



Genome of the human hookworm *N.americanus*



Motluq Alobuied

Introduction to genomics (485)

Thursday 20/1/2022

Outline

- Genome of the human hookworm *N. americanus*
- *N. americanus*:
 - Taxonomy (classification)
 - Physical characteristic
 - Parasitism
 - Geographic distribution
 - Life cycle
 - Significance
- sequenced sample info
- Sequencing strategy
- Sequencing method
- Genome assembly
- Interesting genome outcome

Genome of the human hookworm *N.americanus*

nature
genetics

OPEN

Genome of the human hookworm *Necator americanus*

Yat T Tang^{1,16}, Xin Gao^{1,16}, Bruce A Rosa^{1,16}, Sahar Abubucker¹, Kymberlie Hallsworth-Pepin¹, John Martin¹, Rahul Tyagi¹, Esley Heizer¹, Xu Zhang¹, Veena Bhonagiri-Palsikar¹, Patrick Minx¹, Wesley C Warren^{1,2}, Qi Wang¹, Bin Zhan^{3,4}, Peter J Hotez^{3,4}, Paul W Sternberg^{5,6}, Annette Dougall⁷, Soraya Torres Gaze⁷, Jason Mulvenna⁸, Javier Sotillo⁷, Shoba Ranganathan^{9,10}, Elida M Rabelo¹¹, Richard K Wilson^{1,2}, Philip L Felgner¹², Jeffrey Bethony¹³, John M Hawdon¹³, Robin B Gasser¹⁴, Alex Loukas⁷ & Makedonka Mitreva^{1,2,15}

The hookworm *Necator americanus* is the predominant soil-transmitted human parasite. Adult worms feed on blood in the small intestine, causing iron-deficiency anemia, malnutrition, growth and development stunting in children, and severe morbidity and mortality during pregnancy in women. We report sequencing and assembly of the *N. americanus* genome (244 Mb, 19,151 genes). Characterization of this first hookworm genome sequence identified genes orchestrating the hookworm's invasion of the human host, genes involved in blood feeding and development, and genes encoding proteins that represent new potential drug targets against hookworms. *N. americanus* has undergone a considerable and unique expansion of immunomodulator proteins, some of which we highlight as potential treatments against inflammatory diseases. We also used a protein microarray to demonstrate a postgenomic application of the hookworm genome sequence. This genome provides an invaluable resource to boost ongoing efforts toward fundamental and applied postgenomic research, including the development of new methods to control hookworm and human immunological diseases.

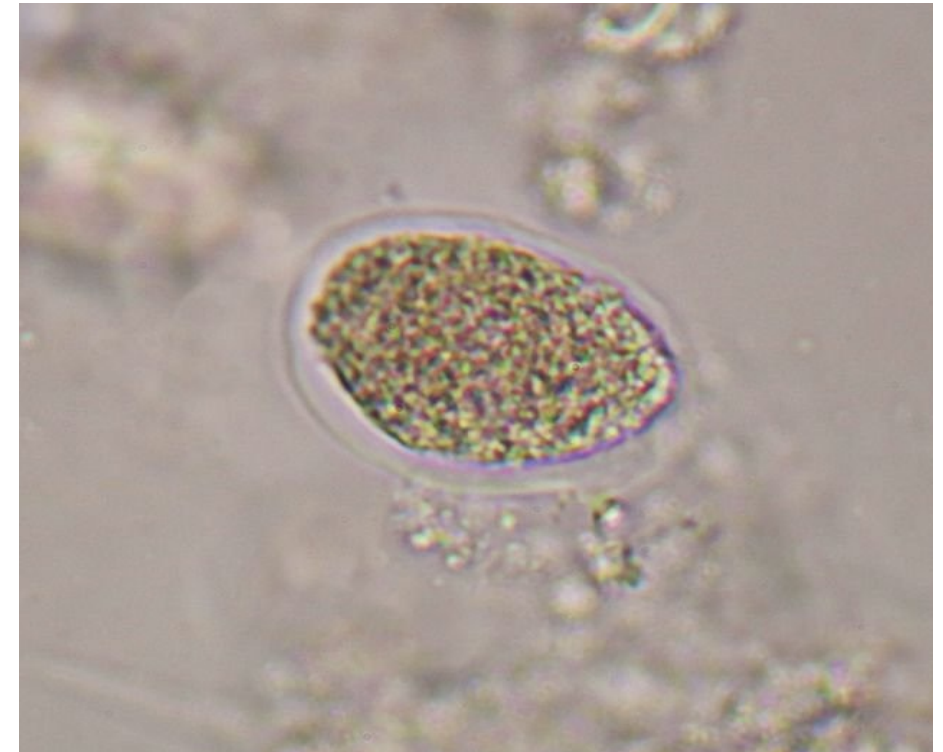
N.americans: Taxonomy

("American murderer"):

