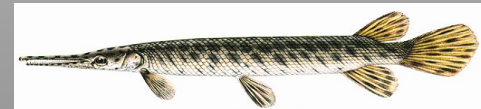
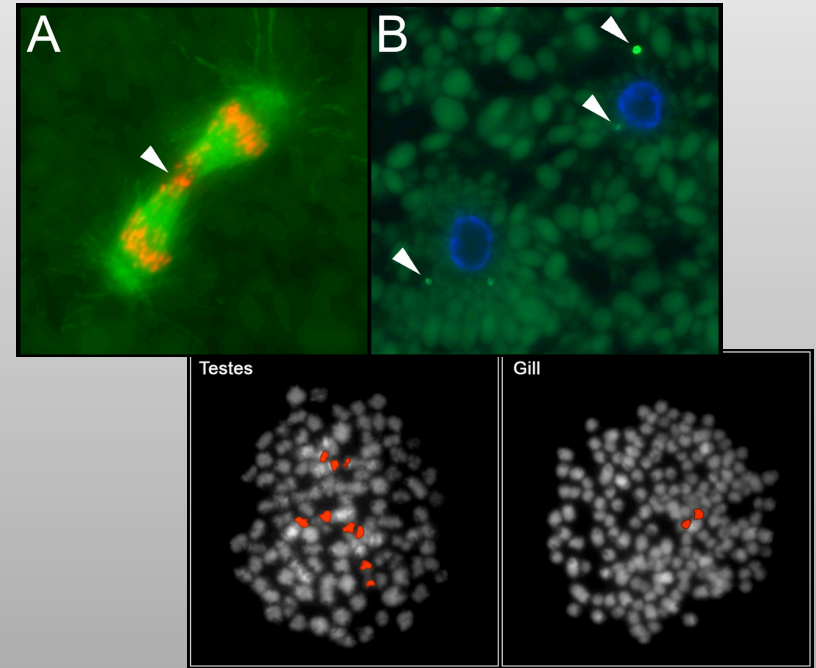
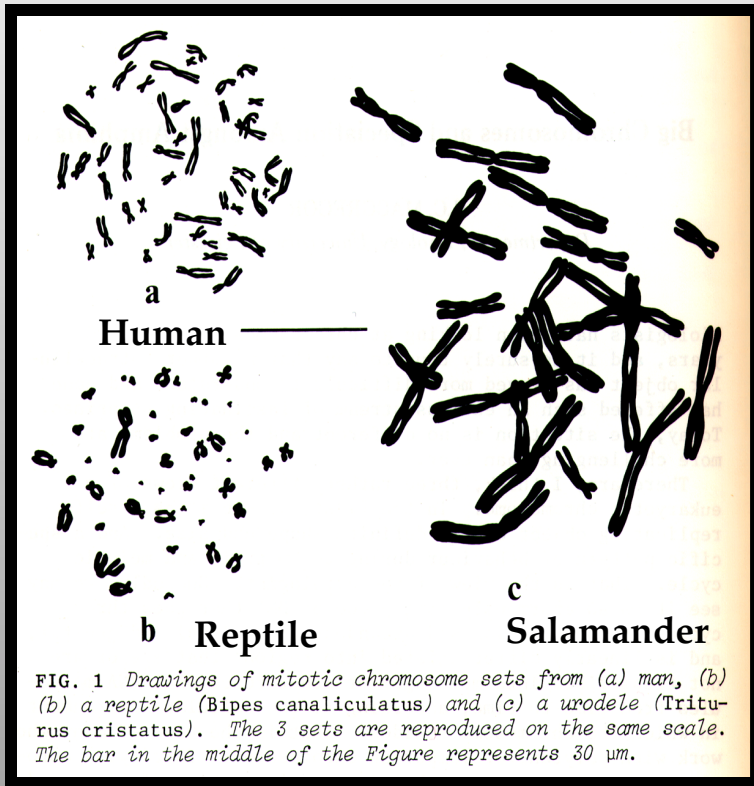
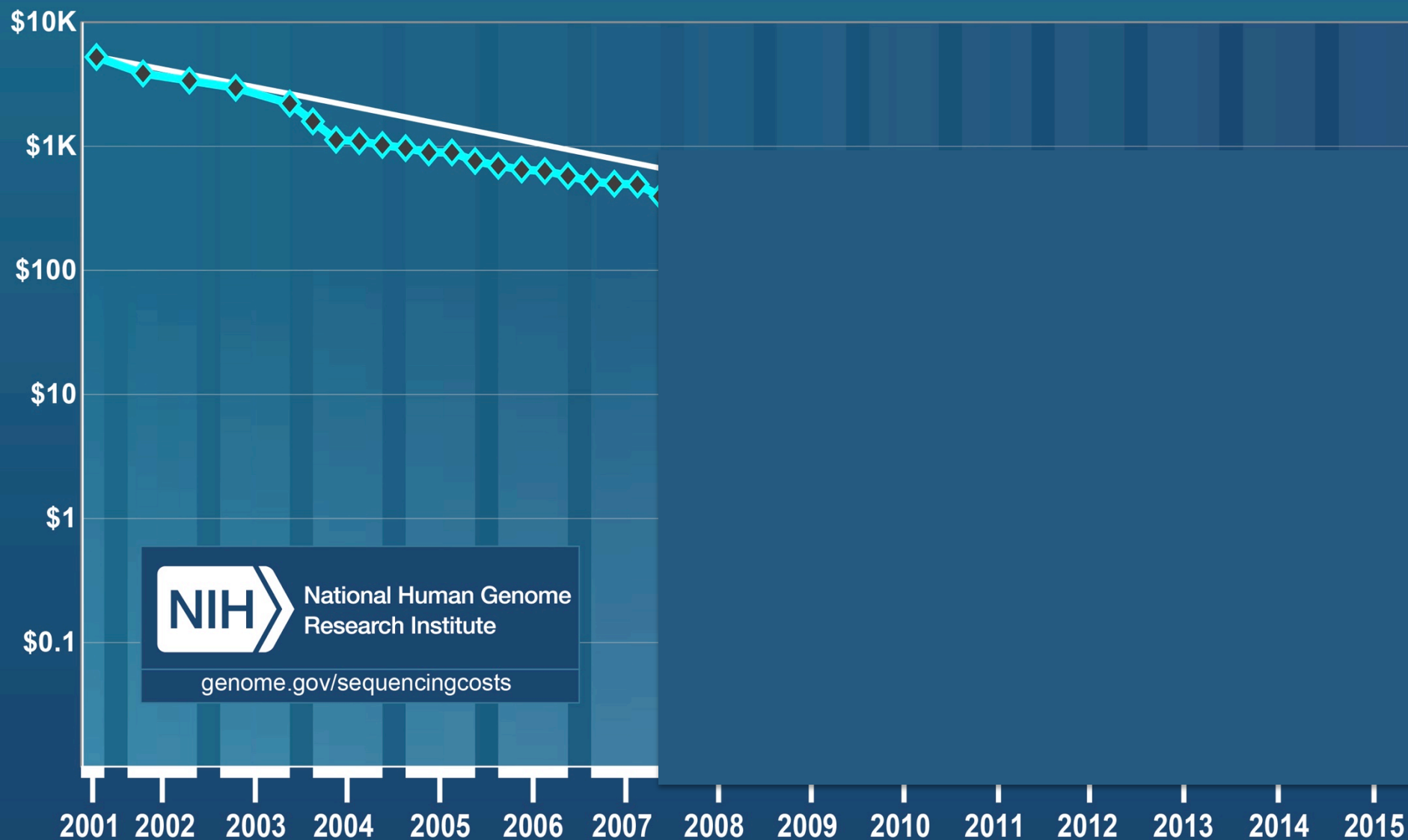


