiosyncrasies of *Caenorhabditis elegans*. *Evol. Dev.* 6, 164–169.
- Coppieters, W., Mes, T.H., Druet, T., Farnir, F., Tamma, N., Schrooten, C., Cornelissen, A.W., Georges, M., Ploeger, H.W., 2009. Mapping QTL influencing gastrointestinal nematode burden in Dutch Holstein-Friesian dairy cattle. *BMC Genomics* 10, 96.
- Crawford, A.M., Paterson, K.A., Dodds, K.G.D., Tascon, C., Williamson, P.A., Roberts, T.M., Bisset, S.A., Beattie, A.E., Greer, G.J., Green, R.S., Wheeler, R., Shaw, R.J., Knowler, K., McEwan, J.C., 2006. Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep. I. Analysis of outcross pedigrees. *BMC Genomics* 7, 178.
- Cristofanilli, M., Bryan, W.J., Miller, L.L., Chang, A.Y., Gradishar, W.J., Kufe, D.W., Hortobagyi, G.N., 1998. Phase II study of adozelesin in untreated metastatic breast cancer. *Anticancer Drugs* 9, 779–782.
- Crofton, H.D., 1971a. A quantitative approach to parasitism. *Parasitology* 62, 179–193.
- Crofton, H.D., 1971b. A model of host–parasite relationships. *Parasitology* 63, 343–364.
- Daniels, R., Volkman, S.K., Milner, D.A., Mahesh, N., Neafsey, D.E., Park, D.J., Rosen, D., Angelino, E., Sabeti, P.C., Wirth, D.F., Wiegand, R.C., 2008. A general SNP-based molecular barcode for *Plasmodium falciparum* identification and tracking. *Malar. J.* 7, 223.
- Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.
- Dieterich, C., Clifton, S.W., Schuster, L.N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., Fulton, L., Fulton, R., Godfrey, J., Minx, P., Mitreva, M., Roeseler, W., Tian, H., Witte, H., Yang, S.P., Wilson, R.K., Sommer, R.J., 2008. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat. Genet.* 40, 1193–1198.
- Donald, R.G., Allocco, J., Singh, S.B., Nare, B., Salowe, S.P., Wiltsie, J., Liberator, P.A., 2002. *Toxoplasma gondii* cyclic GMP-dependent kinase: chemotherapeutic targeting of an essential parasite protein kinase. *Eukaryot. Cell.* 1, 317–328.
- Dopazo, H., Dopazo, J., 2005. Genome-scale evidence of the nematode-arthropod clade. *Genome Biol.* 6, R41.
- Dopazo, H., Santoyo, J., Dopazo, J., 2004. Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. *Bioinformatics* 20 (Suppl. 1), I116–I121.
- Douzery, E.J., Snell, E.A., Baptiste, E., Delsuc, F., Philippe, H., 2004. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc. Natl. Acad. Sci. U.S.A.* 101, 15386–15391.
- Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14338–14343.
- Eisen, J.A., 1998. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. *Genome Res.* 8, 163–167.
- Eisen, J.A., Fraser, C.M., 2003. Phylogenomics: intersection of evolution and genomics. *Science* 300, 1706–1707.
- Emanuel, B., Saitta, S.C., 2007. From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. *Nat. Rev. Gen.* 8, 869–882.
- Feuk, L., Carson, A.R., Scherer, S.W., 2006. Structural variation in the human genome. *Nat. Rev. Gen.* 7, 85–97.
- Fire, A., Xu, S., Montgomery, M., Kostas, S., Driver, S., Mello, C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.
- Frearson, J.A., Wyatt, P.G., Gilbert, I.H., Fairlamb, A.H., 2007. Target assessment for antiparasitic drug discovery. *Trends Parasitol.* 23, 589–595.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., Altshuler, D., 2002. The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229.
- Gasbarre, L.C., Leighton, E.A., Davies, C.J., 1990. Genetic control of immunity to gastrointestinal nematodes of cattle. *Vet. Parasitol.* 37, 257–272.
- Gasbarre, L.C., Leighton, E.A., Davies, C.J., 1993. Influence of host genetics upon antibody responses against gastrointestinal nematode infections in cattle. *Vet. Parasitol.* 46, 81–91.

