Parasitology International 66 (2017) 341–345

Contents lists available at ScienceDirect

Parasitology International

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

Genomics of worms, with an emphasis on *Opisthorchis viverrini* — opportunities for fundamental discovery and biomedical outcomes

Robin B. Gasser [a,*], Patrick Tan [b,c], Bin Tean Teh [c], Sopit Wongkham [d], Neil D. Young [a,*]

[a] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia
[b] Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672, Republic of Singapore
[c] Division of Cancer and Stem Cell Biology, Duke-NUS Graduate Medical School, Singapore 138672, Republic of Singapore
[d] Faculty of Medicine, Department of Biochemistry, Liver Fluke and Cholangiocarcinoma Research Center, Khon Kaen University, Khon Kaen 40002, Thailand

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form 7 January 2016
Accepted 10 January 2016
Available online 12 January 2016

Keywords:
Neglected tropical disease (NTD)
Opisthorchis viverrini
Carcinogen
Molecular biology
Bioinformatics
Genome/transcriptome
Biotechnology

ABSTRACT

Neglected tropical diseases cause substantial morbidity and mortality in animals and people globally. Opisthorchiasis is one such disease, caused by the carcinogenic, Asian liver fluke, *Opisthorchis viverrini*. This hepatobiliary disease is known to be associated with malignant cancer (cholangiocarcinoma, CCA) and affects millions of people in Asia, including Thailand, Lao People's Democratic Republic (PDR) and Cambodia. No vaccine is available, and only one drug (praziquantel) is routinely employed against the parasite. Relatively little is known about the molecular biology of the fluke itself and the disease complex that it causes in humans. With the advent of high-throughput nucleic acid sequencing and bioinformatic technologies, it has now become possible to gain global insights into the molecular biology of parasites. The purpose of this minireview is (i) to discuss recent progress on the genomics of parasitic worms, with an emphasis on the draft genome and transcriptome of *O. viverrini*; (ii) to use results from an integrated, global analysis of the genomic and transcriptomic data, to explain how we believe that this carcinogenic fluke establishes in the biliary system, how it feeds, survives and protects itself in such a hostile, microaerobic environment within the liver, and to propose how this parasite evades or modulates host attack; and (iii) to indicate some of the challenges, and, more importantly, the exciting opportunities that the 'omic resources for *O. viverrini* now provide for a plethora of fundamental and applied research areas. Looking ahead, we hope that this genomic resource stimulates vibrant and productive collaborations within a consortium context, focused on the effective control of opisthorchiasis.

1. Introduction

Compounded by a massive global food shortage, neglected tropical diseases (NTDs) caused by parasitic worms are amongst the world's big challenges [1]. Billions of people are infected with worms (=helminths), which have a comparable socio-economic burden to that of diabetes or lung cancer in disability adjusted life years (DALYs) [2]. These worms include roundworms (=nematodes) and flatworms (=flukes and tapeworms).

Liver flukes, such as *Clonorchis* and *Opisthorchis* spp. (family Opisthorchiidae), are particularly important food-borne pathogens of humans and other fish-eating mammals including canids and felids [3,4]. These parasites are particularly notable because they are classified as a group 1 carcinogen by the International Agency for Research on Cancer (IARC) [5]. *Opisthorchis viverrini* causes opisthorchiasis, which has a major public health impact mainly in countries of the Asia Pacific [4,6]. This pathogen affects tens of millions of people, and millions more are at risk of reccurrent infection [4,6]. Despite control efforts, disease prevalence can be as high as 70% in some Asian countries, including Thailand [6,7]. Chronic infection is linked to cholangitis, bile duct cancer (=cholangiocarcinoma, CCA) and associated complications [6,8]. Although CCA incidence is low in Western countries, this cancer is prevalent in many parts of South East Asia where *O. viverrini* is endemic, including Cambodia, Lao PDR and Thailand, where an age-standardized incidence rate (ASIR) of up to 96 per 100,000 has been reported [9]. Current estimates indicate that chronic opisthorchiasis affects ~10 million people worldwide, and, in Asia, fluke-associated CCA is detected in >2500 people annually [10]. The control of opisthorchiasis relies principally on treating infected people with praziquantel (an anthelmintic compound), as the cultural tradition of eating raw cyprinoid fish (second intermediate host infected with metacercariae) is entrenched. Humans develop only a limited degree of immunity against opisthorchiid liver flukes [11], such that they frequently become re-