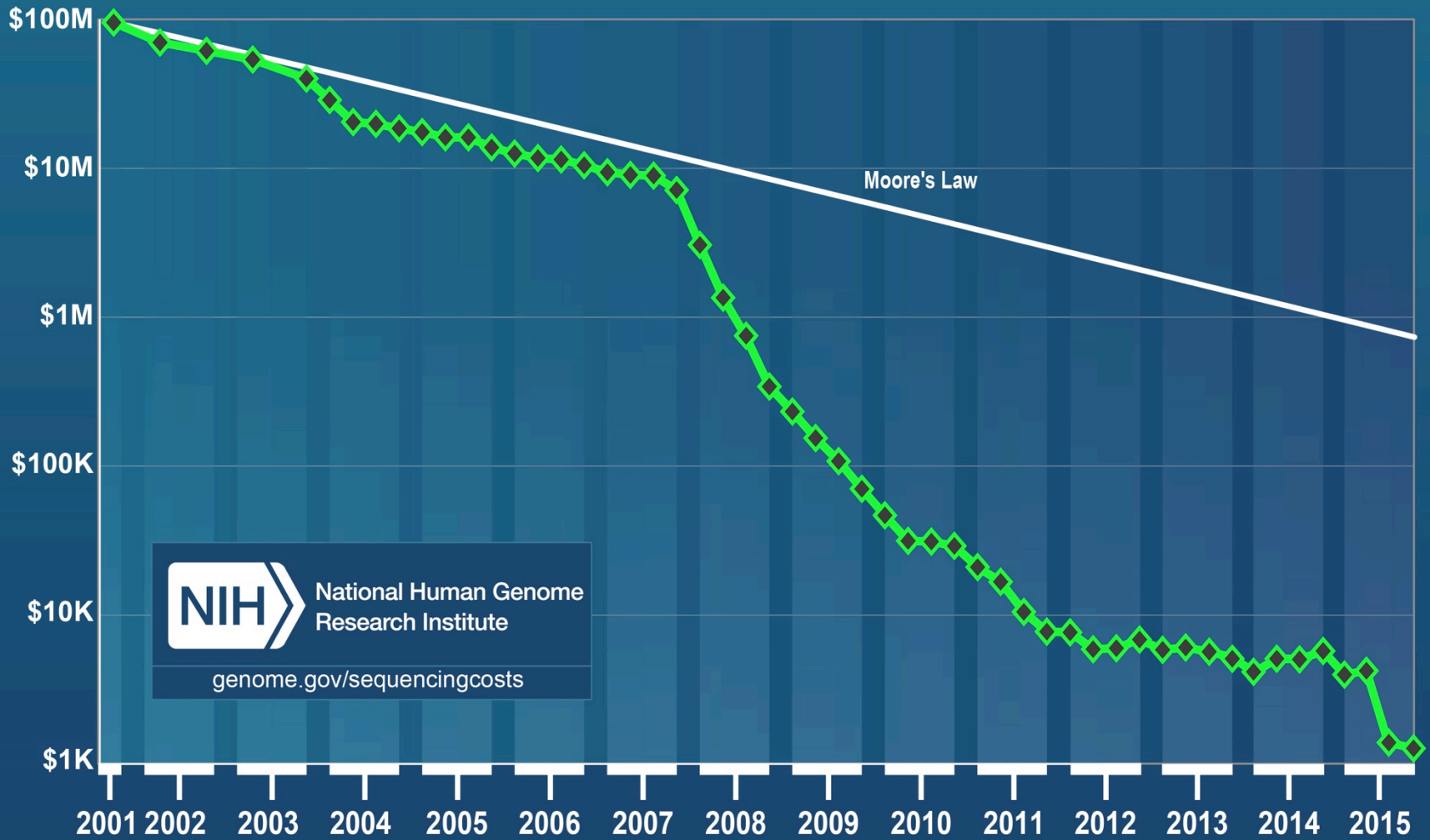
Analysis of Complex Vertebrate Genomes: Computational Challenges and Solutions

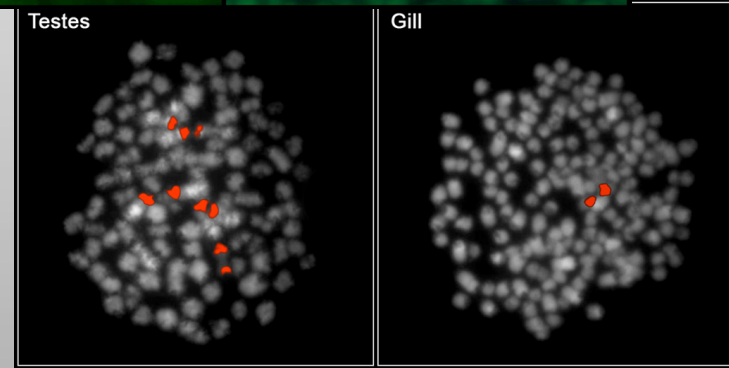
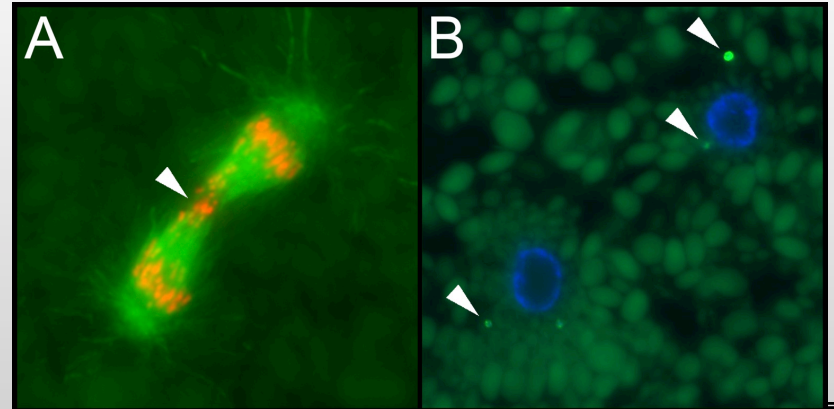
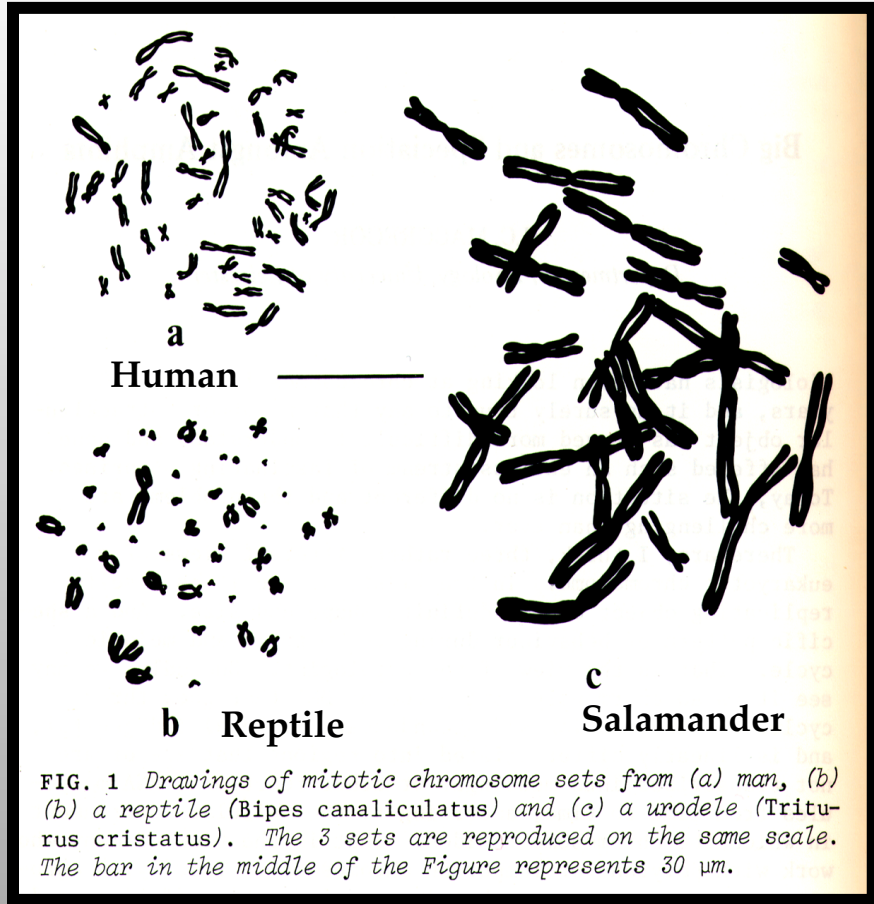