- Gasbarre, L.C., Leighton, E.A., Bryant, D., 1996. Reliability of a single fecal egg per gram determination as a measure of individual and herd values for trichostrongyle nematodes of cattle. *Am. J. Vet. Res.* 57, 168–171.
- Gasbarre, L.C., Miller, J.E., 2000. Genetics of helminth resistance. In: Axford, R.F.E., Bishop, S.C., Nichols, F.W., Owen, J.B. (Eds.), *Breeding for Disease Resistance in Farm Animals*. 2nd ed. CAB International Publishing, New York, pp. 129–152.
- Geldhof, P., Murray, L., Couthier, A., Gilleard, J.S., McLauchlan, G., Knox, D.P., Britton, C., 2006. Testing the efficacy of RNA interference in *Haemonchus contortus*. *Int. J. Parasitol.* 36, 801–810.
- Genchi, C., Madonna, M., Traldi, G., 1989. Epidemiology of *Ostertagia ostertagi* in dairy cows from different breeding systems. *Parasitologia* 31, 123–132.
- Gibbs, J.R., Singleton, A., 2006. Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. *PLoS Genet.* 6, e150.
- Gilleard, J.S., 2006. Understanding anthelmintic resistance: the need for genomics and genetics. *Int. J. Parasitol.* 36, 1227–1239.
- Gissot, M., Kelly, K.A., Ajioka, J.W., Grealley, J.M., Kim, K., 2007. Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog.* 3, e77.
- Gissot, M., Kim, K., 2008. How epigenomics contributes to the understanding of gene regulation in *Toxoplasma gondii*. *J. Eukaryot. Microbiol.* 55, 476–480.
- Grant, W.N., 1992. Transformation of *Caenorhabditis elegans* with genes from parasitic nematodes. *Parasitol. Today* 8, 344–346.
- Grillo, V., Jackson, F., Gilleard, J.S., 2006. Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers. *Mol. Biochem. Parasitol.* 148, 181–189.
- Gronvold, J., Nansen, P., Gasbarre, L.C., Christensen, C.M., Larsen, M., Monrad, J., Midtgaard, N., 1992. Development of immunity to *Ostertagia ostertagi* (Trichostrongylidae: Nematoda) in pastured young cattle. *Acta Vet. Scand.* 33, 305–316.
- Harris, M.P., Hasso, S.M., Ferguson, M.W., Fallon, J.F., 2006. The development of archosaurian first-generation teeth in a chicken mutant. *Curr. Biol.* 16, 371–377.
- Hasan, S., Daugelat, S., Rao, P.S., Schreiber, M., 2006. Prioritizing genomic drug targets in pathogens: application to *Mycobacterium tuberculosis*. *PLoS Comput. Biol.* 2, e61.
- Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.L., Hickenbotham, M., Huang, W., Magrini, V.J., Richt, R.J., Sander, S.N., Stewart, D.A., Stromberg, M., Tsung, E.F., Wylie, T., Schedl, T., Wilson, R.K., Mardis, E.R., 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods* 5, 183–188.
- Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., Craig, D.W., 2008a. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 4, e1000167.
- Homer, N., Tembe, W.D., Szelinger, S., Redman, M., Stephan, D.A., Pearson, J.V., Nelson, S.F., Craig, D.W., 2008b. Multimarker analysis and imputation of multiple platform pooling-based genome-wide association studies. *Bioinformatics* 24, 1896–1902.
- Huang, J., Mullapudi, N., Lancto, C.A., Scott, M., Abrahamsen, M.S., Kissinger, J.C., 2004a. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in *Cryptosporidium parvum*. *Genome Biol.* 5, R88.
- Huang, J., Mullapudi, N., Sicheritz-Ponten, T., Kissinger, J.C., 2004b. A first glimpse into the pattern and scale of gene transfer in Apicomplexa. *Int. J. Parasitol.* 34, 265–274.
- Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231.
