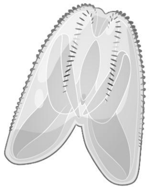
In silico structural and functional annotation of the ctenophore *Mnemiopsis leidyi* genome

Gabriel Dall'Alba (MSc Candidate)
Supervisor: Dr. Steven Plotkin
Genome Science and Technology Program
B.I.G. 2023

Motivation

Early-branching groups (non-bilaterian animals: sponges, ctenophores, placozoans and cnidarians) provide unique opportunities to investigate the origins of mechanisms that allow multicellularity to emerge and be sustained.

High-quality genome assemblies with a well-characterized gene repertoire facilitates the exploration of such topics.



Mnemiopsis leidyi

Sequenced in 2013

5100 scaffolds
N50 = 187 kb
Genome size: 156 Mb

Pleurobrachia bachei

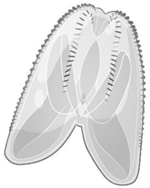
Sequenced in 2014

21979 scaffolds
N50 = 23 kb
Genome size: 157 Mb

Hormiphora californensis

Sequenced in 2017

13 Chromosome-scale scaffolds
N50 = 8.5 Mb
Genome size: 140 Mb

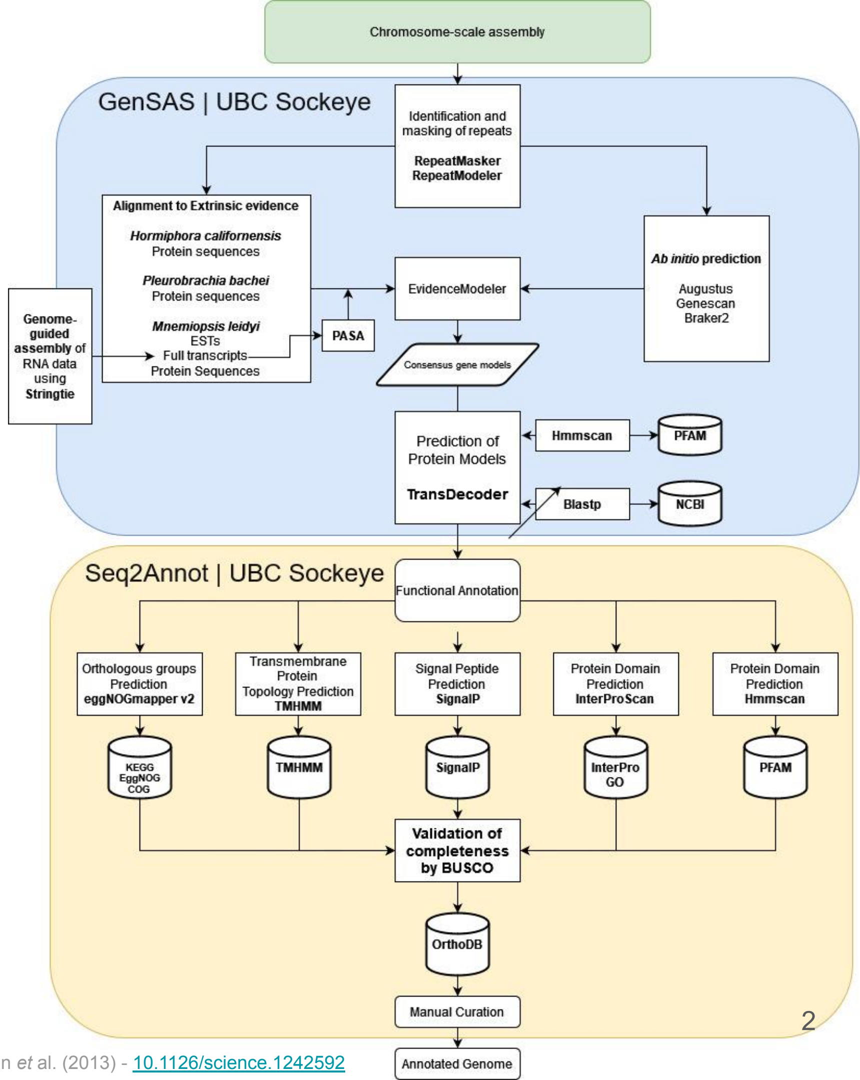


Mnemiopsis leidyi

Resequenced in 2022

13 Chromosome-scale scaffolds
N50 = ~16Mb
Genome size: 203 Mb

Using PacBio + HiC, we resequenced and are currently annotating the genome of *Mnemiopsis leidyi*.