Cost per Raw Megabase of DNA Sequence

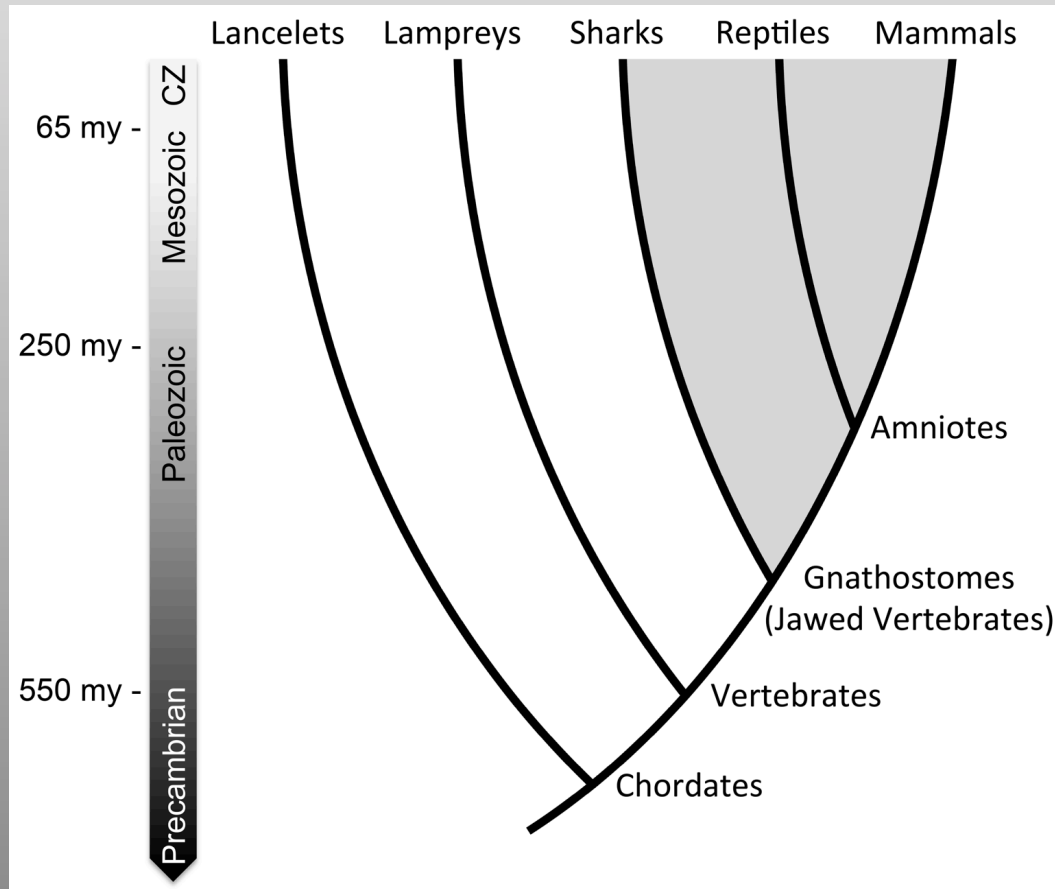


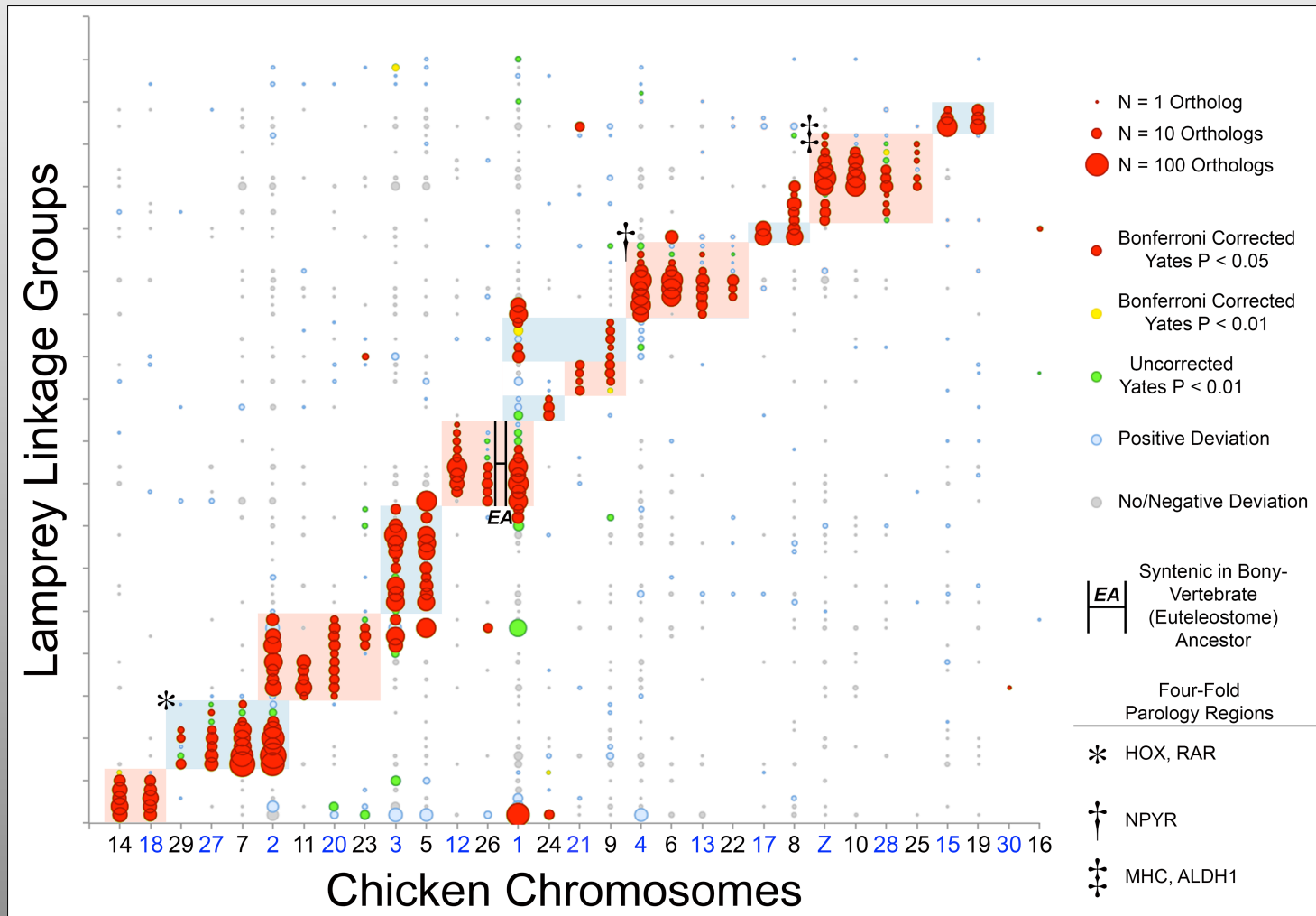
Cost per Genome



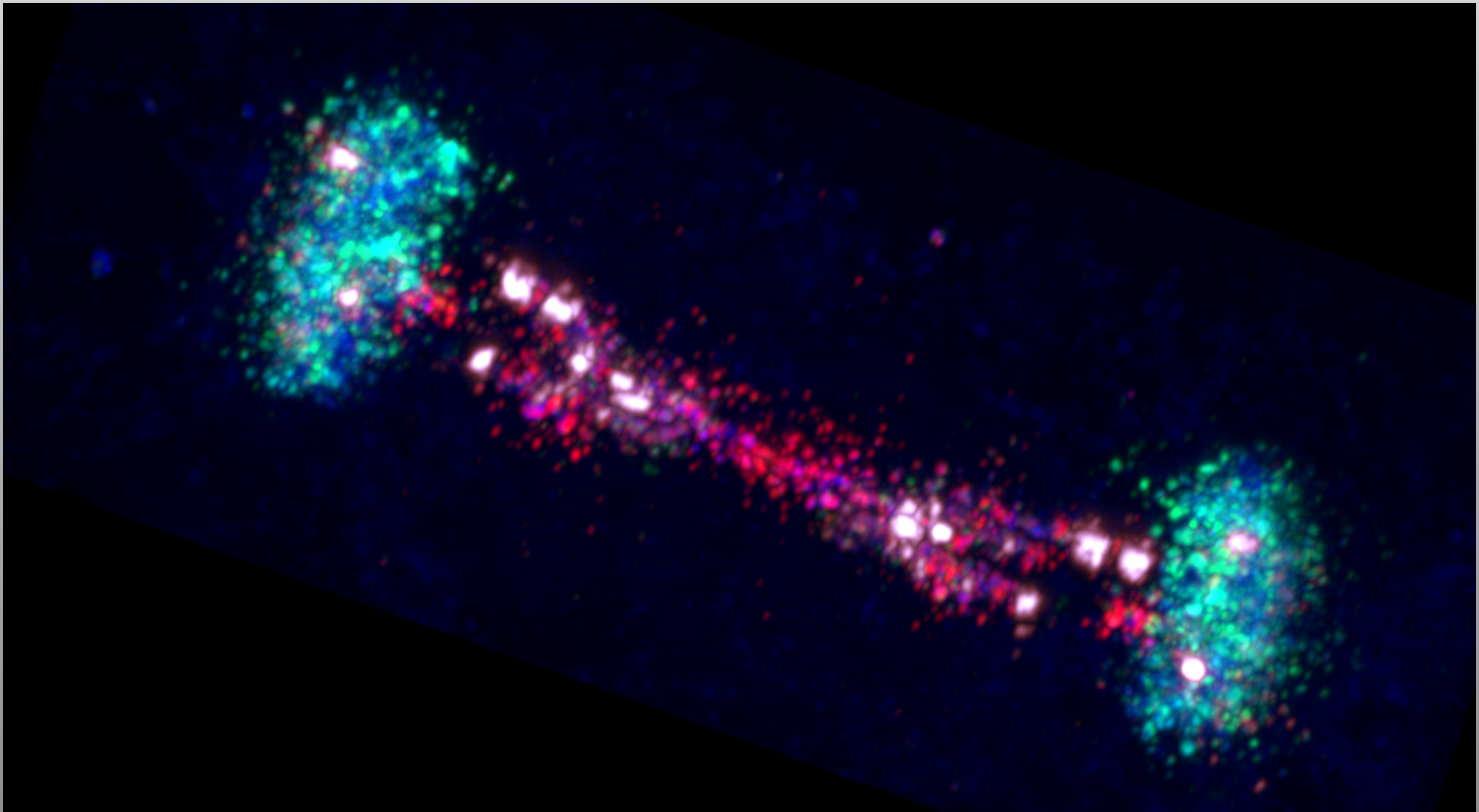


Lamprey

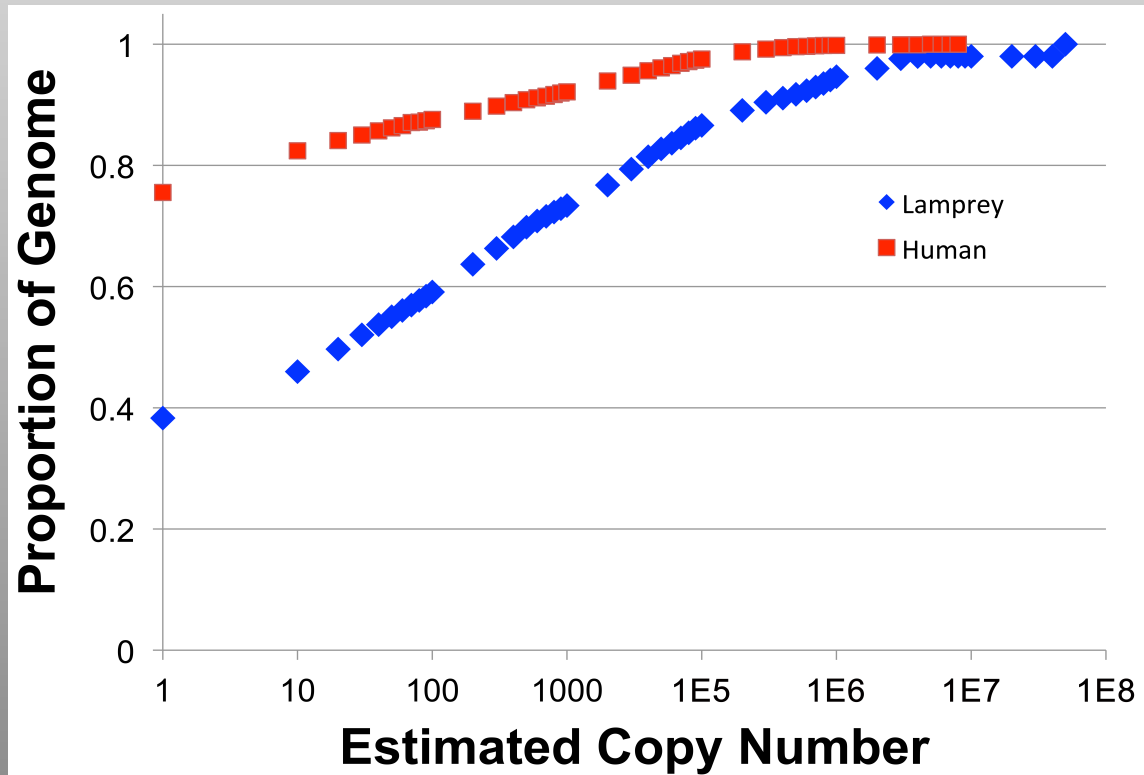




Lamprey



A Painful Genome to Assemble



Assembly Attempts

(each internally optimized)