- K: Metazoa
 - P: Nematoda
 - C: Chromadorea
 - O: Strongylida
 - F: Uncinariidae
 - G: Necator
 - S: Necator americanus

Physical characteristic

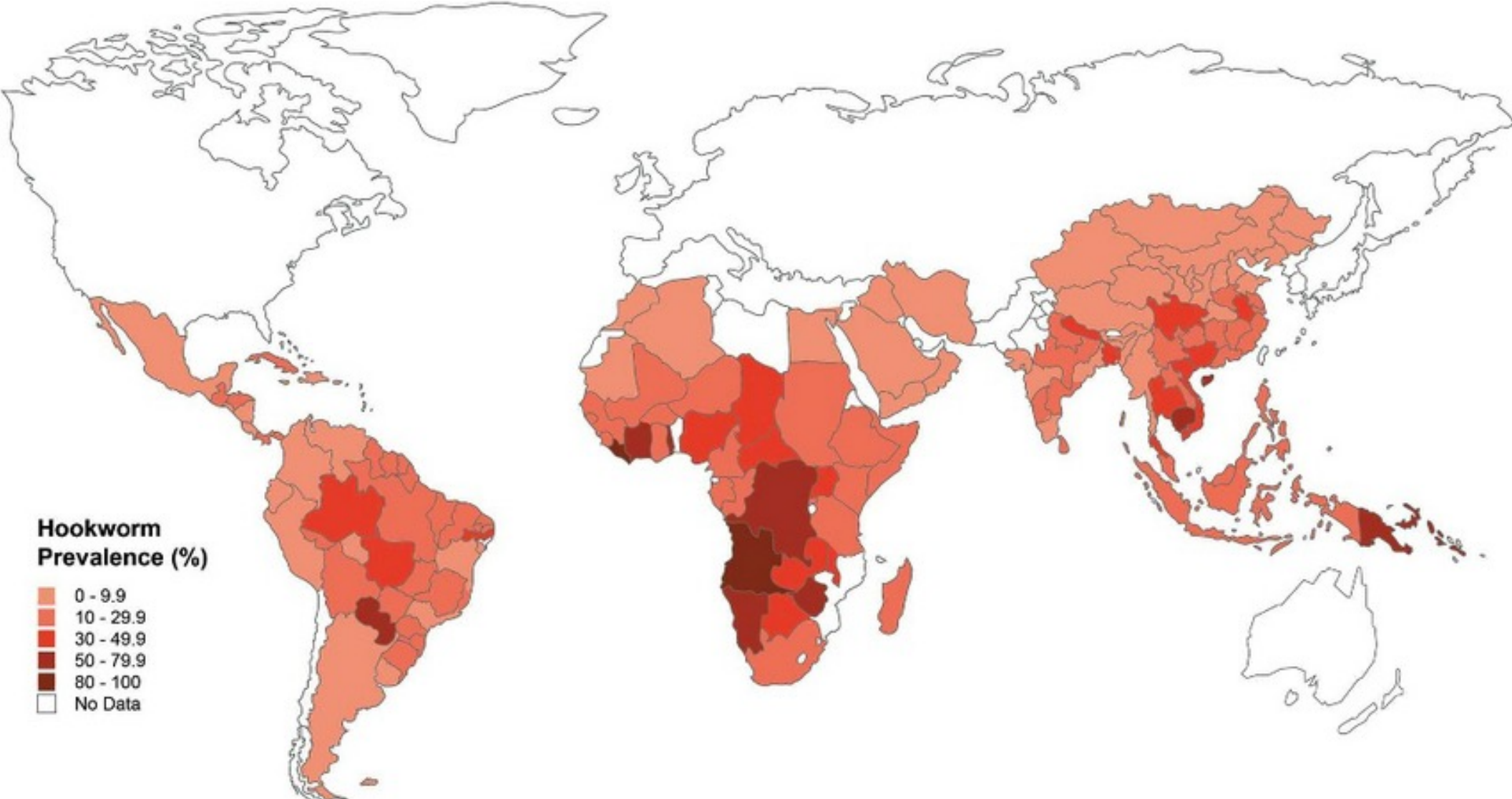
- Adult male worm measures 7mm to 9mm long by 0.4-0.5 mm wide ,
- the female is 9 to 11mm long by 0.4-0.5 mm wide.(both around 1cm long)
- Eggs have a thin shell and are oval shaped, measuring ~56-74 long by 36-40 mm wide.



