No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*

Georgios Koutsovoulosa, Sujai Kumara, Dominik R. Laetschа,b, Lewis Stevensa, Jennifer Dauba, Claire Conlona, Habib Maroona, Fran Thomasa, Aziz A. Aboobakerc, and Mark Blaxtera
PNAS 2016 113 (18) 5053-5058

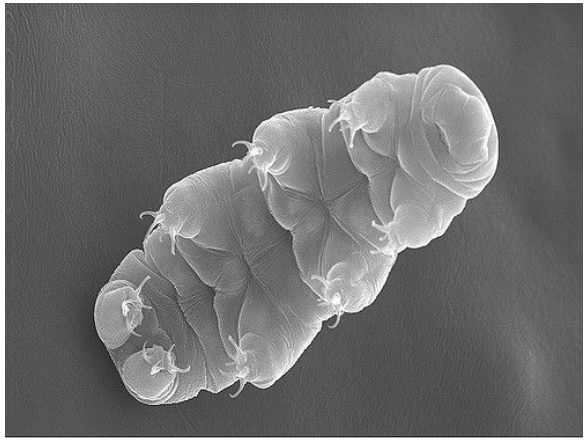
Presentation by Olivia Dai, Ben Hong, and Max Jentzsch

Outline

1. Background
2. Compare/contrast methods
3. Discussion of claims

What are tardigrades?

- Also known as water bears or moss piglets
- Microscopic animals with ability to survive under extreme conditions through a process called cryptobiosis:
  - Temperatures ranging from -272 to 151 ˚C
  - Extreme radiation intensities
  - Deep space vacuum
- Cryptobiotic mechanism has significant implications for biotechnology and medicine



https://s-media-cache-ak0.pinimg.com/736x/dd/34/bc/dd34bc96ff9032d00950eec3fdf3532e.jpg
https://c2.staticflickr.com/2/1231/1138945495_2d77a33baa_z.jpg?zz=1

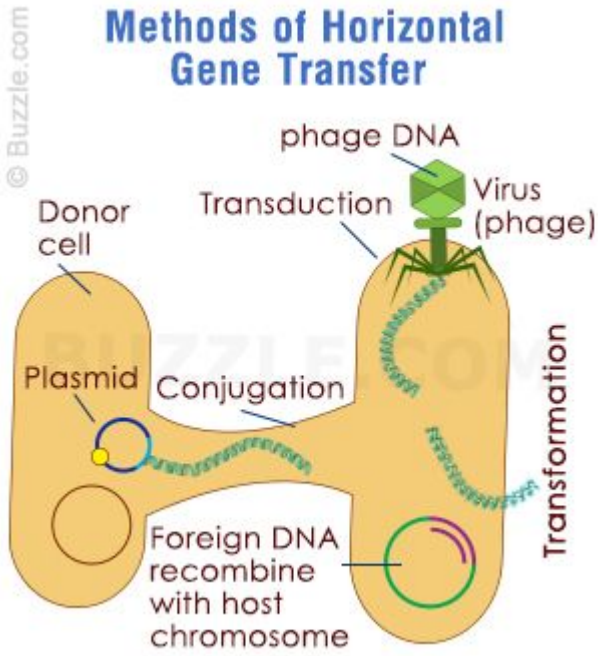
Horizontal Gene Transfer

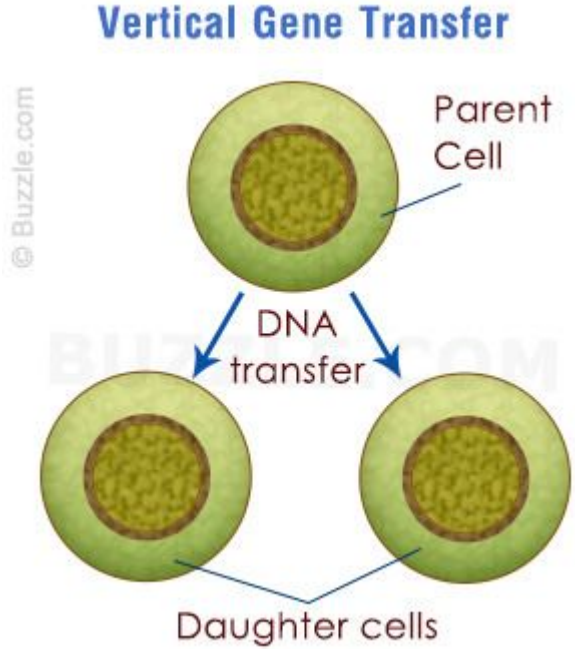
- Transfer of genetic material between unrelated individuals
- Fuels pathogen evolution, which is responsible for spread of antibiotic resistance
- Natural mechanisms:
  - Transduction
  - Transformation
  - Conjugation

