GENOMICS OF PARASITIC FLATWORMS
(Class Trematoda)

Neil D. Young, Aaron R. Jex, Stefano Mangiola, Ross S. Hall, Pasi Korhonen and Robin B. Gasser

Gasser Laboratory, The University of Melbourne
# Neglected Tropical Disease

- Of the 17 WHO-listed NTDs, 8 are caused by helminths *(nematodes/trematodes/cestodes)*.

<table>
<thead>
<tr>
<th>Buruli Ulcer</th>
<th>Rabies</th>
<th>helminthiasis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chagas disease</td>
<td>Trachoma</td>
<td>Cysticercosis</td>
</tr>
<tr>
<td>Dengue</td>
<td>Yaws</td>
<td>Echinococcosis</td>
</tr>
<tr>
<td>Human African trypanosomiasis</td>
<td>Dracunculiasis</td>
<td>Food-borne trematode infections</td>
</tr>
<tr>
<td>Leishmaniasis</td>
<td>Lymphatic filariasis</td>
<td>Schistosomiasis</td>
</tr>
<tr>
<td>Leprosy</td>
<td>Onchocerciasis</td>
<td>Trematodiasis, caused by parasitic flatworms (Class Trematodota)</td>
</tr>
<tr>
<td>Soil transmitted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Trematodiasis

- **Human parasites**
  - Water-borne: *Schistosoma* spp.

- **Parasites of other mammals**
  - Food-borne: *Fasciola* spp., *Fascioloides magna*, *Dicrocoelium* spp.

- **Reliance on praziquantel**

**INSIGHTS INTO TREMATODE BIOLOGY** BY CHARACTERISING THEIR GENOMES AND TRANSCRIPTOMES
Genomic resources for flatworms
NCBI submissions (Nov 2013)

Deuterostomia
- Echinodermata
  - 5,373,609 proteins
  - 30,314 bioprojects

Chordata
- 385,568 proteins
- 347 bioprojects

Protostomia
- Lophotrochozoa
  - 2,408,587 proteins
  - 2,933 bioprojects

- Ecdysozoa
  - 35%
  - > 35% of proteins unannotated
  - More data needed

- Trematodes/cestodes
- Biologically diverse
- Evolutionarily divergent

Mouse
- C. elegans
- Fly
Update on current programs

- Transcriptome assembly and curation of a trematode sequence database
- Assembly and annotation of trematode genomes (e.g. *Schistosoma haematobium*)
## Published food-borne trematode transcriptomes

<table>
<thead>
<tr>
<th></th>
<th>Opisthoriidae</th>
<th>Fasciolidae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Opisthorchis viverrini</td>
<td>Clonorchis sinensis</td>
</tr>
<tr>
<td>Roche 454</td>
<td>642,918</td>
<td>574,448</td>
</tr>
<tr>
<td>Illumina</td>
<td>22.4 million</td>
<td></td>
</tr>
<tr>
<td>Transcripts</td>
<td>21,026</td>
<td>50,769</td>
</tr>
</tbody>
</table>

**Adult *F. magna* secreted proteins** *(Cantacessi et al. 2012 Mol Cell Proteom 11: 1340)*

**Opisthorchis viverrini** juvenile-enriched peptidases

- Asparaginyl endopeptidase (legumain or C13) – 236 transcripts
- Lysosomal endopeptidase but component of the ES
- Transactivates cathepsin B – but what else???

Published food-borne trematode transcriptomes

<table>
<thead>
<tr>
<th></th>
<th>Opisthorchiidae</th>
<th>Fascioliidae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Opisthorchis</td>
<td>Fasciola</td>
</tr>
<tr>
<td></td>
<td>viverrini</td>
<td>hepatica</td>
</tr>
<tr>
<td>Roche 454</td>
<td>642,918</td>
<td>590,927</td>
</tr>
<tr>
<td>Illumina</td>
<td>22.4 million</td>
<td>21.9 million</td>
</tr>
<tr>
<td>Transcripts</td>
<td>21,026</td>
<td>30,513</td>
</tr>
</tbody>
</table>

Adult *F. magna* secreted proteins (Cantacessi et al. 2012 Mol Cell Proteom 11: 1340)


Curation of “orthologues”, conserved functional domains, gene ontology vocabulary and KEGG biological pathways/protein family classification

Schistosomiasis (bilharziasis/snail fever)

- 200 million infected; > 200,000 deaths p.a.
- Water-borne infection, chronic
- Treatment – praziquantel, no vaccine

Gryseels et al. 2006 Lancet; 368: 1106
Urogenital schistosomiasis
*Schistosoma haematobium*

- Important NTD (Rollinson 2009 Parasitology; 136:1593)
- Difficult to culture/passage (snail host)
- Laboratory infection (hamster – atypical)

Sequence one pair of worms (male/female)

- 400 ng genomic DNA
- Whole genome amplification (WGA)
  - > 20 ug of genomic DNA