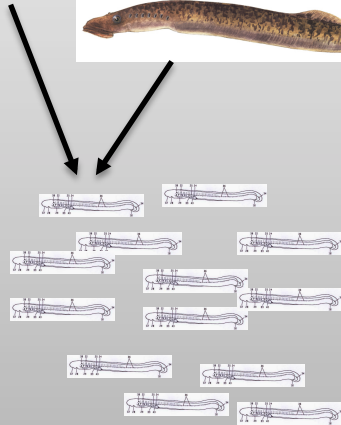
	N50		Assembly	
	Length	Number	Length	Number
AllPaths	642	269,078	1.095 Gb	2.3 million

Assembly Attempts

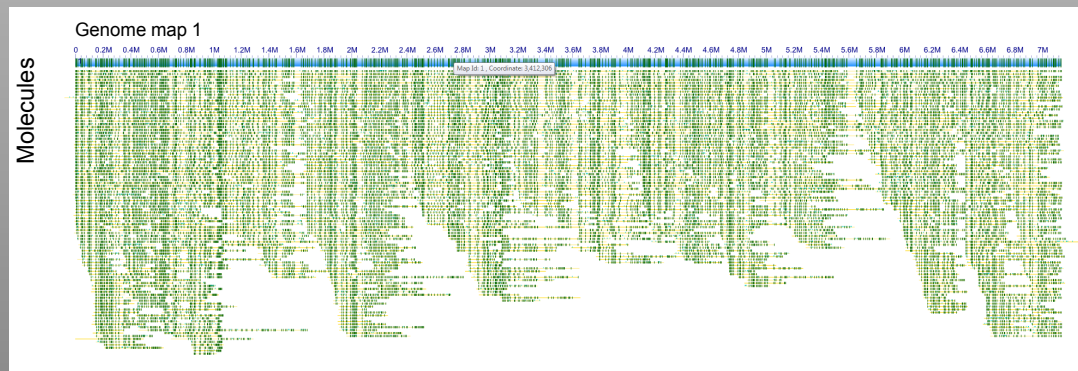
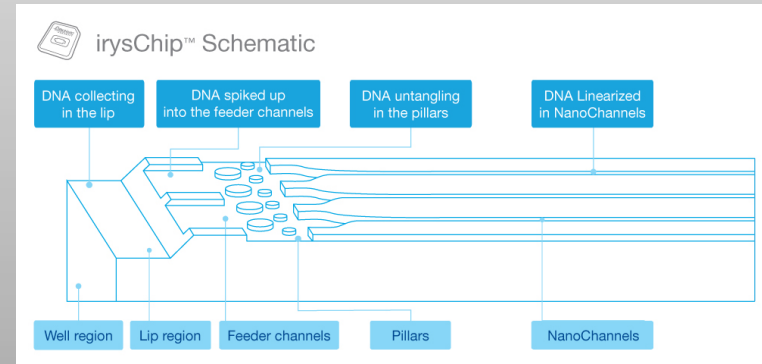
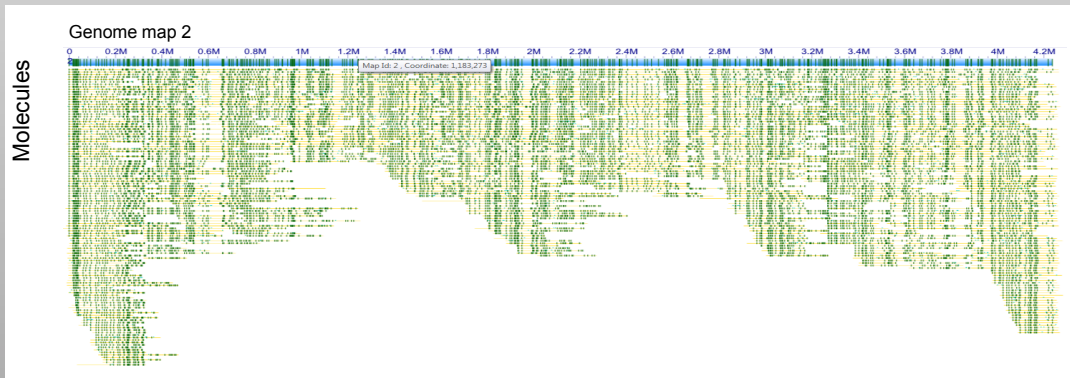
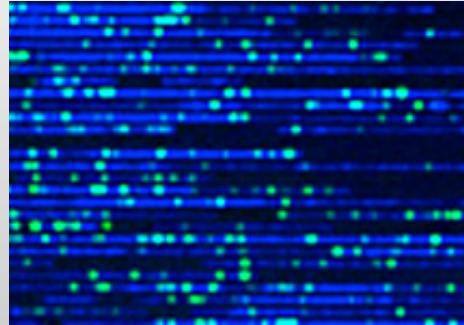
(each internally optimized)

	N50		Assembly	
	Length	Number	Length	Number
AllPaths	642	269,078	1.095 Gb	2.3 million
pacBioToCA+SSPACE	11,721	23,858	1.174 Gb	201,586
DBG2OLC+SSPACE	193,210	1,266	1.043 Gb	22,820