Vertical Gene Transfer

- Transmission of genetic material from parent to offspring during reproduction
- Natural mechanism:
  - Reproduction

Horizontal vs. Vertical Gene Transfer





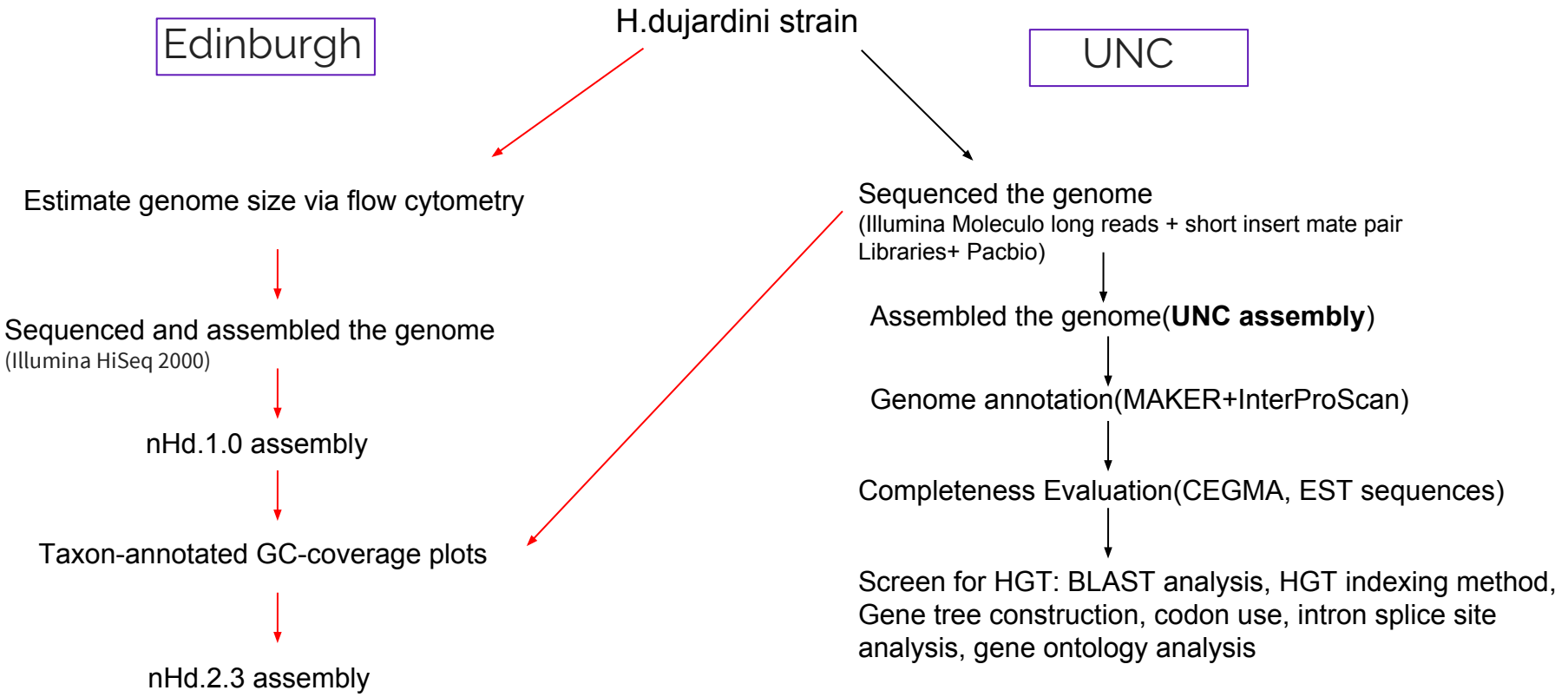
http://www.buzzle.com/images/diagrams/horizontal-gene-transfer.jpg
http://www.buzzle.com/images/diagrams/vertical-gene-transfer.jpg

Results of a previous publication (UNC)

- Boothby et al (2015)
- Sequenced the entire genome
- Determined proportion of fHGT: 17%!