- Johnson, P.C., Webster, L.M., Adam, A., Buckland, R., Dawson, D.A., Keller, L.F., 2006. Abundant variation in microsatellites of the parasitic nematode *Trichostrongylus tenuis* and linkage to a tandem repeat. *Mol. Biochem. Parasitol.* 148, 210–218.
- Kamath, R.S., Ahringer, J., 2003. Genome-wide RNAi screening in *Caenorhabditis elegans*. *Methods* 30, 313–321.
- Kaminsky, R., Ducray, P., Jung, M., Clover, R., Rufener, L., Bouvier, J., Weber, S.S., Wenger, A., Wieland-Berghausen, S., Goebel, T., Gauvry, N., Pautrat, F., Skripsky, T., Froelich, O., Komoin-Oka, C., Westlund, B., Sluder, A., Mäser, P., 2008. A new class of anthelmintics effective against drug-resistant nematodes. *Nature* 452, 176–180.
- Kappes, S.M., Keele, J.W., Stone, R.T., McGraw, A.A., Sonstegard, T.S., Smith, T.P., Lopez-Corrales, N.L., Beattie, C.W., 1997. A second-generation linkage map of the bovine genome. *Genome Res.* 7, 235–249.
- Karunaweera, N.D., Ferreira, M.U., Hartl, D.L., Wirth, D.F., 2007. Fourteen polymorphic microsatellite DNA markers for the human malaria parasite *Plasmodium vivax*. *Mol. Ecol. Notes* 7, 172–175.
- Karunaweera, N.D., Ferreira, M.U., Munasinghe, A., Barnwell, J.W., Collins, W.E., King, C.L., Kawamoto, F., Hartl, D.L., Wirth, D.F., 2008. Extensive microsatellite diversity in the human malaria parasite *Plasmodium vivax*. *Gene* 410, 105–112.
- Kim, K., Weiss, L.M., 2004. *Toxoplasma gondii*: the model apicomplexan. *Int. J. Parasitol.* 34, 423–432.
- Knox, D.P., Geldhof, P., Visser, A., Britton, C., 2007. RNA interference in parasitic nematodes of animals: a reality check? *Trends Parasitol.* 23, 105–107.
- Kolpakov, R., Bana, G., Kucherov, G., 2003. Mreps: efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res.* 31, 3672–3678.
- Korbel, J.O., Snel, B., Huynen, M.A., Bork, P., 2002. SHOT: a web server for the construction of genome phylogenies. *Trends Genet.* 18, 158–162.
- Kumar, S., Chaudhary, K., Foster, J.M., Novelli, J.F., Zhang, Y., Wang, S., Spiro, D., Ghedin, E., Carlow, C.K., 2007. Mining predicted essential genes of *Brugia malayi* for nematode drug targets. *PLoS ONE* 2, e1189.
- Kunieda, M., Tsuji, T., Abbasi, A.R., Khalaj, M., Ikeda, M., Miyadera, K., Ogawa, H., Kunieda, T., 2005. An insertion mutation of the bovine Fii gene is responsible for factor XI deficiency in Japanese black cattle. *Mamm. Genome* 16, 383–389.
- Kuo, C.H., Wares, J.P., Kissinger, J.C., 2008. The Apicomplexan whole-genome phylogeny: an analysis of incongruence among gene trees. *Mol. Biol. Evol.* 25, 2689–2698.
- Kwa, M.S., Veenstra, J.G., Roos, M.H., 1994. Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1. *Mol. Biochem. Parasitol.* 63, 299–303.
- Lau, A.O., 2009. An overview of the *Babesia*, *Plasmodium* and *Theileria* genomes: a comparative perspective. *Mol. Biochem. Parasitol.* 164, 1–8.
- Lee, J., Ahn, J., Bae, S.C., 2004. Homologs of RUNX and CBF beta/PEBP2 beta in *C. elegans*. *Oncogene* 23, 4346–4352.
- Lee, S., Horn, V., Julien, E., Liu, Y., Wysocka, J., Bowerman, B., Hengartner, M.O., Herr, W., 2007. Epigenetic regulation of histone H3 serine 10 phosphorylation status by HCF-1 proteins in *C. elegans* and mammalian cells. *PLoS ONE* 2, e1213.
