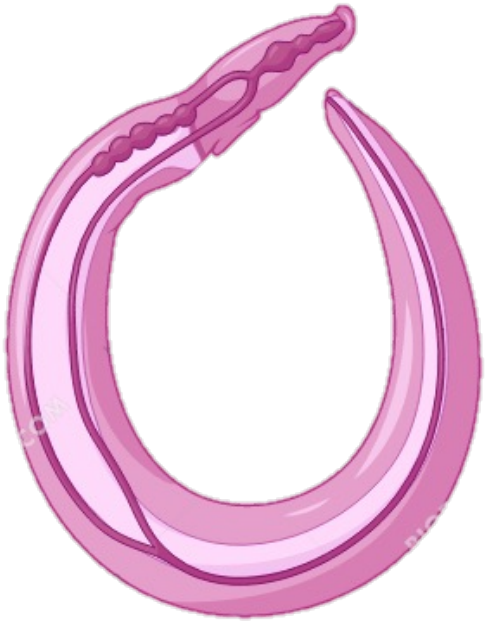
A look into the genome of *Schistosoma haematobium*

Ali Ahmad

Introduction to genomics (485)

20/1/2022



Outline

- Selected paper
- *S.haematobium:*
    - Taxonomy
    - Physical characteristics
    - Parasitism
    - Geographic distribution
    - Life cycle
    - Significance
- Sequenced sample info
- Sequencing strategy
- Sequencing method
- Genome assembly
- Interesting genome outcome

Selected paper

nature genetics

Whole-genome sequence of *Schistosoma haematobium*

Neil D Young[1,11], Aaron R Jex[1,11], Bo Li[2,11], Shiping Liu[2], Linfeng Yang[2], Zijun Xiong[2], Yingrui Li[2], Cinzia Cantacessi[1], Ross S Hall[1], Xun Xu[2], Fangyuan Chen[2], Xuan Wu[2], Adhemar Zerlotini[3], Guilherme Oliveira[3], Andreas Hofmann[1,4], Guojie Zhang[2], Xiaodong Fang[2], Yi Kang[2], Bronwyn E Campbell[1], Alex Loukas[5], Shoba Ranganathan[6,7], David Rollinson[8], Gabriel Rinaldi[9,10], Paul J Brindley[10], Huanming Yang[2], Jun Wang[2], Jian Wang[2] & Robin B Gasser[1]

Schistosomiasis is a neglected tropical disease caused by blood flukes (genus *Schistosoma*; schistosomes) and affecting 200 million people worldwide[1]. No vaccines are available, and treatment relies on one drug, praziquantel[2]. *Schistosoma haematobium* has come into the spotlight as a major cause of urogenital disease, as an agent linked to bladder cancer[1,3] and as a predisposing factor for HIV/AIDS[4,5]. The parasite is transmitted to humans from freshwater snails[1]. Worms dwell in blood vessels and release eggs that become embedded in the bladder wall to elicit chronic immune-mediated disease[6] and induce squamous cell carcinoma[7]. Here we sequenced the 385-Mb genome of *S. haematobium* using Illumina-based technology at 74-fold coverage and compared it to sequences from related parasites[8,9]. We included genome annotation based on function, gene ontology, networking and pathway mapping. This genome now provides an unprecedented resource for many fundamental research areas and shows great promise for the design of new disease interventions.