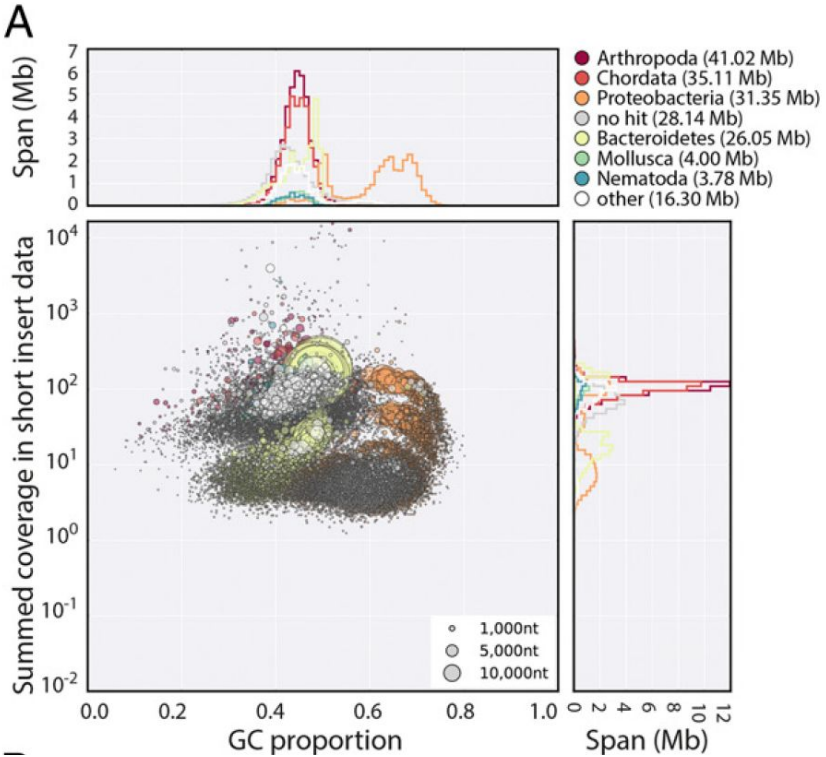
… but wait! This challenges widely accepted notions of the phylogenetic independence of animal genomes and changes how we normally think of evolutionary biology. That's weird! This is where the Edinburgh paper comes along.

Methods


H.dujardini strain
Edinburgh
UNC
Estimate genome size via flow cytometry
Sequenced and assembled the genome
(Illumina HiSeq 2000)
nHd.1.0 assembly
Taxon-annotated GC-coverage plots
nHd.2.3 assembly
Sequenced the genome
(Illumina Moleculo long reads + short insert mate pair Libraries+ Pacbio)
Assembled the genome(**UNC assembly**)
Genome annotation(MAKER+InterProScan)
Completeness Evaluation(CEGMA, EST sequences)
Screen for HGT: BLAST analysis, HGT indexing method, Gene tree construction, codon use, intron splice site analysis, gene ontology analysis

Raw data contained many contaminants

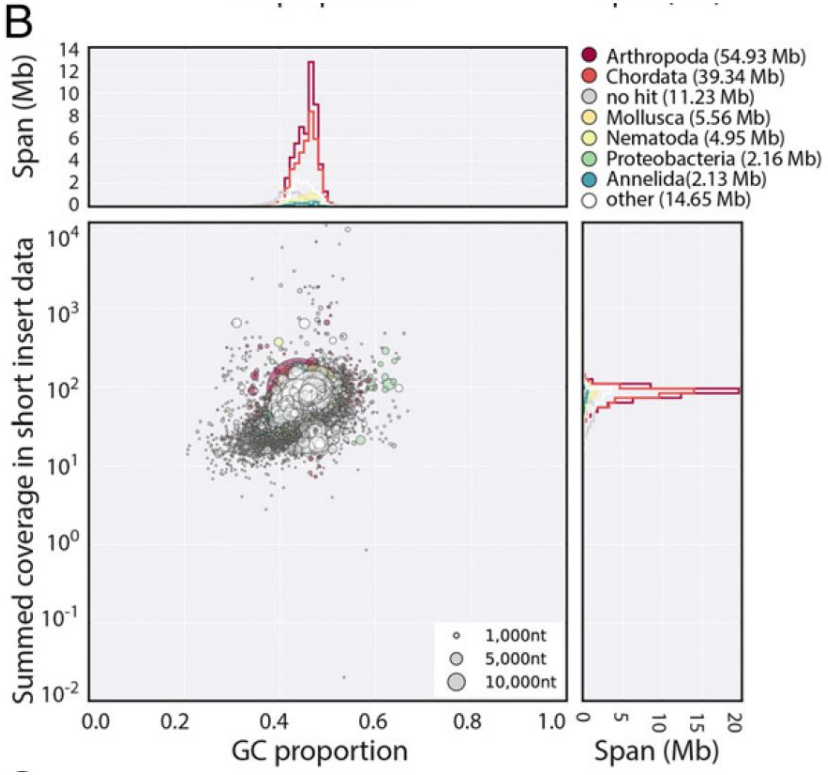
nHd.1.0 Assembly



Koutsovoulos et al (2016)