- Lendner, M., Doligalska, M., Lucius, R., Hartmann, S., 2008. Attempts to establish RNA interference in the parasitic nematode *Heligmosomoides polygyrus*. *Mol. Biochem. Parasitol.* 161, 21–31.
- Li, R.W., Gasbarre, L.C., 2009. A temporal shift in regulatory networks and pathways in the bovine small intestine during *Cooperia oncophora* infection. *Int. J. Parasitol.* 39, 813–824.
- Lingle, J.K., Koch, C.R., 2001. *The Breed of Noble Bloods*, 2nd ed. University of Maryland Foundation, Adelphi, MD, p. 160.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., Ecker, J.R., 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536.
- Lu, F., Jiang, H., Ding, J., Mu, J., Valenzuela, J.G., Ribeiro, J.M., Su, X.Z., 2007. cDNA sequences reveal considerable gene prediction inaccuracy in the *Plasmodium falciparum* genome. *BMC Genomics* 8, 255.
- Mackinnon, K.M., Burton, J.L., Zajac, A.M., Notter, D.R., 2009. Microarray analysis reveals difference in gene expression profiles of hair and wool sheep infected with *Haemonchus contortus*. *Vet. Immunol. Immunopathol.* doi:10.1016/j.vetimm.2009.02.013 [Epub ahead of print].
- McCarter, J.P., 2004. Genomic filtering: an approach to discovering novel antiparasitics. *Trends Parasitol.* 20, 462–468.
- Mitreva, M., Blaxter, M.L., Bird, D.M., McCarter, J.P., 2005. Comparative genomics of nematodes. *Trends Genet.* 21, 573–581.
- Mitreva, M., McCarter, J.P., Martin, J., Dante, M., Wylie, T., Chiappelli, B., Pape, D., Clifton, S.W., Nutman, T.B., Waterston, R.H., 2004. Comparative genomics of gene expression in the parasitic and free-living nematodes *Strongyloides stercoralis* and *Caenorhabditis elegans*.
- Morgan, C., LaCourse, E.J., Rushbrook, B.J., Gretham, D., Hamilton, J.V., Barrett, J., Bailey, K., Brophy, P.M., 2006. Plasticity demonstrated in the proteome of a parasitic nematode within the intestine of different host strains. *Proteomics* 6, 4633–4645.
- Morlais, I., Ravel, S., Grébaud, P., Dumas, V., Cuny, G., 2001. New molecular marker for *Trypanosoma (Duttonella) vivax* identification. *Acta Trop.* 80, 207–213.
- Nairz, K., Zipperlen, P., Schneider, M., 2007. FLP-Mapping: a universal, cost-effective, and automatable method for gene mapping. *Methods Mol. Biol.* 396, 419–432.
- Natalang, O., Bischoff, E., Deplaine, G., Proux, C., Dillies, M.A., Sismeiro, O., Guigon, G., Bonnefoy, S., Patarapotikul, J., Mercereau-Puijalon, O., Coppée, J.Y., David, P.H., 2008. Dynamic RNA profiling in *Plasmodium falciparum* synchronized blood stages exposed to lethal doses of artesunate. *BMC Genomics* 9, 388.

- Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner Jr., D.A., Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J.F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M.U., Moraes, S.D., Dash, A.P., Chitnis, C.E., Wiegand, R.C., Hartl, D.L., Birren, B.W., Lander, E.S., Sabeti, P.C., Wirth, D.F., 2008. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. *Genome Biol.* 9, R171.
- Nejsum, P., Frydenberg, J., Roepstorff, A., Parker Jr., E.D., 2005a. Population structure in *Ascaris suum* (Nematoda) among domestic swine in Denmark as measured by whole genome DNA fingerprinting. *Heredity* 142, 7–14.
- Nejsum, P., Parker Jr., E.D., Frydenberg, J., Roepstorff, A., Boes, J., Haque, R., Astrup, I., Prag, J., Skov Sørensen, U.B., 2005b. Ascariasis is a zoonosis in Denmark. *J. Clin. Microbiol.* 43, 1142–1148.