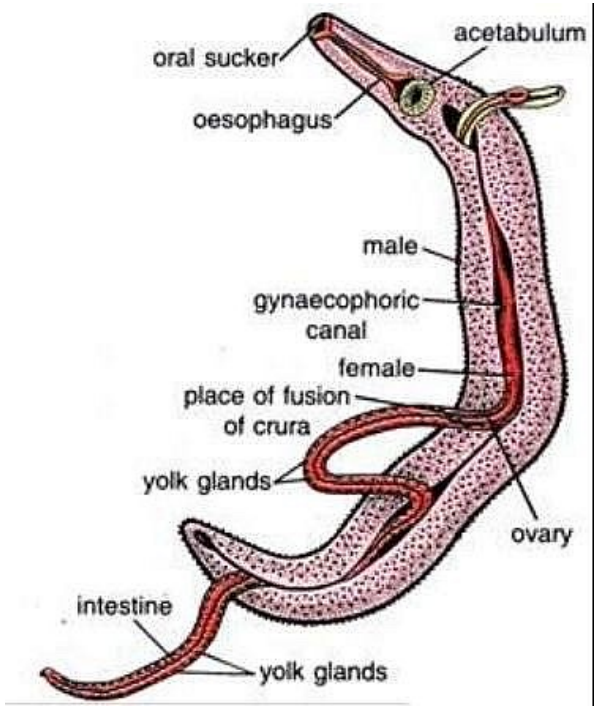
amplification, a result that is consistent with published information[11,12] (**Supplementary Note, Supplementary Figs. 3** and **4**).

Comparison of the *S. haematobium* and *S. mansoni* genomes showed a similar percentage and composition of repetitive elements (**Table 1**, **Supplementary Note**, **Supplementary Tables 4–6**). Using both homology-based and *de novo* predictions, we estimated that 43% of the *S. haematobium* genome comprises repetitive elements, consistent with the *S. mansoni* genome (40%)[8]. More than half (58.5%) of the repeats were retrotransposons (at least 20 types, including LINE/RTE-BovB and LTR/*Gypsy*); 37% were unknown repeats, including satellites (1.9%), simple repeats (1.2%) and DNA transposons (five types; <1%). On the basis of homology, *de novo* predictions and evidence of transcription (in adult and egg stages), we inferred 13,073 protein-coding genes from the genome and included data for *S. mansoni* and *S. japonicum* for comparisons (**Supplementary Note**, **Supplementary Table 7**, **Supplementary Figs. 5–7**). The number of *S. haematobium* genes was consistent with those of *S. mansoni* (13,184) and *S. japonicum* (13,469), as were the gene structures. Most (9,714)

Taxonomy

- SK: Eukaryota
  - K: Metazoa
    - P: Platyhelminthes
      - C: Trematoda
        - O: Strigeidida
          - F: Schistosomatidae
            - G: *Schistosoma*
              - S: *haematobium*