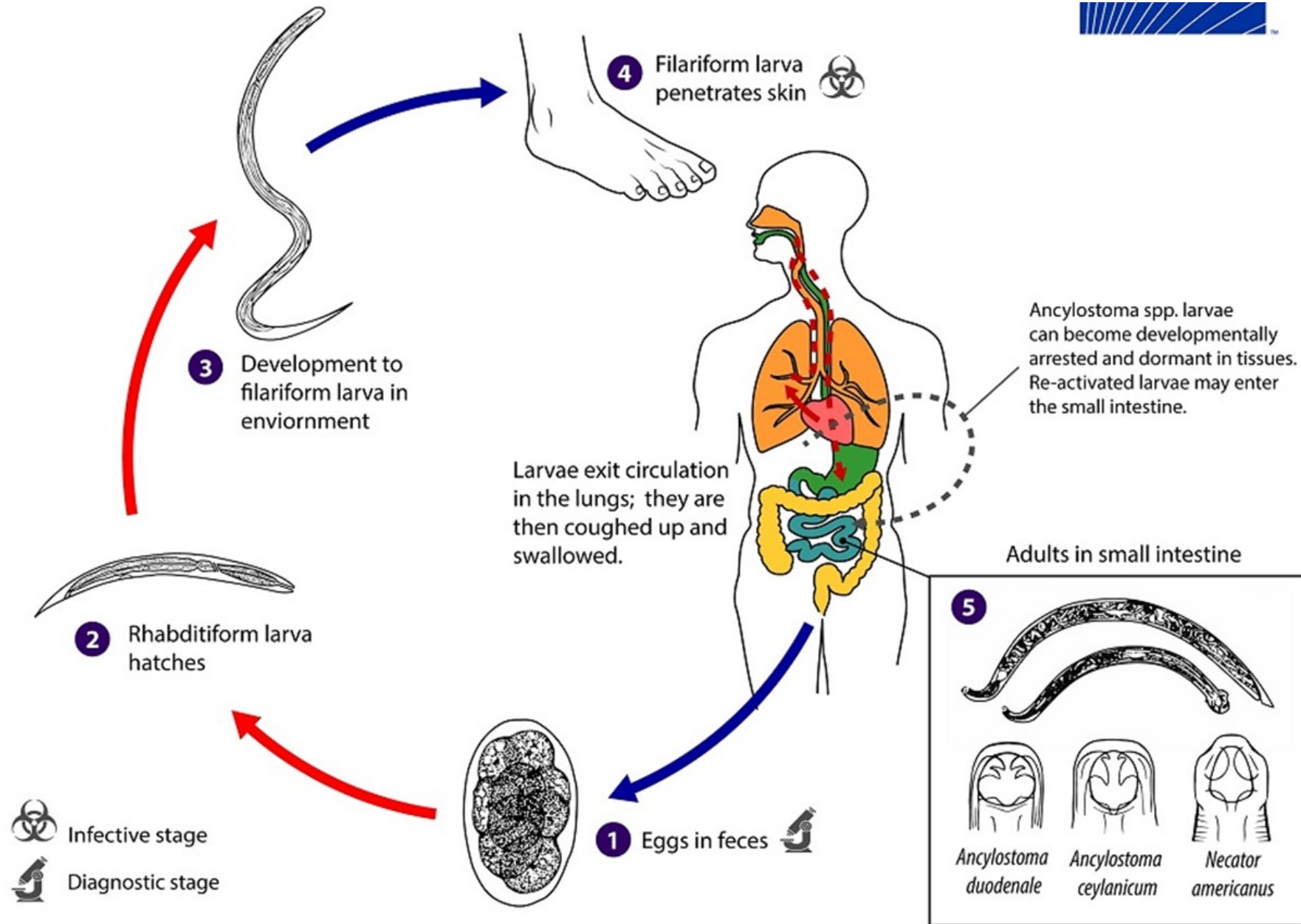
Parasitism

- infect the human host by skin penetration and travel via the circulation to the lungs.
- The iL3 migrate via the trachea to the oropharynx, after which they are swallowed and travel to the small intestine, where they develop to become dioecious adults.
- The adult worms feed on blood (up to 30 μ l per day per worm), and can survive in the human host for up to a decade.