* Corresponding authors at: Faculty of Veterinary and Agricultural Sciences, Parkville, Victoria 3010, Australia.
E-mail addresses: robinbg@unimelb.edu.au (R.B. Gasser), nyoung@unimelb.edu.au (N.D. Young).

http://dx.doi.org/10.1016/j.parint.2016.01.005
1383-5769/

infected. Although there has also been a focus on developing alternative control methods [4,12], no vaccines are yet available. In our opinion, detailed fundamental molecular biological investigations of *O. viverrini*/opisthorchiasis could be useful to underpin the development of novel intervention methods. However, until recently, most molecular studies of flatworms had focused predominantly on human blood flukes (schistosomes) [13–15], complemented by sustained efforts to establish in vitro systems for functional genomic analyses (e.g., [16–18]).

In 2009, when the genomes of two blood flukes (*Schistosoma mansoni* and *S. japonicum*) were published [13,14], there was very little molecular information available for opisthorchid flukes. In 2010, we undertook the first transcriptomic surveys of *C. sinensis* and *O. viverrini* using 454 sequencing technology [19,20], followed by a study of differential transcription between immature and adult *O. viverrini* [21]. To do this, we combined previous 454 sequence data with RNA-sequence (RNA-seq) data to achieve an enhanced assembly and annotation of the transcriptome, in order to underpin the differential transcription analysis. In spite of this enhancement, there were some issues relating to transcript redundancy, incomplete annotations and suboptimal mapping of transcripts. In order to overcome these limitations, in 2012, we decided to sequence the nuclear genome of *O. viverrini* using Illumina technology, at a time when this technology was revolutionizing the sequencing of many other animal genomes (cf. [22,23]). With the completion of this draft genome [24], our intent was to provide the scientific community with a molecular resource for future transcriptomic, proteomic and functional genomic and a plethora of other studies of a carcinogenic liver fluke. In 2011, a draft genome of *C. sinensis* was reported from China [25], and an enhanced version presented by the same research team two years later [26].

2. The draft genome of *O. viverrini*, its features and functional annotation

In 2014, our international team published this draft genome [24]; here, we summarize some salient features. We generated 79.9 Gb of short-read sequence data (>130-fold genome coverage) from seven libraries (170 bp to 20 kb) constructed from genomic DNA from multiple adult specimens of *O. viverrini*. Following the verification of low sequence heterozygosity within and among some libraries, we assembled the genome into scaffolds to produce a draft genome of 634.5 Mb (N50 = 1,323,951 bp; repeat content: 31%; GC-content = 44%), in which we found ~86% of 248 core essential genes. We predicted 16,379 protein-encoding genes using transcriptomic evidence and sequence data for *C. sinensis* (see [25]) and blood flukes [13–15]. Most (87%) genes were supported by published RNA-seq data [19,21], and >99% of assembled transcripts mapped to the genome. The estimated number of genes, proportion of coding regions (3.4%), and mean total gene length (18,231 bp), exon length (254 bp), intron length (3,531 bp), and mean number of exons per gene (6) were similar to *C. sinensis* [26], but distinct from other flukes. More recently, comparative genomics revealed a number of conserved (n = 16) and novel (n = 19) microRNA in *O. viverrini* [27], most of which lack functional annotation.

Structurally, the *O. viverrini* draft genome is very divergent from all other published genomes of flukes, including *C. sinensis*, *S. haematobium*, *S. japonicum* and *S. mansoni* [13–15,26]. In particular, only 22% of *O. viverrini* scaffolds could be aligned to 26% of *C. sinensis* scaffolds (at the nucleotide level). We interpreted this lack of conservation to relate to karyotypic differences, with *O. viverrini* having 12 chromosomes (2n) [28], and *C. sinensis* having 14 (Russian isolate) [29] or 58 (Korean isolate) [30], and all human schistosomes having eight chromosomes [31].