Chromosome-scale assembly

GenSAS | UBC Sockeye

Identification and masking of repeats
RepeatMasker
RepeatModeler

Alignment to Extrinsic evidence
Hormiphora californensis
Protein sequences
Pleurobrachia bachei
Protein sequences
Mnemiopsis leidyi
ESTs
Full transcripts
Protein Sequences

Genome-guided assembly of RNA data using **Stringtie**

PASA

EvidenceModeler

***Ab initio* prediction**
Augustus
Genescan
Braker2

Consensus gene models

Prediction of Protein Models
TransDecoder

Hmmscan
PFAM
Blastp
NCBI

Seq2Annot | UBC Sockeye

Functional Annotation

Orthologous groups Prediction
eggNOGmapper v2
KEGG
EggNOG
COG

Transmembrane Protein Topology Prediction
TMHMM
TMHMM

Signal Peptide Prediction
SignalP
SignalP

Protein Domain Prediction
InterProScan
InterPro
GO

Protein Domain Prediction
Hmmscan
PFAM

Validation of completeness by BUSCO

OrthoDB

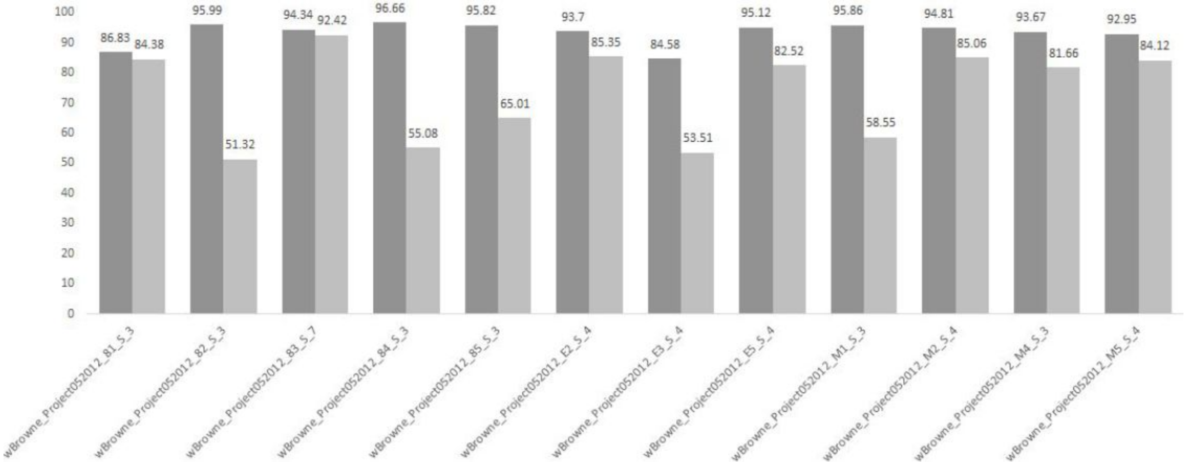
Manual Curation

Annotated Genome

Schultz *et* al. (2021) - https://doi.org/10.1093/g3journal/jkab302, Levin *et* al. (2014) - http://dx.doi.org/10.1038/nature13400, Ryan *et* al. (2013) - 10.1126/science.1242592

Dataset	**Biological Material**	**Avg. Alignment rate (%) of raw reads to assembly**		**# Transcripts generated** (stringtie, guided using 2022 assembly)	**BUSCO completeness**
		2022	**2013**		
Davidson *et* al. (2017)	Single-cell embryo, 2, 4, 8 cells embryo (including Macromeres/Micromeres);	97.71%	83.11%	18932	C:91.4%[S:70.2%,D:21.2%],F:1.6%,M:7.0%
Babonis *et* al. (2018)	Tentacle bulbs; Comb Rows	90.92%	88.10%	46075	C:93.0%[S:67.1%,D:25.9%],F:1.6%,M:5.4%

Davidson *et* al. (2017) % of raw reads that map to transcriptome generated using Stringtie (genome-guided).