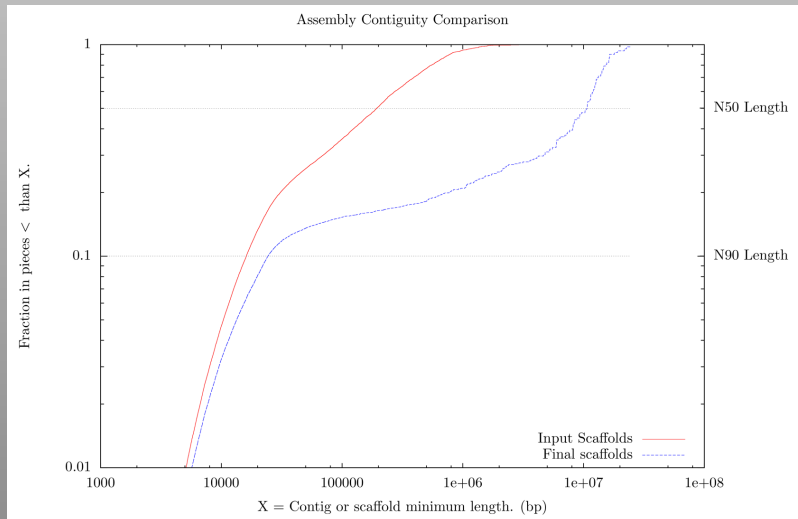
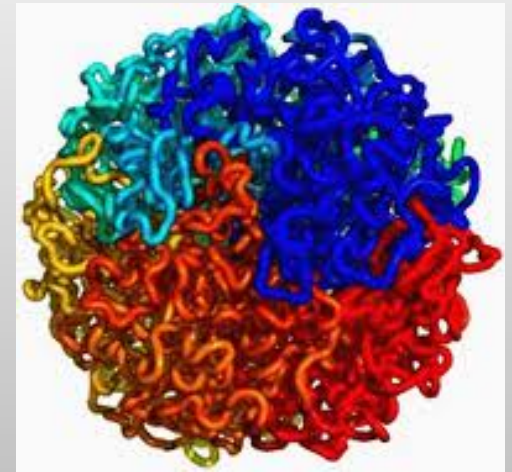
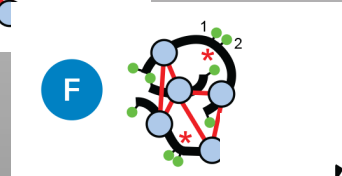
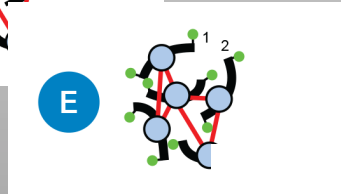
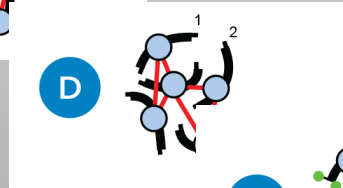
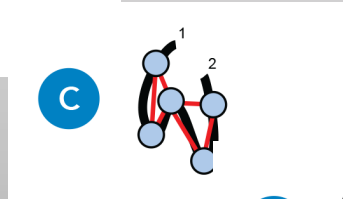
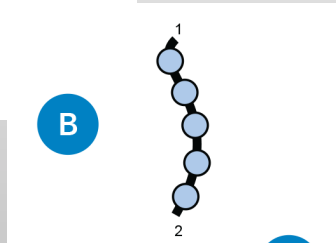
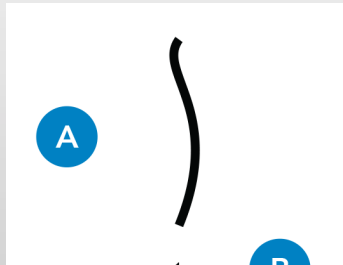
Meiotic Maps