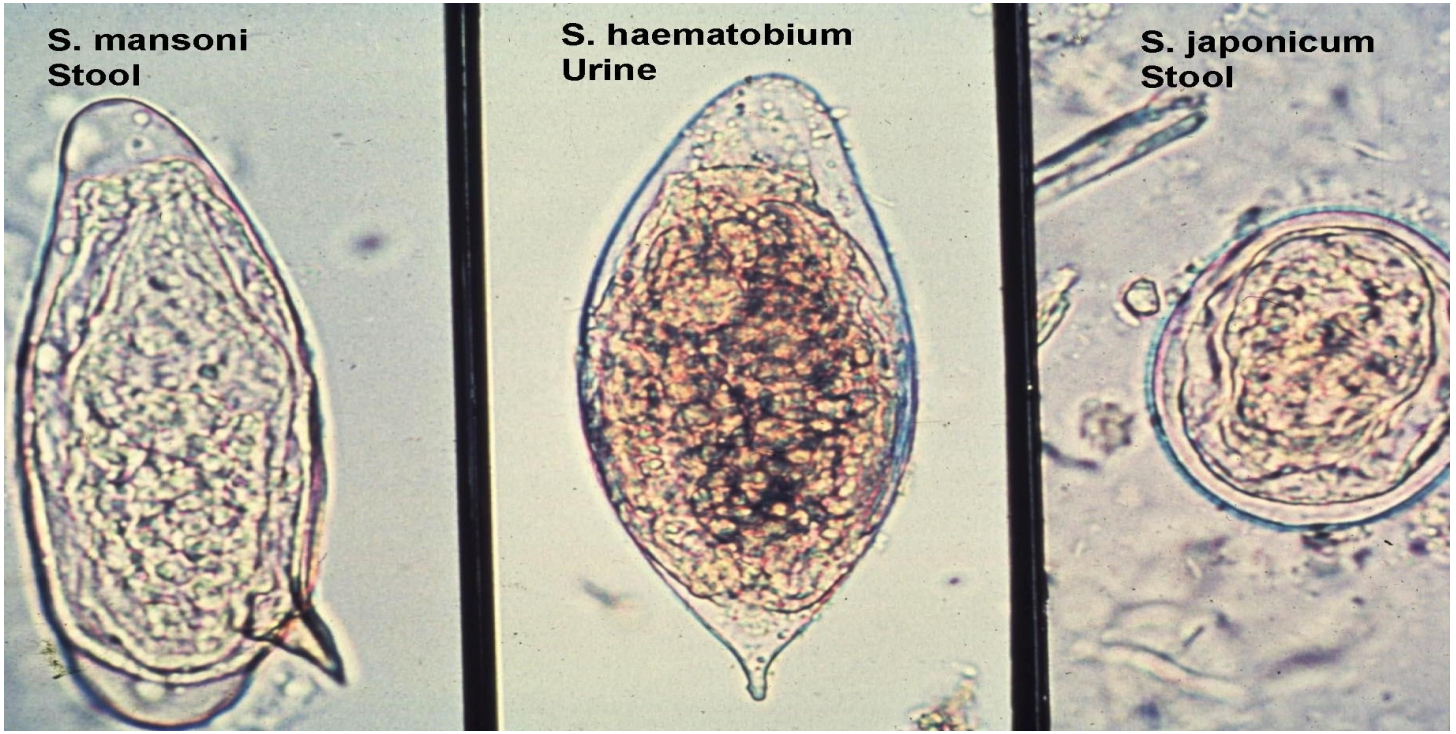
Physical characteristics



	S. mansoni	*S. japonicum*	*S. haematobiun*
Adult male			
• Length (mm)	6–12	12–20	10–14
• Breadth (mm)	2.00	0.50–0.55	0.75–1.00
• No of testes	4–13	6–9	4–5
Adult female			
• Length (mm)	7–17	16–28	16–20
• Breadth (mm)	1.00	0.30	0.25
• No of eggs in uterus	Usually 1	50 or more	10–100

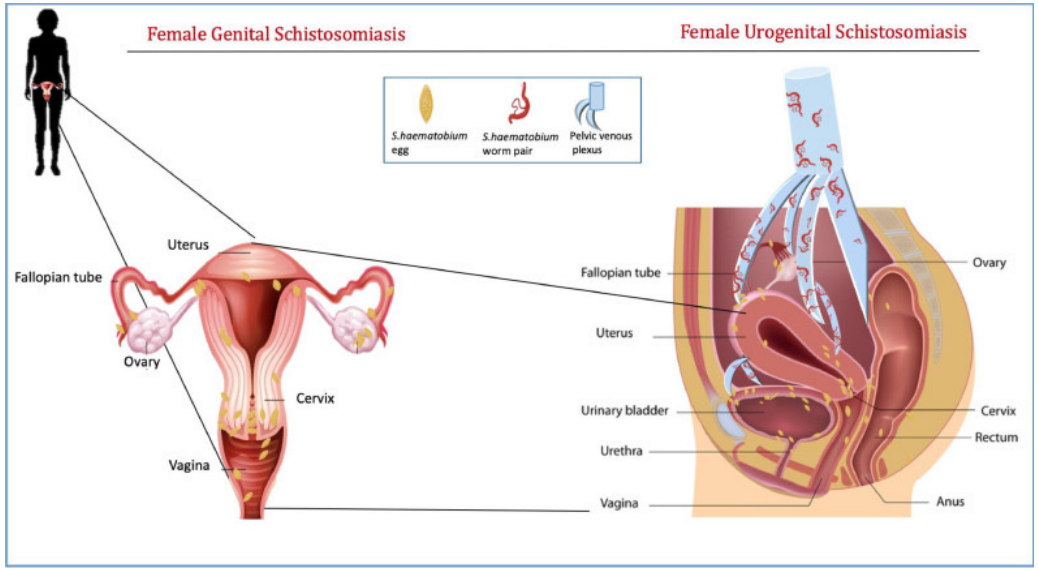
Physical Characteristic

- *S.haematobium* eggs have an oval shape with conspicuous terminal spine
- Size: 110-170 µm long by 40-70 µm wide