<table>
<thead>
<tr>
<th>Insert size (bp)</th>
<th>170</th>
<th>500</th>
<th>800</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small insert</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mate pair</td>
<td>2000</td>
<td>5000</td>
<td>10000</td>
</tr>
</tbody>
</table>

Illumina sequencing (~35 Gb)
*de novo* genome assembly
## Genome assembly

<table>
<thead>
<tr>
<th></th>
<th><em>S. haematobium</em></th>
<th><em>S. mansoni v5</em></th>
<th><em>S. japonicum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaffolds</td>
<td>99,953</td>
<td>885</td>
<td>25,048</td>
</tr>
<tr>
<td>Genome size (Mb)</td>
<td>385</td>
<td>365</td>
<td>403</td>
</tr>
<tr>
<td>N50 scaffold length</td>
<td>306,738</td>
<td>32,100,000</td>
<td>176,868</td>
</tr>
<tr>
<td>Number of genes</td>
<td>13,073</td>
<td>10,852</td>
<td>13,469</td>
</tr>
<tr>
<td>GC content</td>
<td>35.9%</td>
<td>35.3%</td>
<td>34.1%</td>
</tr>
</tbody>
</table>

### Draft *S. haematobium* genome from a single pair of worms

Nature Genetics 2012; 44, 221-225

Whole-genome sequence of *Schistosoma haematobium*

Genome added to SchistoDB: [http://schistodb.net/](http://schistodb.net/)
Transcriptome – annotation and gene expression

RNA-Seq
RPKM = expression

Gene density

Cell growth
(cadherin, innexin, connexin)

Muscle development
(titin, myosin, paramyosin)

Oogenesis
(superoxide dismutase, egg-shell proteins, female specific protein 800)

Tegument
(tetraspanin-1)

Egg proteins (CP391S-like)

Growth/Development
(tubulin, annexin, sphingosine kinase, calcium binding proteins)

Immune-related?
(TNF-associated factor, secreted frizzled-related protein, VALs)
GENOMIC RESOURCES FOR TREMATODES

- **F. hepatica**
  - Transcriptomic
  - Genomic
  - Small RNA

- **F. gigantica**
  - Proteomic

- **S. haematobium**
  - Transcriptomic
  - Genomic
  - Small RNA

- **F. magna**
  - Proteomic

- **S. mansoni**
  - Transcriptomic
  - Genomic

- **S. japonicum**
  - Proteomic

- **C. sinensis**
  - Genomic

- **D. dendriticum**
  - Small RNA

**Classification**

- **Food-borne**
  - Fasciolidae
  - Opisthorchiidae
  - Dicrocoelidae

- **Water-borne**
  - Schistosomatidae
GENOMIC RESOURCES FOR TREMATODES

<table>
<thead>
<tr>
<th>Genus</th>
<th>Transcriptomic</th>
<th>Genomic</th>
<th>Proteomic</th>
<th>Small RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>F. hepatica</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>F. gigantica</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>F. magna</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. haematobium</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. mansoni</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. japonicum</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Food-borne | Oppisthorchiidae | Dicrocoelidae |
Water-borne | Schistosomatidae |

Fasciolidae

“2% of the human genome contains genes........ around 80% of the genome is actively doing something...... involved in controlling when and where proteins are produced”
Ewan Birney ENCODE PROJECT
**COMPARE and CONTRAST**
Vaccine candidates (e.g. Tetraspanin 2)
Helminth proteins (HDMs, SCP/TAPS)
Trematode proteins (AEPs)
Drug discovery/repurposing
Adaptations to life in blood/bile duct

**CONFIRM FUNCTION**
Transcription = Expression
Phenotype (RNAi)
Localisation (IHC)
Enzyme activity
Receptor specificity

New intervention strategies for trematodiasis
Future work

• Sequencing trematode genomes (using WGA)
• Population genetics
• Drugs / vaccines – predict, prioritise and consolidate
• Curate resource for proteomics and functional genomics
• Non-coding elements (conserved/novel)
Acknowledgements

The University of Melbourne
Members of the Gasser Laboratory
BGI-Shenzhen
Jun Wang, Bo Li, Shiping Liu and their bioinformatic teams
Natural History Museum
David Rollinson
Khon Kaen University
Thewarach Laha and Banchob Sripa
Gyeongsang National University
Woon-Mok Sohn

George Washington University
Gabriel Rinaldi and Paul Brindley
Fundação Oswaldo Cruz (FIOCRUZ)
Adhemar Zerlotini and Guilherme Oliveira
James Cook University
Cinzia Cantacessi and Alex Loukas
Griffith University
Andreas Hoffman

FUNDING/SUPPORT
Australian Research Council (ARC-linkage)
NHMRC (Early Career Research Fellow, Project Grant)
Victorian Life Sciences Computation Initiative (VLSCI)

Visit the Gasser lab: http://www.gasserlab.org/