Optical Mapping



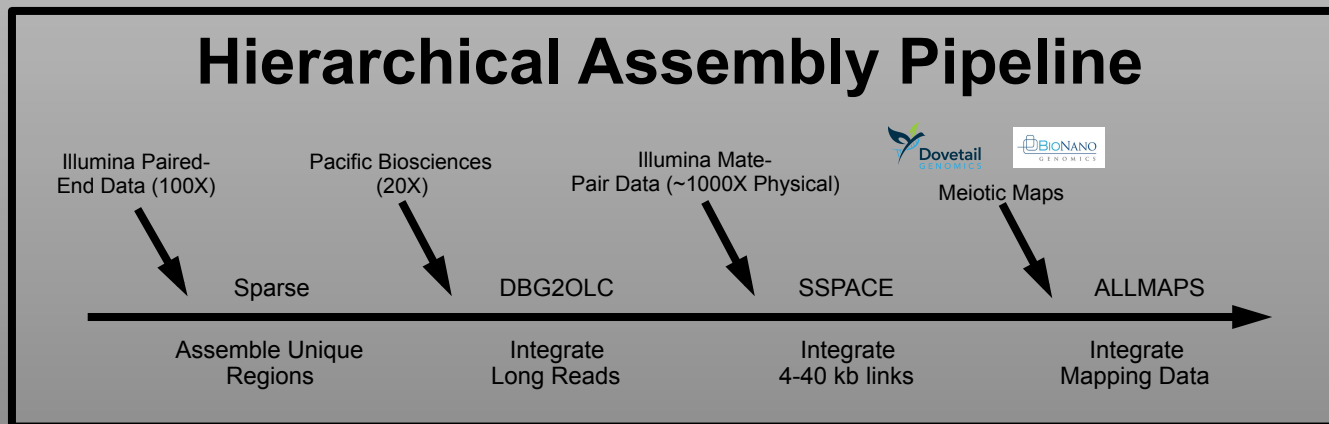
Chromatin Contacts

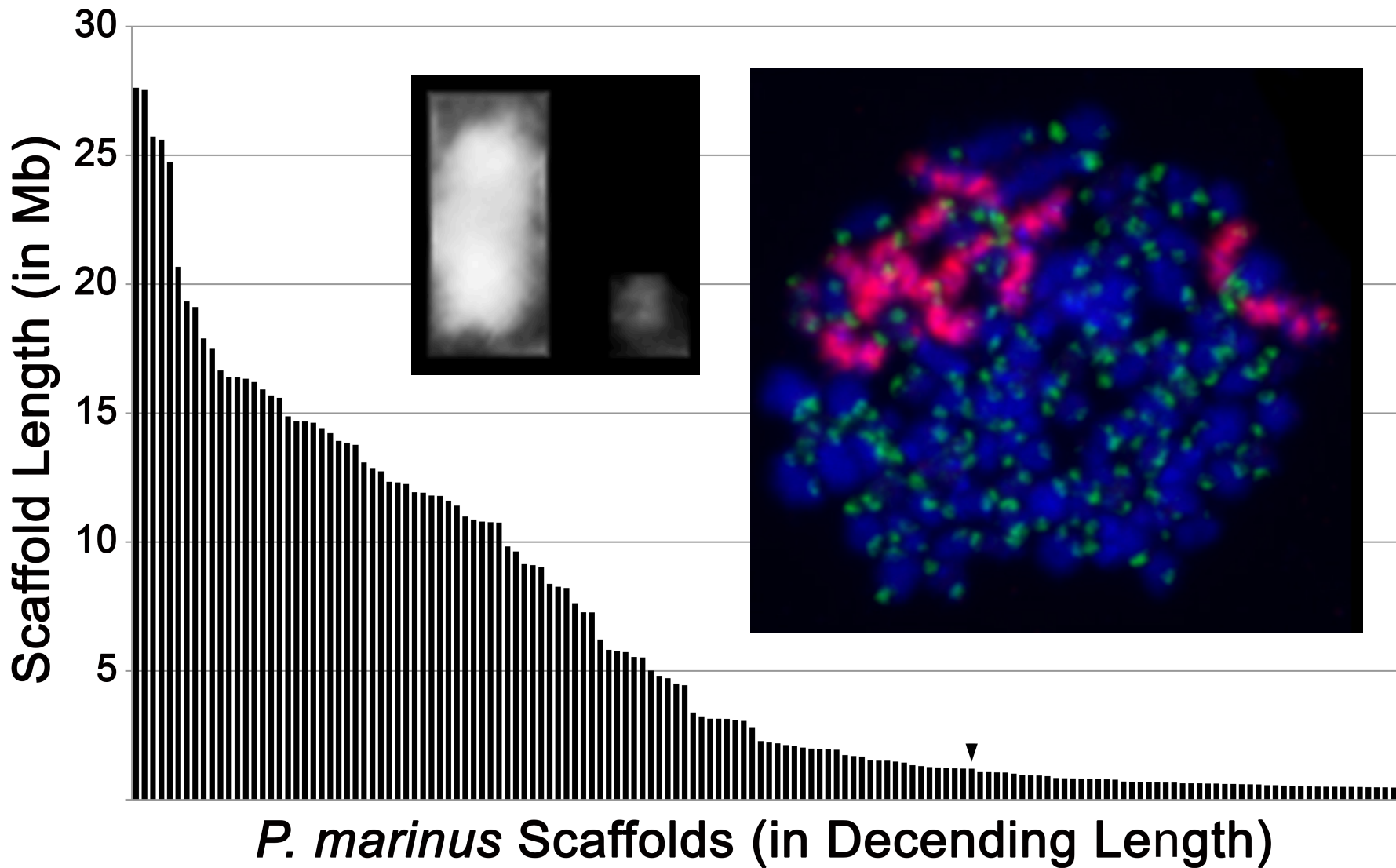


Assembly Attempts

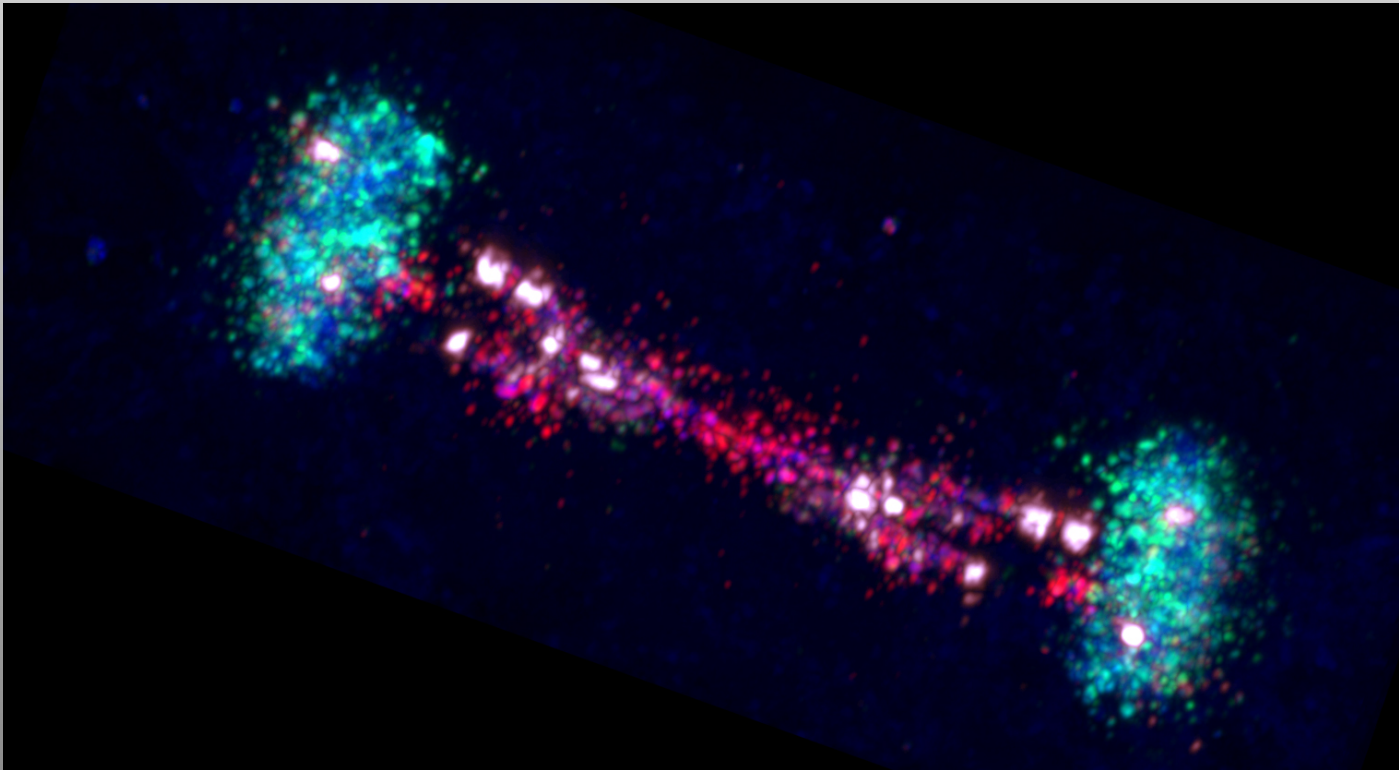
(each internally optimized)

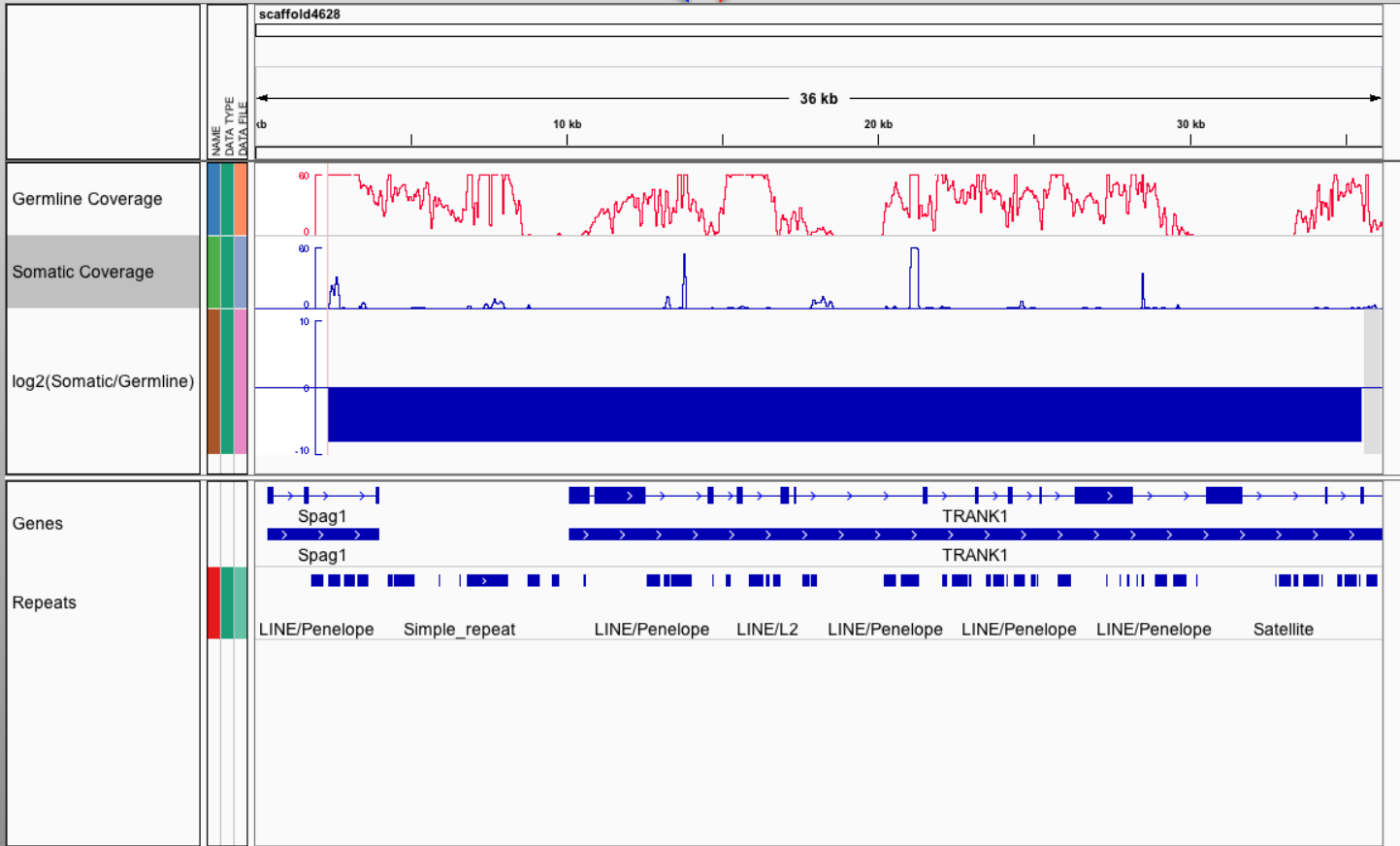
	N50		Assembly	
	Length	Number	Length	Number
AllPaths	642	269,078	1.095 Gb	2.3 million
pacBioToCA+SSPACE	11,721	23,858	1.174 Gb	201,586
DBG2OLC+SSPACE	193,210	1,266	1.043 Gb	22,820
Meiotic Map	348,730	173	1.052 Gb	21,246
BioNano	910,000	232	1.173 Gb	21,780
Dovetail	10.7 Mb	37	1.053 Gb	12,144
BN -> DT	7.4 Mb	46	1.174 Gb	12,373
DT -> BN	10.7 Mb	~37	1.081 Gb	21,239
AllMaps	11.9 Mb	34	1.130 Gb	12,077