- O'Brien, S.J., Stanton, R., 1999. Phylogenomics: ancestral primate viewed. *Nature* 402, 365–366.
- Orlandi, P.A., Carter, L., Brinker, A.M., da Silva, A.J., Chu, D.M., Lampel, K.A., Monday, S.R., 2003. Targeting single-nucleotide polymorphisms in the 18S rRNA gene to differentiate *Cyclospora* species from *Eimeria* species by multiplex PCR. *Appl. Environ. Microbiol.* 69, 4806–4813.
- Parisi, V., De Fonzo, V., Aluffi-Pentini, F., 2003. STRING: finding tandem repeats in DNA sequences. *Bioinformatics* 19, 1733–1738.
- Philippe, H., Delsuc, F., Brinkmann, H., Lartillot, N., 2005a. Phylogenomics. *Ann. Rev. Ecol. Evol. Syst.* 36, 541–562.
- Philippe, H., Lartillot, N., Brinkmann, H., 2005b. Multigene analyses of bilateral animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol. Biol. Evol.* 22, 1246–1253.
- Philippe, H., Snell, E.A., Baptiste, E., Lopez, P., Holland, P.W., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 21, 1740–1752.
- Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21, 1455–1458.
- Prieto, J.H., Koncarevic, S., Park, S.K., Yates 3rd, J., Becker, K., 2008. Large-scale differential proteome analysis in *Plasmodium falciparum* under drug treatment. *PLoS ONE* 3, e4098.
- Rawat, D.S., Lumb, V., Sharma, Y.D., Pasha, S.T., Singh, G., 2007. Histone as future drug target for malaria. *J. Commun. Dis.* 39, 119–128.
- Rodrigues, N.B., Coura Filho, P., de Souza, C.P., Jannotti Passos, L.K., Dias-Neto, E., Romanha, A.J., 2002. Populational structure of *Schistosoma mansoni* assessed by DNA microsatellites. *Int. J. Parasitol.* 32, 843–851.
- Rokas, A., Williams, B.L., King, N., Carroll, S.B., 2003. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 425, 798–804.
- Roos, D.S., Crawford, M.J., Donald, R.G., Fohl, L.M., Hager, K.M., Kissinger, J.C., Reynolds, M.G., Striepen, B., Sullivan Jr., W.J., 1999. Transport and trafficking: *Toxoplasma* as a model for *Plasmodium*. *Novartis Found. Symp.* 226, 176–195.
- Rosenthal, B.M., La Rosa, G., Zarlenga, D., Dunams, D., Chunyu, Y., Mingyuan, L., Pozio, E., 2008. Human dispersal of *Trichinella spiralis* in domesticated pigs. *Infect. Genet. Evol.* 8, 799–805.
- Roy, S.W., Gilbert, W., 2005. Resolution of a deep animal divergence by the pattern of intron conservation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4403–4408.
- Schnack, H.G., Bakker, S.C., van't Slot, R., Groot, B.M., Sinke, R.J., Kahn, R.S., Pearson, P.L., 2004. Accurate determination of microsatellite allele frequencies in pooled DNA samples. *Eur. J. Hum. Genet.* 12, 925–934.
- Scholl, E.H., Thorne, J.L., McCarter, J.P., Bird, D.M., 2003. Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biol.* 4, R39.
- Schweitzer, M.H., Suo, Z., Avci, R., Asara, J.M., Allen, M.A., Arce, F.T., Horner, J.R., 2007. Analyses of soft tissue from *Tyrannosaurus rex* suggest the presence of protein. *Science* 316, 277–280.
- Silva, L.K., Liu, S., Blanton, R.E., 2006. Microsatellite analysis of pooled *Schistosoma mansoni* DNA: an approach for studies of parasite populations. *Parasitology* 132, 331–338.
- Silveira, H., Ramos, S., Abrantes, P., Lopes, L.F., do Rosario, V.E., Abrahamson, M.S., 2007. Effect of chloroquine on gene expression of *Plasmodium yoelii nigeriensis* during its sporogonic development in the mosquito vector. *Malar. J.* 6, 84.
- Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J., Plasterk, R.H., 2003. Genome-wide RNAi of *C. elegans* using the hypersensitive rrf-3 strain reveals novel gene functions. *PLoS Biol.* 1, e12.
- Simões, M., Bahia, D., Zerlotini, A., Torres, K., Artiguenave, F., Neshich, G., Kuser, P., Oliveira, G., 2007. Single nucleotide polymorphisms identification in expressed genes of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 154, 134–140.
- Simonet, T., Dulerio, R., Schott, S., Palladino, F., 2007. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in *C. elegans* development. *Dev Biol.* 312, 367–383.
- Shendure, J., 2008. The beginning of the end for microarrays? *Nat. Methods* 5, 585–587.
- Snel, B., Bork, P., Huynen, M.A., 1999. Genome phylogeny based on gene content. *Nat. Genet.* 21, 108–110.
- Sobreira, T.J., Durham, A.M., Gruber, A., 2006. TRAP: automated classification, quantification and annotation of tandemly repeated sequences. *Bioinformatics* 22, 361–362.
- Soria, J.M., Almasy, L., Souto, J.C., Sabater-Lleal, M., Fontcuberta, J., Blangero, J., 2005. The F7 gene and clotting factor VII levels: dissection of a human quantitative trait locus. *Hum. Biol.* 77, 561–575.
- Stadler, M., Mayer, A., Anke, H., Sterner, O., 1994. Fatty acids and other compounds with nematicidal activity from cultures of *Basidiomycetes*. *Planta Med.* 60, 128–132.
- Stear, M.J., Fitton, L., Innocent, G.T., Murphy, L., Rennie, K., Matthews, L., 2007. The dynamic influence of genetic variation on the susceptibility of sheep to gastrointestinal nematode infection. *J. R. Soc. Interface* 4, 767–776.
- Stear, M.J., Hetzel, D.J.S., Brown, S.C., Gershwin, L.J., Mackinnon, M.J., Nicholas, F.W., 1990. The relationships among ecto- and endoparasite levels, class I antigens of the bovine major histocompatibility system, immunoglobulin E levels and weight gain. *Vet. Parasitol.* 34, 303–321.
- Stear, M.J., Nicholas, F.W., Brown, S.C., Tierney, T., Rudder, T., 1984. The relationship between the bovine major histocompatibility system and faecal worm egg counts. In: Dineen, J.K., Outteridge, P.M. (Eds.), *Immunogenetic Approaches to the Control of Endoparasites with Particular Reference to Parasites of Sheep*. CSIRO, Australia, pp. 126–133.
- Striepen, B., Kissinger, J.C., 2004. Genomics meets transgenics in search of the elusive *Cryptosporidium* drug target. *Trends Parasitol.* 20, 355–358.
- Striepen, B., Pruijssers, A.J., Huang, J., Li, C., Gubbels, M.J., Umejiego, N.N., Hedstrom, L., Kissinger, J.C., 2004. Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3154–3159.
- Striepen, B., White, M.W., Li, C., Guerini, M.N., Malik, S.B., Logsdon Jr., J.M., Liu, C., Abrahamson, M.S., 2002. Genetic complementation in apicomplexan parasites. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6304–6309.
- Swan, K.A., Curtis, D.E., McKusick, K.B., Voinov, A.V., Mapa, F.A., Cancilla, M.R., 2002. High-throughput gene mapping in *Caenorhabditis elegans*. *Genome Res.* 12, 1100–1105.
- The International SNP Map Working Group, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933.
- Van der Veer, M., de Vries, E., 2004. A single nucleotide polymorphism map of the mitochondrial genome of the parasitic nematode *Cooperia oncophora*. *Parasitology* 128, 421–431.
- Viney, M.E., Thompson, F.J., 2008. Two hypotheses to explain why RNA interference does not work in animal parasitic nematodes. *Int. J. Parasitol.* 38, 43–47.
- Visser, A., Geldhof, P., de Maere, V., Knox, D.P., Vercruysse, J., Claerebout, E., 2006. Efficacy and specificity of RNA interference in larval life-stages of *Ostertagia ostertagi*. *Parasitology* 133, 777–783.