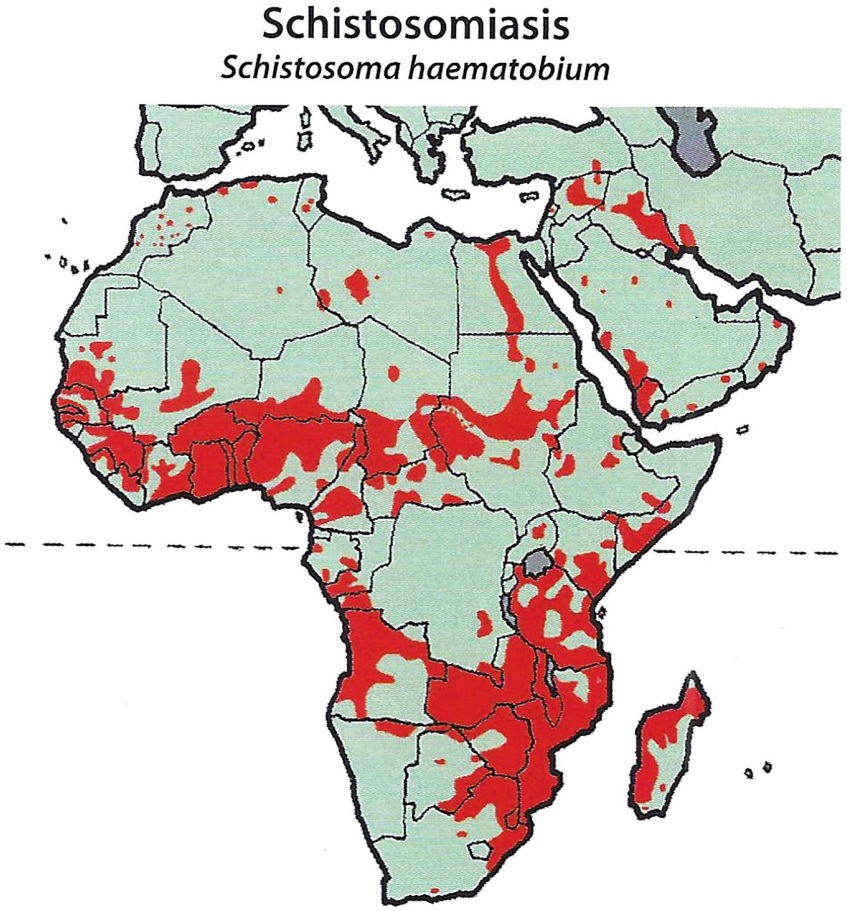


Parasitism

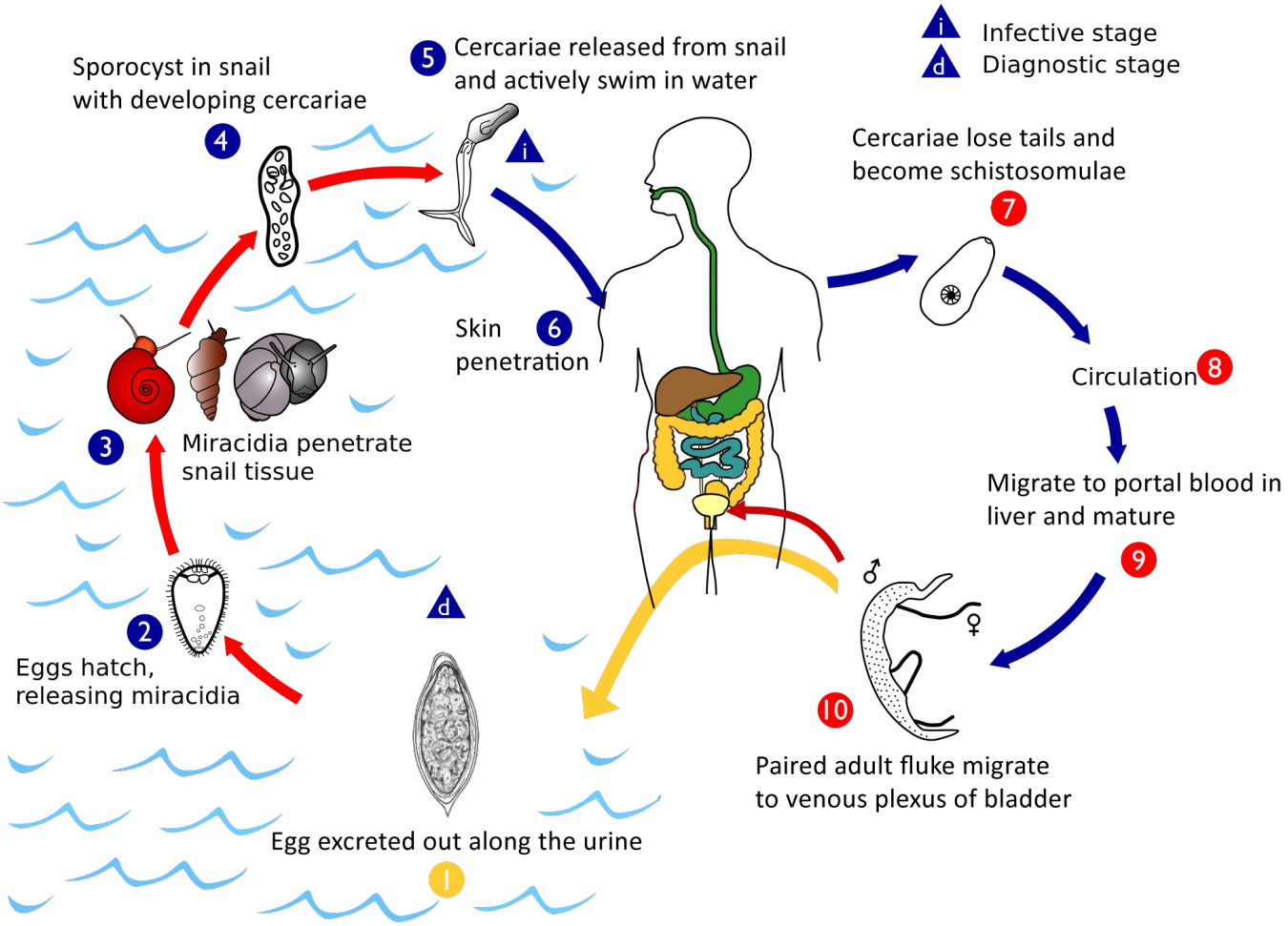
- Infect 200 million people worldwide with schistosomiasis
- Bladder cancer and urogenital disease agents
- Intermediate host: freshwater snails
- Definitive host: human



Geographic distribution



Life cycle


i Infective stage
d Diagnostic stage
1 Egg excreted out along the urine
2 Eggs hatch, releasing miracidia
3 Miracidia penetrate snail tissue
4 Sporocyst in snail with developing cercariae
5 Cercariae released from snail and actively swim in water
6 Skin penetration
7 Cercariae lose tails and become schistosomulae
8 Circulation
9 Migrate to portal blood in liver and mature
10 Paired adult fluke migrate to venous plexus of bladder
♂
♀