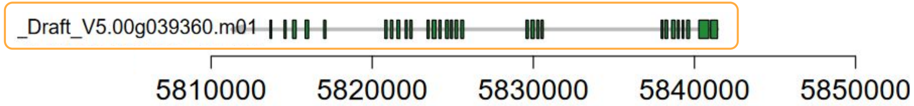
2013 avg. = 73%

2022 avg .= 93%

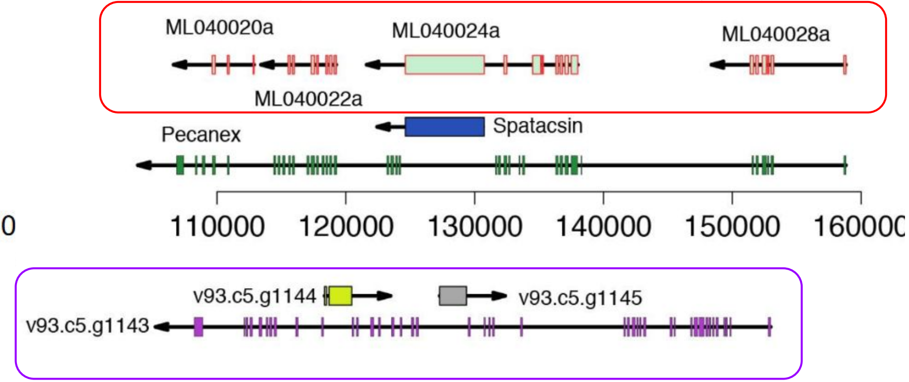
Davidson *et* al. (2017) - 10.1002/mrd.22926, Babonis *et* al. (2018)

	2022 Assembly	2013 Assembly
Number of Genes	15,273	16,548
% of exonic sequence	11.31%	14%
Intron Median Size (bp)	273	275
Intergenic distance Q10 (bp)	435	287
Intergenic distance Q50 (bp)	3127	1990
Intergenic distance Q90 (bp)	18348	8806

Pecanex model (2022) assembly



Incorrectly split gene models in *Mle* 2013 assembly



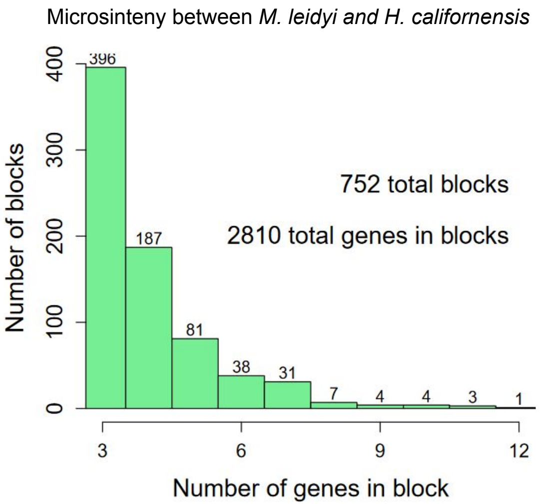
H. californensis models

Right side image modified from Schultz *et* al. (2021) - https://doi.org/10.1093/g3journal/jkab302

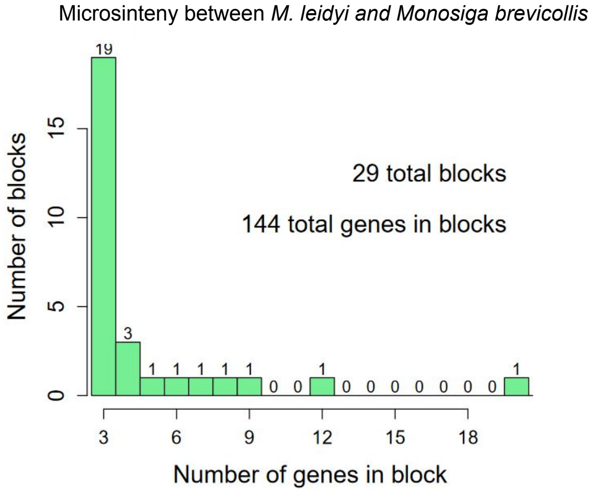
There is a clear 1-to-1 correspondence of chromosomes between *M. leidyi* and *H. californensis*.