- Volkman, S.K., Sabeti, P.C., DeCaprio, D., Neafsey, D.E., Schaffner, S.F., Milner Jr., D.A., Daily, J.P., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Duraisingh, M.T., Lukens, A., Derr, A., Stange-Thomann, N., Waggoner, S., Onofrio, R., Ziaugra, L., Mauceli, E., Gnerre, S., Jaffe, D.B., Zainoun, J., Wiegand, R.C., Birren, B.W., Hartl, D.L., Galagan, J.E., Lander, E.S., Wirth, D.F., 2007. A genome-wide map of diversity in *Plasmodium falciparum*. *Nat. Genet.* 39, 113–119.
- Von Samson-Himmelstjerna, G., Blackhall, W., 2005. Will technology provide solutions for drug resistance in veterinary helminths? *Vet. Parasitol.* 132, 223–239.
- Von Samson-Himmelstjerna, G., Blackhall, W.J., McCarthy, J.S., Skuce, P.J., 2007. Single nucleotide polymorphism (SNP) markers for benzimidazole resistance in veterinary nematodes. *Parasitology* 134, 1077–1086.
- Wang, Z., Martin, J., Abubucker, S., Yin, Y., Gasser, R.B., Mitreva, M., 2009. Systematic analysis of insertions and deletions specific to nematode proteins and their proposed functional and evolutionary relevance. *BMC Evol. Biol.* 9, 23, doi:10.1186/1471-2148-9-23 [Epub ahead of print].
- Webster, L.M., Johnson, P.C., Adam, A., Mable, B.K., Keller, L.F., 2008. Absence of three known benzimidazole resistance mutations in *Trichostrongylus tenuis*, a nematode parasite of avian hosts. *Vet. Parasitol.* 158, 302–310.
- Whitlock, J.H., 1955. A study of inheritance of resistance to trichostrongylosis in sheep. *Cornell Vet.* 45, 422–439.

- Wicks, S.R., Yeh, R.T., Gish, W.R., Waterston, R.H., Plasterk, R.H., 2001. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat. Genet.* 28, 160–164.
- Williams, D.J., Kloek, A.P., Hresko, M.C., 2005. Nematicidal compositions and methods. U.S. Patents 6,903,052 and 6,887,900.
- Wolf, Y.I., Rogozin, I.B., Koonin, E.V., 2004. Coelomata and not Ecdysozoa: evidence from genome-wide phylogenetic analysis. *Genome Res.* 14, 29–36.
- Wolf, M.Y., Wolf, Y.I., Koonin, E.V., 2008. Comparable contributions of structural-functional constraints and expression level to the rate of protein sequence evolution. *Biol. Direct.* 3, 40.
- Woynarowski, J.M., Krugliak, M., Ginsburg, H., 2007. Pharmacogenomic analyses of targeting the AT-rich malaria parasite genome with AT-specific alkylating drugs. *Mol. Biochem. Parasitol.* 154, 70–81.
- Wu, Y., Wang, X., Liu, X., Wang, Y., 2003. Data-mining approaches reveal hidden families of proteases in the genome of malaria parasite. *Genome Res.* 13, 601–616.
- Yanow, S.K., Purcell, L.A., Pradel, G., Sato, A., Rodriguez, A., Lee, M., Spithill, T.W., 2008. Potent antimalarial and transmission-blocking activities of centanamycin, a novel DNA-binding agent. *J. Infect. Dis.* 197, 527–534.
- Yanow, S.K., Purcell, L.A., Spithill, T.W., 2006. The A/T-specific DNA alkylating agent adozelesin inhibits *Plasmodium falciparum* growth in vitro and protects mice against *Plasmodium chabaudi adami* infection. *Mol. Biochem. Parasitol.* 148, 52–59.
- Yin, Y., Martin, J., Abubucker, S., Scott, A.L., McCarter, J.P., Wilson, R.K., Jasmer, D.P., Mitreva, M., 2008. Intestinal transcriptomes of nematodes: comparison of the parasites *Ascaris suum* and *Haemonchus contortus* with the free-living *Caenorhabditis elegans*. *PLoS Negl. Trop. Dis.* 2, e269.