Geographic distribution



Life cycle



Significance

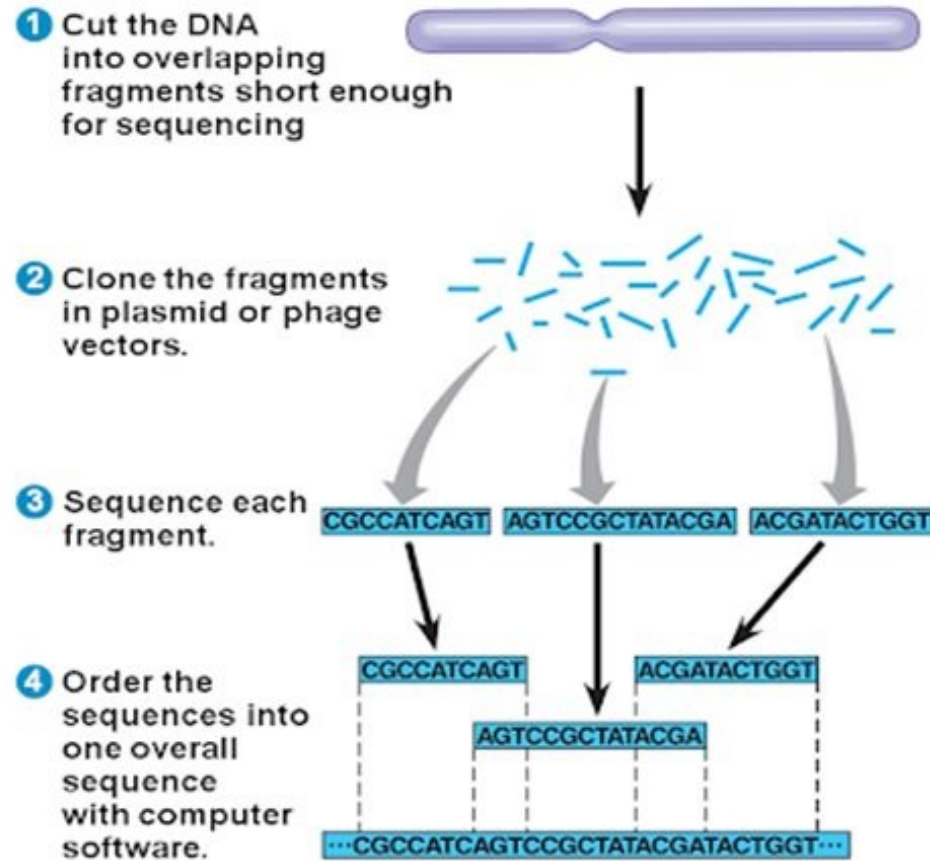
- hookworms suppress the production of pro-inflammatory molecules and promote anti-inflammatory and wound-healing properties
- hookworm recombinant proteins have been tested in clinical trials for non-infectious diseases

sequenced sample info

- Adult worms were collected from intestines of hamsters infected subcutaneously with *N. americanus* iL3 for 8 weeks
- Causes necatoriasis, impairment of cognitive/ physical development in children and malnutrition in pregnant woman