Genes seems to be highly rearranged intra-chromosomally.

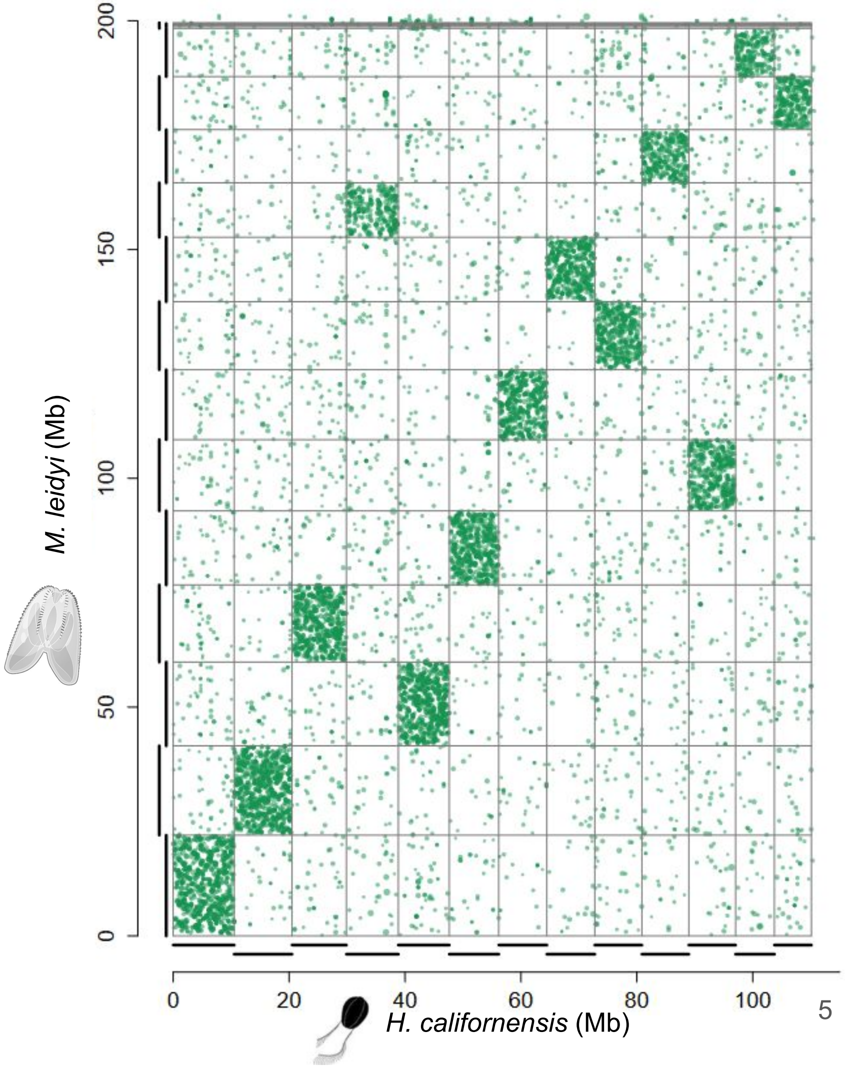
The extension of the changes provoked by this pattern remains to be investigated.



Microsinteny between *M. leidyi and H. californensis*



Microsinteny between *M. leidyi and Monosiga brevicollis*



Acknowledgements


Steven Plotkin

Supervisor


Pranav Garg

Sequencing and genome assembly


Charlotte Barclay

HiC sequencing and genome assembly

BIG23

BIOINFORMATICS
INTERDISCIPLINARY ONCOLOGY
GENOME SCIENCE AND TECHNOLOGY

UBC ARC SOCKEYE

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

UBC
Genome Science + Technology Graduate Program