Significance

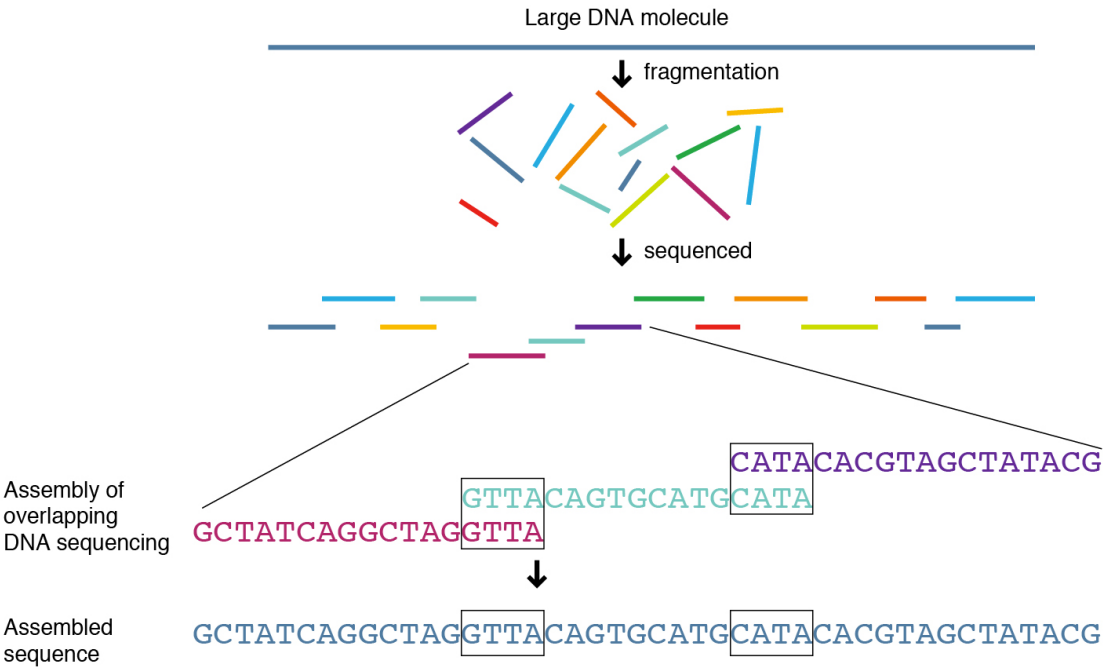
- The sequence were done to provide a library data bases of genome that can be later use in control several diseases cased by *S.haematobium* and an unprecedented resource for many fundamental research areas.

Sequencing sample information

- Hamsters were each infected with 1,000 cercariae and after 90 days paired adults of *S.haematobium* were collected from the hamsters.

