Biotechnology Advances 34 (2016) 663–686

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Making sense of genomes of parasitic worms: Tackling bioinformatic challenges

Pasi K. Korhonen *, Neil D. Young, Robin B. Gasser *

Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 5 November 2015
Received in revised form 25 February 2016
Accepted 1 March 2016
Available online 5 March 2016

Keywords:
Parasitic worms
Bioinformatics
Genome
Assembly
Annotation

ABSTRACT

Billions of people and animals are infected with parasitic worms (helminths). Many of these worms cause diseases that have a major socioeconomic impact worldwide, and are challenging to control because existing treatment methods are often inadequate. There is, therefore, a need to work toward developing new intervention methods, built on a sound understanding of parasitic worms at molecular level, the relationships that they have with their animal hosts and/or the diseases that they cause. Decoding the genomes and transcriptomes of these parasites brings us a step closer to this goal. The key focus of this article is to critically review and discuss bioinformatic tools used for the assembly and annotation of these genomes and transcriptomes, as well as various post-genomic analyses of transcription profiles, biological pathways, synteny, phylogeny, biogeography and the prediction and prioritisation of drug target candidates. Bioinformatic pipelines implemented and established recently provide practical and efficient tools for the assembly and annotation of genomes of parasitic worms, and will be applicable to a wide range of other parasites and eukaryotic organisms. Future research will need to assess the utility of long-read sequence data sets for enhanced genomic assemblies, and develop improved algorithms for gene prediction and post-genomic analyses, to enable comprehensive systems biology explorations of parasitic organisms.

Contents

* Corresponding authors.
E-mail addresses: pasi.korhonen@unimelb.edu.au (P.K. Korhonen), robinbg@unimelb.edu.au (R.B. Gasser).

http://dx.doi.org/10.1016/j.biotechadv.2016.03.001
0734-9750/

1. Introduction — importance of nematode genomics

Parasitic worms (helminths) of humans and animals cause diseases of major socioeconomic importance globally (Fenwick, 2012; Utzinger, 2012). In humans, the global disease burden caused by worms in disability-adjusted life years (DALYs) is comparable to that caused by malaria and tuberculosis (Hotez et al., 2006, 2009), and represents ~1% of DALYs worldwide (Murray et al., 2012). These diseases can be transmitted via water, food, soil or other means, and a considerable number of them are zoonotic (Pozio, 2003; Bethony et al., 2006). In livestock, the annual economic losses due to death, poor health and reduced productivity caused by roundworms (nematodes), such as *Haemonchus* spp., are estimated at billions of dollars per annum worldwide (cf. Newton and Munn, 1999; Preston et al., 2015). In Australia alone, endemic and gastrointestinal worms of sheep and cattle cause economic losses of 410 to 530 million dollars (Sackett et al., 2006; Lane et al., 2015).

As an example of an important zoonotic parasite, the intestinal nematode *Toxocara* has major health importance in carnivores (dogs or cats), humans and other hosts. Indeed, millions of people are infected with *Toxocara canis*, and the worldwide prevalence of serum antibodies against *Toxocara* in humans ranges from 7% to 82% (Maizels, 2013). Also foodborne nematodes, such as *Trichinella*, are of human and animal health importance (Pozio and Zarlenga, 2013). The prevalence of trichinellosis in humans is estimated at 11 million (Dupouy-Camet, 2000), and the incidence (annual) is estimated at 2500–10,000 (Pozio, 2007; Gottstein et al., 2009; Murrell and Pozio, 2011). In addition, *Toxocara spiralis* causes a major economic burden (>$1.5 billion per annum in Europe and the USA) associated with preventative measures in the human food chain (Murrell and Pozio, 2000). Clearly, these examples illustrate, how parasitic nematodes impact on human and animal health.

Despite current knowledge of parasitic nematodes and the diseases that they cause (Anderson, 2001), little is known about their molecular biology, genetics and evolution; how they invade and establish in the host, cause disease, interact, suppress and evade their host's immune responses; and how resistance develops and evolves against current treatments. For instance, although resistance to most commonly used anthelmintics is widespread (Kaplan and Vidyashankar, 2012) and seriously compromises nematode control, it is not possible to accurately estimate the prevalence of drug resistant genotypes of nematodes (Howell et al., 2008; Sczesny-Moraes et al., 2010; Kaplan and Vidyashankar, 2012). The related annual cost of anthelmintic treatment worldwide is estimated at $3 billion, and this cost, together with the production losses caused by nematode infections/diseases, has stimulated research toward developing improved intervention methods. Thus far, the development of vaccines against parasitic worms has been relatively unsuccessful (Knox, 2011, 2013), and the development of novel anthelmintic drugs has been relatively slow, with emodepside (patent WO 1997/02256) (Harder and von Samson-Himmelstjerna, 2001), monepantel (patent WO2005/44784) (Kaminsky et al., 2008) and derquantel (patent WO1997/03988) (Lee et al., 2001) being new anthelmintics developed in the past 20 years.

There is a clear imperative to gain an improved understanding of the biology of socioeconomically important parasitic nematodes at the molecular level, in order to underpin new methods of treatment, diagnosis and control. Advanced genomic and bioinformatic technologies should provide useful tools to investigate nematode genomes and transcriptomes, to elucidate the global "molecular landscapes" of nematodes as well as key aspects of their biochemistry and physiology. In spite of technological advances, there have been considerable challenges in the sequencing and annotation of nematode genomes and post-genomic analyses of genomic data sets. In the present article, we (i) critically appraise genome sequencing technologies as well as algorithms and approaches used for genome assembly and annotation (highlighting their advantages and disadvantages), (ii) summarise developments in the sequencing and drafting of genomes of parasitic nematodes using various bioinformatics tools; (iii) describe enhanced pipelines constructed and employed in our laboratory, which have proven to be practical and efficient for the assembly and annotation of worm genomes, and, finally, (iv) provide a perspective on future research toward improved genomics/bioinformatics of parasitic nematodes.

2. Appraisal of genomic and bioinformatics approaches

In terms of bioinformatics, a draft genome project for a parasite encompasses genome assembly and annotation (Stein, 2001; Yandell and Ence, 2012). The success of a genome assembly depends on the sequencing technology and assembly algorithm(s) used. However, typically, short-read, shot-gun assemblies do not lead to complete chromosomal assemblies of eukaryote genomes, mainly because of the difficulties in the resolution of repeat regions (Alkan et al., 2011), and a lack of uniform read coverage usually associated with the suboptimal quality of genomic DNA and of library construction (Gasser et al., 1993, 2006; Raghunathan et al., 2005; Gasser, 2013).

Genome annotation can be divided into structural (*computational*) and functional (*annotation*) phases (Stein, 2001; Ekblom and Wolf, 2014). During structural annotation, genomic features, such as genes, repeat regions and RNA species, are predicted, and their composition and location in the genome recorded. For gene prediction, usually the