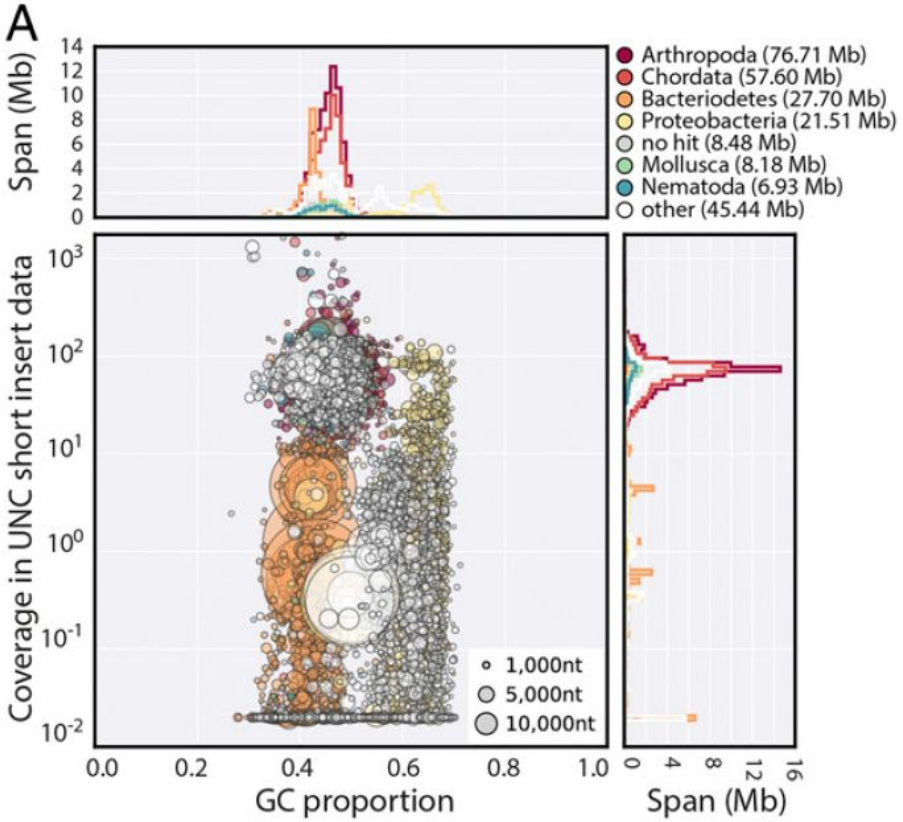
Cleaned data

nHd.2.3 Assembly

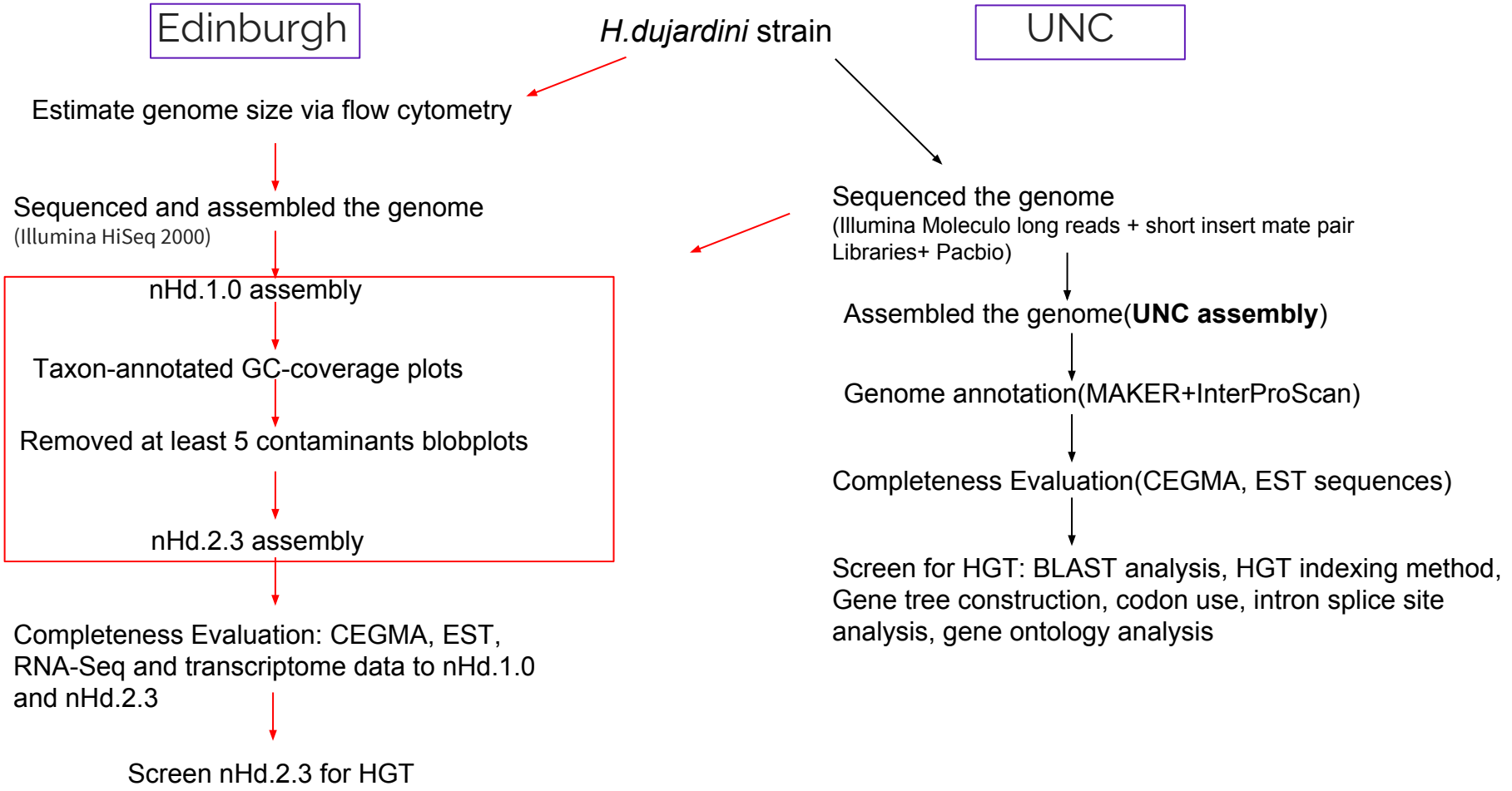


Koutsovoulos et al (2016)

UNC assembly



Koutsovoulos et al (2016)


Edinburgh
H.dujardini strain
UNC
Estimate genome size via flow cytometry
Sequenced and assembled the genome
(Illumina HiSeq 2000)
nHd.1.0 assembly
Taxon-annotated GC-coverage plots
Removed at least 5 contaminants blobplots
nHd.2.3 assembly
Completeness Evaluation: CEGMA, EST, RNA-Seq and transcriptome data to nHd.1.0 and nHd.2.3
Screen nHd.2.3 for HGT
Sequenced the genome
(Illumina Moleculo long reads + short insert mate pair Libraries+ Pacbio)
Assembled the genome(UNC assembly)
Genome annotation(MAKER+InterProScan)
Completeness Evaluation(CEGMA, EST sequences)
Screen for HGT: BLAST analysis, HGT indexing method, Gene tree construction, codon use, intron splice site analysis, gene ontology analysis

Sequencing

Table 1. *H. dujardini* assembly comparison

Genome assembly	nHd.2.3	UNC (13)
Scaffold metrics		
No. scaffolds	13,202	22,497
Span (Mb)	134.96	252.54*
Min length (bp)	500	2,000
N50 length (bp)	50,531	15,907
Scaffolds in N50	701	4,078
GC proportion	0.452	0.469
Quality assessment		
CEGMA completeness	97.2%	94.8%
CEGMA average copies	1.55	3.52
RNA-Seq mapping	92.8%	89.5%
Genome content		
Protein-coding genes	23,021	39,532*
Contaminant span (Mb)	1.5 (1.1%)	68.9 (27.3%)
Initial bacterial HGT loci	554	6,663
Bacterial contaminants	355	9,872†
HGT with expression	196	NA

Koutsovoulos et al (2016)