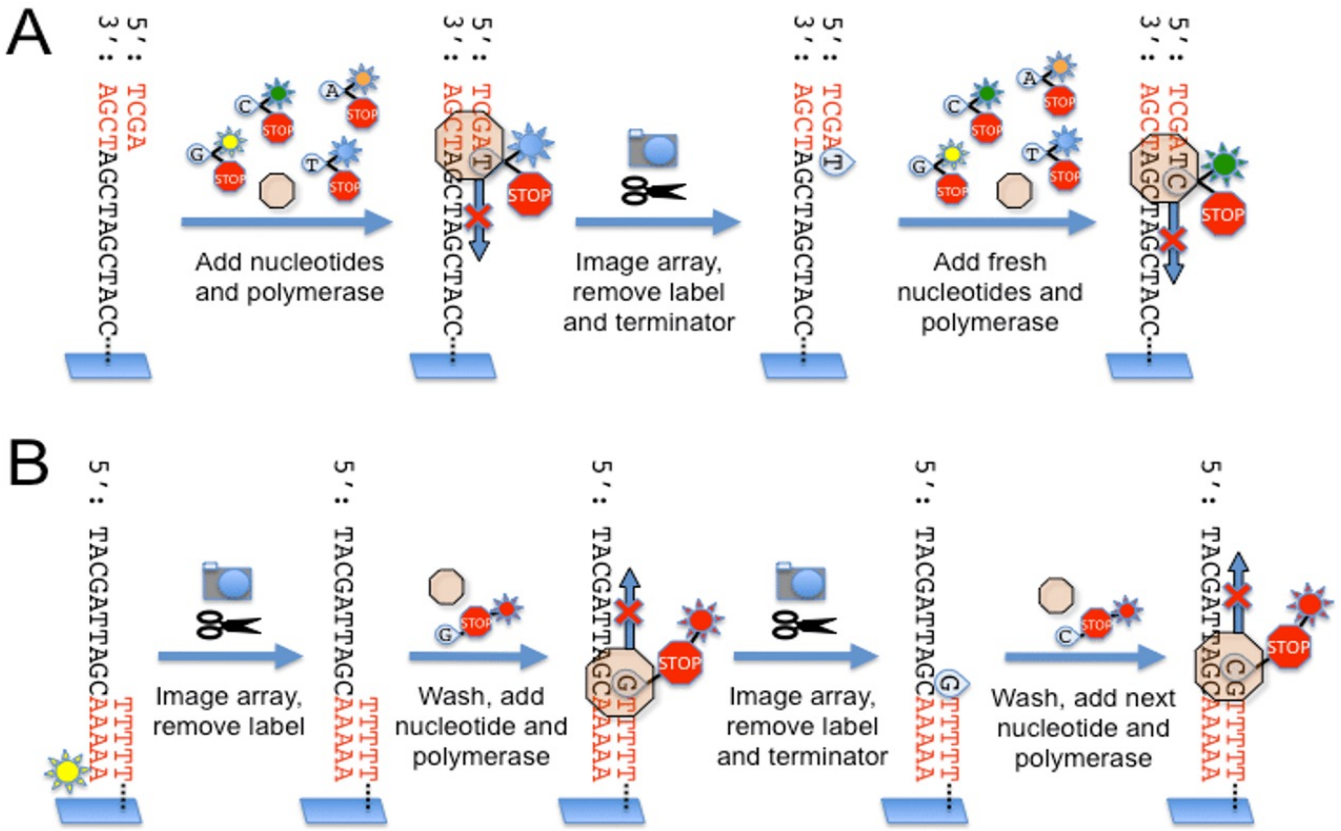
Sequencing strategy

- Whole genome shotgun



Sequencing method

- Illumina



Genome assembly

Table 1 Comparison of the *Schistosoma haematobium* genome with those of *S. mansoni* and *S. japonicum*

Genomic features	*Schistosoma haematobium*	*Schistosoma mansoni*	*Schistosoma japonicum*
Estimate of genome size (Mb)	385	381	403
Chromosome number (2*n*)	8[a]	8	8
Total number of base pairs within assembled contigs	361,903,089	374,944,597	369,039,322
N50 contig (length (bp); total number >500 bp)	21,744; *n* = 36,826	16,320; *n* = 50,292	6,121; *n* = 95,265
Total number of base pairs within assembled scaffolds	385,110,549	381,096,674	402,705,545
N50 scaffold (length (bp); total number >1,000 bp in length)	306,738; *n* = 7,475	832,5415; *n* = 19,022	176,869; *n* = 25,048
Proportion of genome that is coding (%)	4.43	4.72	4.32
Number of putative coding genes	13,073	13,184	13,469
Gene size (average bp ± s.d.; range)	11,952 ± 16,273; 30–20	13,397 ± 18,029; 84–240,193	10,003 ± 12,980; 150–173,394
Average coding domain length (average bp ± s.d.; range)	1,319 ± 1,502; 30–28,212	1,344 ± 1,447; 60–22,983	1,179 ± 1,201; 147–24,180
Average exon number per gene (average bp ± s.d.; range)	5.4 ± 5.80; 1–136	6.2 ± 6.24; 1–94	5.3 ± 4.70; 1–65
Gene exon length (average bp ± s.d.; range)	246 ± 287; 1–9,737	218 ± 236; 1–9,291	223 ± 256; 6–6,326
Gene intron length (average bp ± s.d.; range)	2,442 ± 2,958; 1–68,754	2,331 ± 3,200; 1–67,221	2,058 ± 2,679; 15–59,770
Total GC content (%)	34.3	34.7	33.5
Repeat rate (%)	47.2	45	40.1

[a]Estimate derived from a karyological study of *S. haematobium*[16].

- Coverage = 74x
- Genome size = 385 Mb
- N50 contig = 21 Kb
- N50 scaffolds = 306 Kb

Interesting genome outcome

- High synteny between *S.haematobium* and *S.mansoni* (89.4%) were revealed by a genome-wide analysis.
- 33.5 Gb of usable sequence data were produced from a single mated pair of adult worms.
- The GC content of *S.haematobium* were almost the same to that from *S.mansoni* and *S.japonicum.*

Questions

1. What is synteny ?

2. Why illumina method were used ?

Thank you for listening