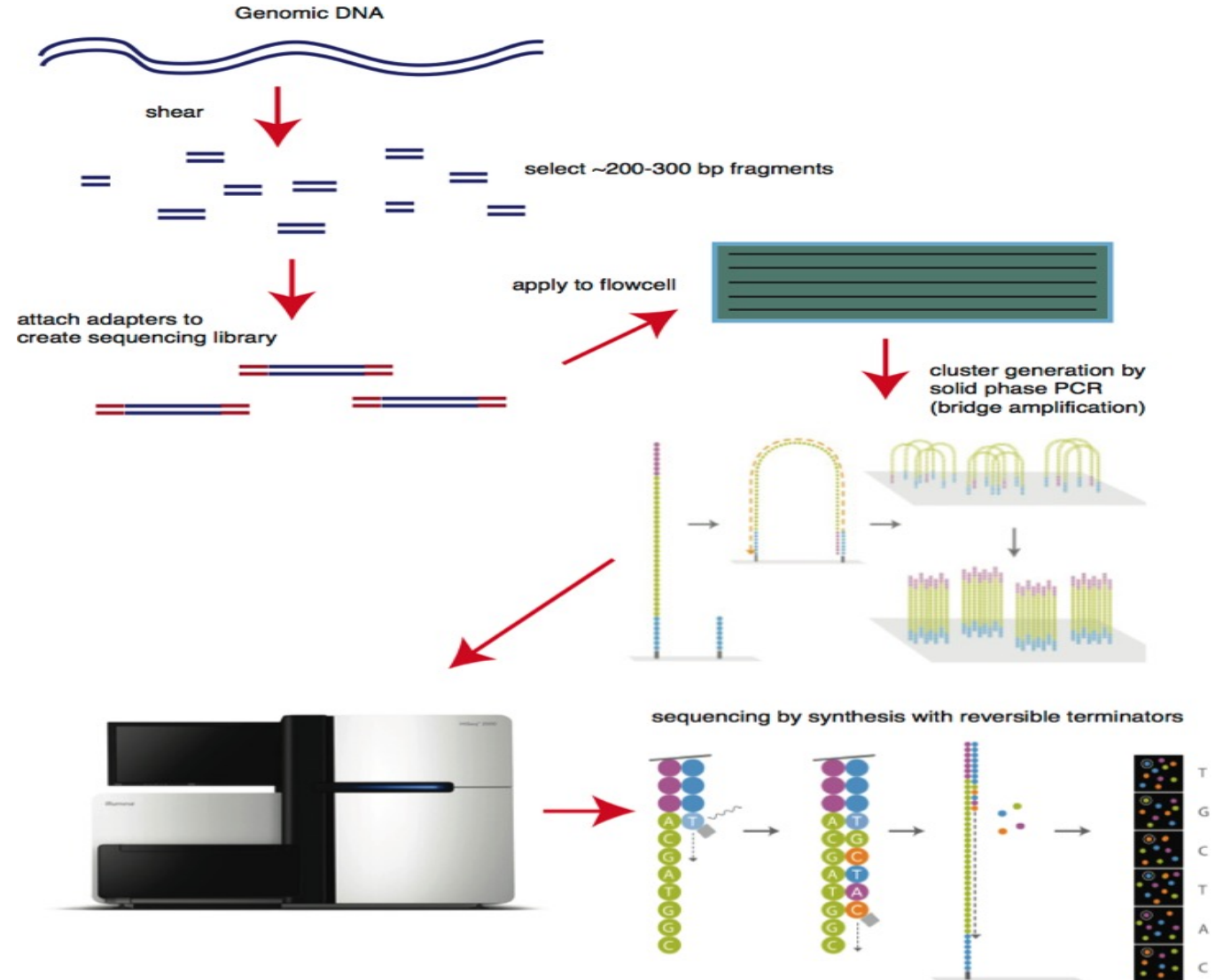
Sequencing strategy

- Paired end whole genome shotgun libraries (3Kb and 8Kb inserts sizes)



Sequencing method

- Illumina was used to sequence cDNA
- Use special ni (has diff colored florescent tag + terminator)
- In this case, Terminator is reversible



Genome assembly

Table 1 Summary of *N. americanus* genomic features

Estimated genome size (Mb)	244
Assembly statistics	
Total number of supercontigs (≥1 kb)	11,713
Total number of base pairs (bp) in supercontigs	244,009,025
Number of N50 supercontigs ^a	283
N50 supercontig length (bp) ^a	213,095
Number of N90 supercontigs ^a	1,336
N90 supercontig length (bp) ^a	29,214
GC content of whole genome	40.20%
Repetitive sequences	23.50%
Protein-coding loci	
Total number of protein-coding genes	19,151
Avg. gene locus footprint (bp)	4,289
Avg. number of exons per gene	6.4
Avg. exon size (bp)	125
Avg. intron size (bp)	642
Avg. intergenic space (bp)	6,631

^aN50 and N90 respectively denote 50% and 90% of all nucleotides in the assembly. 50% of the genome is in 283 supercontigs and in supercontigs with a minimum length of 213 kb; 90% of the genome is in 1,336 supercontigs and in supercontigs with a minimum length of 29 kb.

Interesting genome outcome

- identified molecules that have low similarity to proteins in other species but are recognized by all infected individuals and therefore have high diagnostic potential