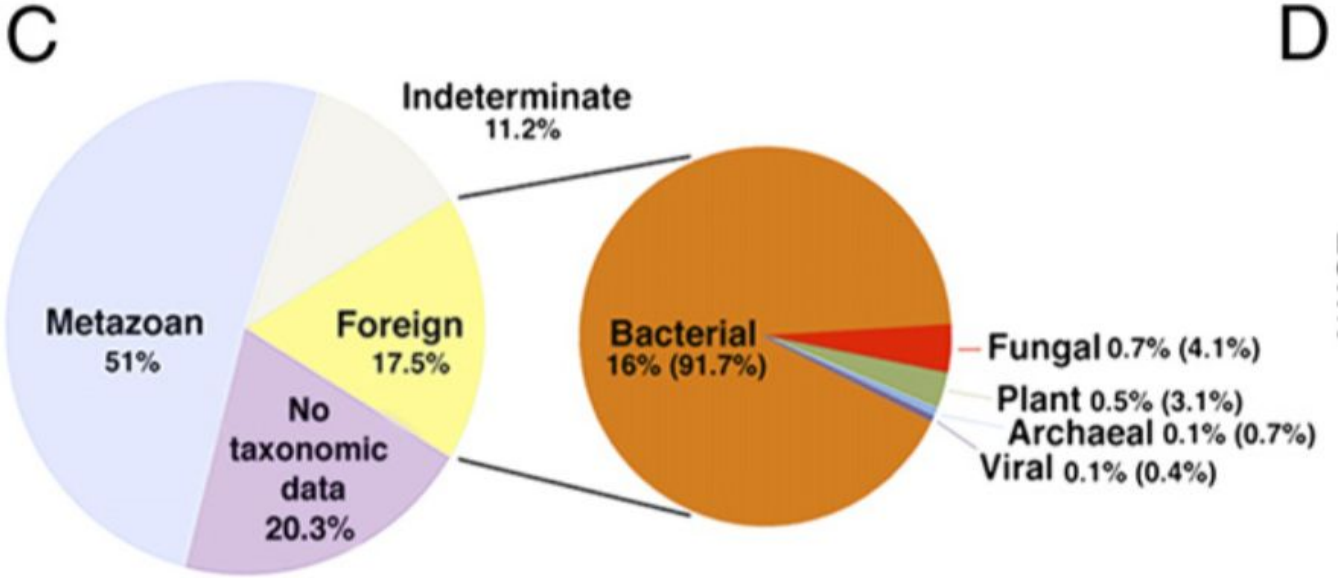
Claims to be refuted

1. HGT index calculations revealed bacterial sequence origins

2. Phylogenies of candidate genes were from non-metazoan taxa

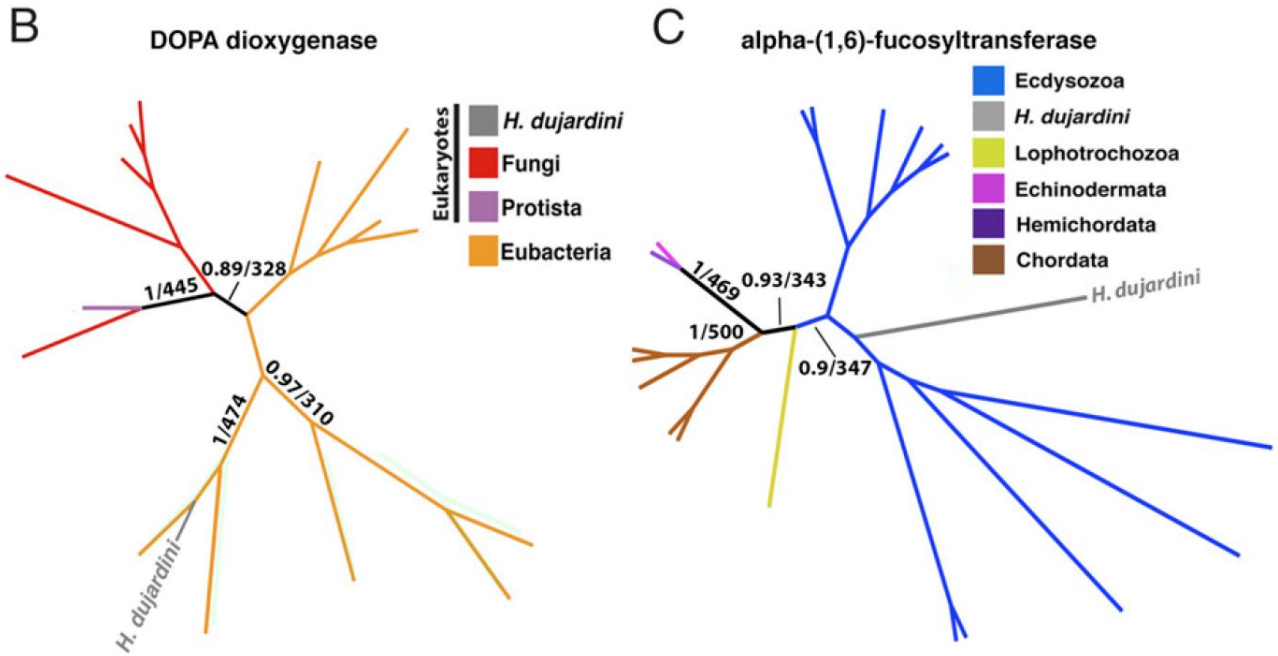
3. Potential fHGT loci had introns

HGT index calculations revealed bacterial sequence origins



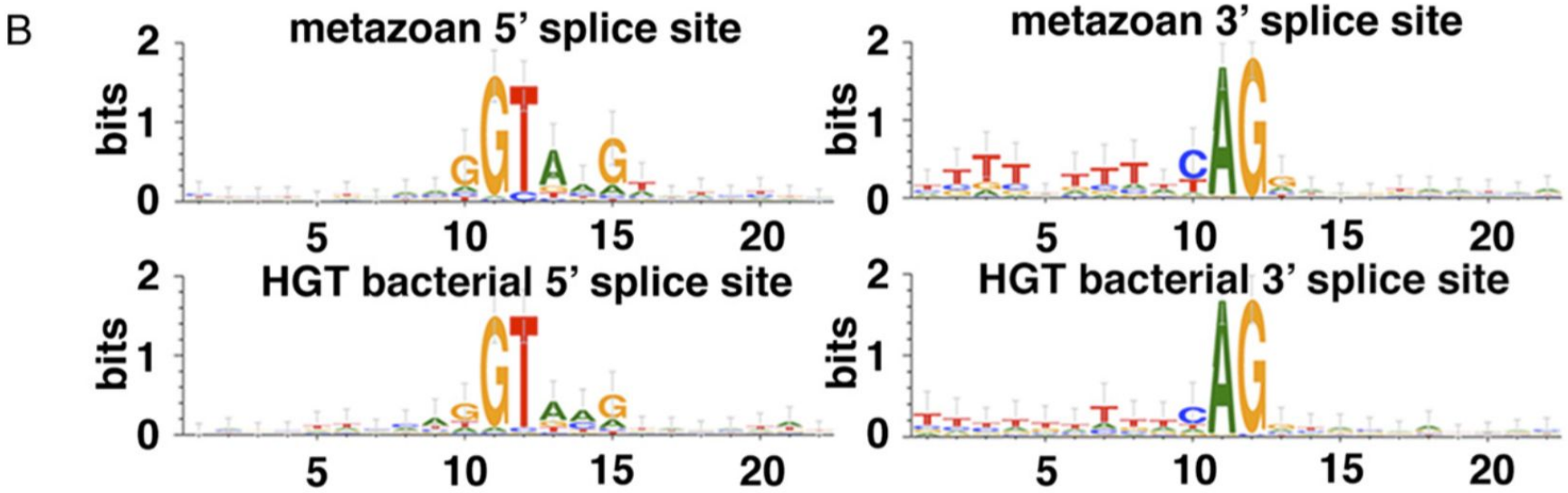
Boothby et al (2015)

Phylogenies of candidate genes were from non-metazoan taxa



Boothby et al (2015)

Potential fHGT loci had introns



Boothby et al (2015)

fHGT claims, refuted

1. HGT index calculations revealed bacterial sequence origins
2. Phylogenies of candidate genes were from non-metazoan taxa
3. Potential fHGT loci had introns

1. HGT index is uninformative if assembly is contaminated. Not itself proof of HGT.
2. While a strong indicator of fHGT, requires solid evidence for genome integration
3. Eukaryotic gene finders can artificially create introns.