3. What can we learn about this carcinogenic fluke by inference from genomic and published information?

Although the main morphological changes that take place during the life cycle of *O. viverrini* are well known, little is understood about the molecular and biochemical processes underlying developmental changes and survival as well as parasite–host interactions and disease. Insights into these fundamental processes are critically important, and could provide a basis for the identification of targets for the design of new interventions. Using advanced bioinformatic pipelines for parasites, we integrated all of the genomic, transcriptomic, inferred proteomic data as well as published information to characterize some of the molecular landscape of *O. viverrini* [24]. This effort allowed us to address some key biological questions regarding fundamental molecular biology of this pathogen, infer essential pathways associated with the fluke-host interplay and to suggest some genes/gene products that might contribute to CCA development.

3.1. How does the fluke migrate to and establishment in the biliary duct?

When ingested, metacercariae of *O. viverrini* pass through the digestive tract, where they excyst, migrate to and then establish in the biliary tract [32]. Proteases, including aspartic and cysteine peptidases, appear to play a key role in excystment [24,33,34], being reflected in high transcription in opisthorchiid metacercariae [32]. The large number of GPCRs and ion channels encoded in the genome might enable chemotaxis-mediated migration to the biliary duct; molecules such as the rhodopsin biogenic amine receptors and ion channels are conserved for opisthorchiids and divergent from flukes that live external to the biliary system [24].

3.2. How does the worm feed and nourish itself?

The newly excysted juvenile (NEJ) stage of *O. viverrini* relies initially on energy stored within glycogen granules and lipid droplets in the excretory bladder of the worm [35]. However, these energy reserves are rapidly depleted once the fluke reaches the bile duct. Thus, the developing fluke must rapidly acquire nutrients and energy from its surrounding environment for survival, development and reproduction. Because bile contains extremely low levels of glucose [36], the fluke cannot rely on this sugar for energy. However, bile is rich in high (HDLs), intermediate (IDLs), low (LDLs) and very-low density lipoproteins (VLDLs), which all contain differing proportions of triglycerides, phospholipids, cholesterol and amphipathic proteins [37]; it is also rich in branched-chain amino acids [38] and long-chain saturated palmitic acid (C16:0) as well as unsaturated linoelaidic (C18:2) and arachidonic (C20:4) fatty acids [39,40].

Therefore, the fluke produces enzymes and accessory proteins to process bile and blood constituents [24,33,34,41]. Using its vast array of peptidases with broad substrate specificity, the fluke likely degrades lipoprotein complexes and proteins. Free amino acids are then taken up via amino acid transporters, after which they are processed for energy via acetyl CoA [24]). As the fluke is unable to synthesize cholesterol, the fluke likely uses a scavenger-like receptor (SR-B1) to transport cholesterol from HDLs into cells [42], and LDL receptor (LDLR), LDLR-related protein 1 receptor (LRP1) and CD36-like receptors for LDL, IDL and fatty acid uptake, respectively. In addition, the fluke also has an expanded group of lipid-binding proteins (n = 25) with a MD-2-related lipid-binding domain, which are homologous to the human Niemann–Pick C2 protein (NPC2) [43] and enable intracellular and extracellular sterol transport [43,44]. Fifteen of these 25 NPC2-like proteins appear to be expressed in fluke stages within the bile duct [24]. Remarkably, other eukaryotes appear to express only one such protein [43]. Clearly, this expansion in *O. viverrini* appears to reflect the significance of the binding and/or transportation of sterols and/or lipids and intracellular cholesterol. In addition to an adaptation to a lipid-rich diet, *O. viverrini* also likely degrades cholangiocyte components using galactosylceramidase/galactocerebrosidase (GALC) and sphingomyelin phosphodiesterases (SMPDs), which (within lysosomes) catabolize sphingomyelin, a highly enriched constituent of cholangiocytes [45]. In summary, *O. viverrini* has an extensive repertoire of enzymes and