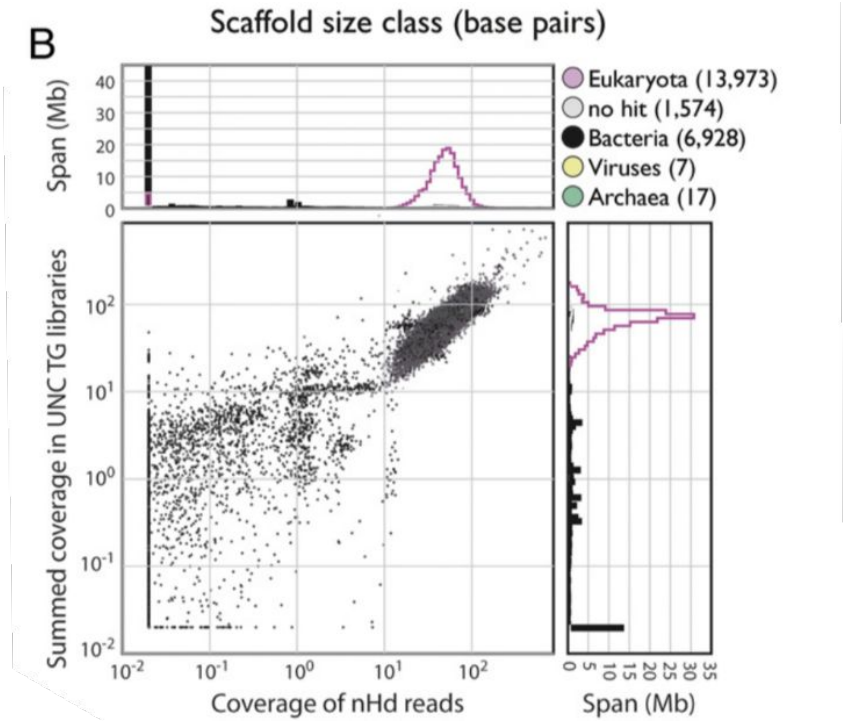
Take away points

1. Be very rigorous about data cleaning
2. Make sure to have controls at every step in your analysis
3. Many necessary pieces of evidence do not form into one sufficient piece of evidence

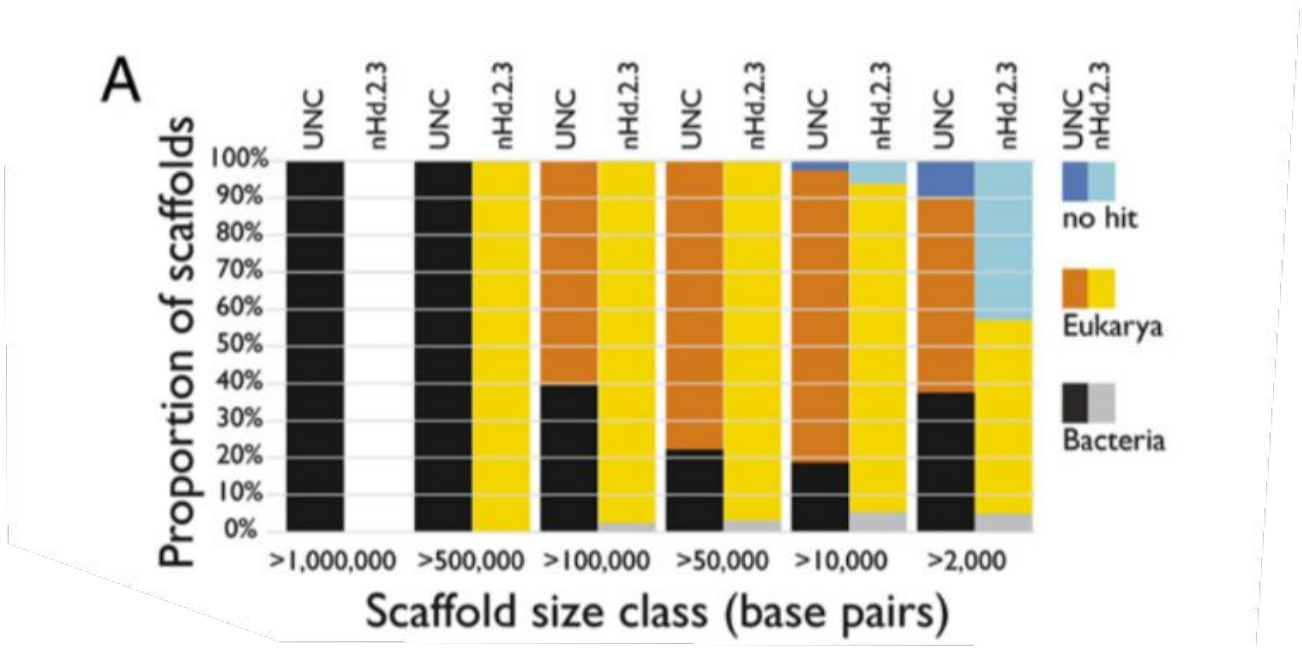
Questions?

Extra Slides

Bacterial read coverage does not overlap in UNC and Edinburgh raw data



Presence of bacterial and other features in assembly of sequences



Cryptobiosis

- As defined by David Keilin, “the state of an organism when it shows no visible signs of life and when its metabolic activity becomes hardly measurable, or comes reversibly to a standstill”
- Induced by:
  - Desiccation (anhydrobiosis)
  - Low temperatures (cryobiosis)
  - Lack of oxygen (anoxybiosis)
- Through cryptobiosis, tardigrades survived and even reproduced after a journey into outer space in 2007!