Using the Genome





← ⓘ GitHub, Inc. (US) | <https://github.com/timnat/DifCover> | 🔍 Search

📁 Most Visited 🌐 Getting Started 📧 Gmail 🗺️ Google Maps 📖 UW Library 🧬 NCBI 📰 Google News 📺 Space 📄 Science Daily 🇬🇧 UKY Libraries

📖 README.md

DifCover

Description

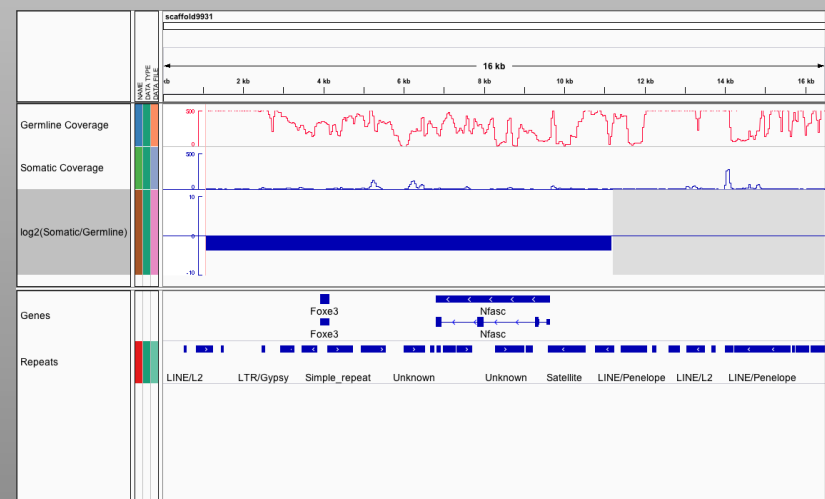
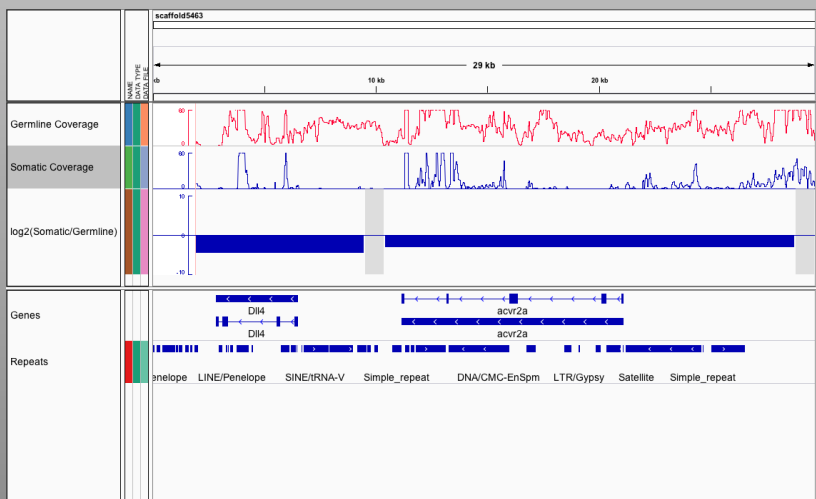
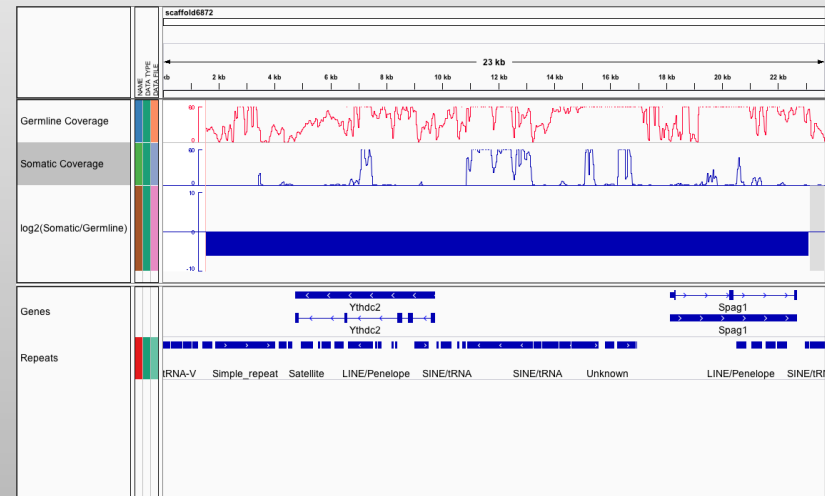
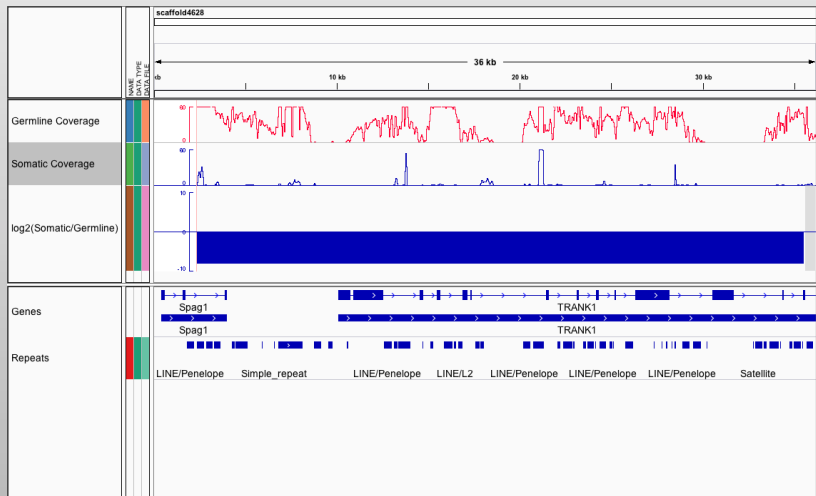
The DifCover pipeline aims to identify regions in a reference genome for which the read coverage of one sample (sample1) is significantly different from the read coverage of another sample (sample2) when aligned to a common reference genome. “Significantly different” is determined by user-defined thresholds. The pipeline allows exclusion of regions from consideration based on read coverage. These include regions with low sequence coverage in both samples (regions that are undersampled due to nucleotide content) and regions with exceedingly high sequence coverage (i.e. repetitive sequences). Both cases can be misleading with respect to coverage analyses. The DifCover pipeline is specifically oriented to the analysis of large genomes and can handle very fragmented assemblies.

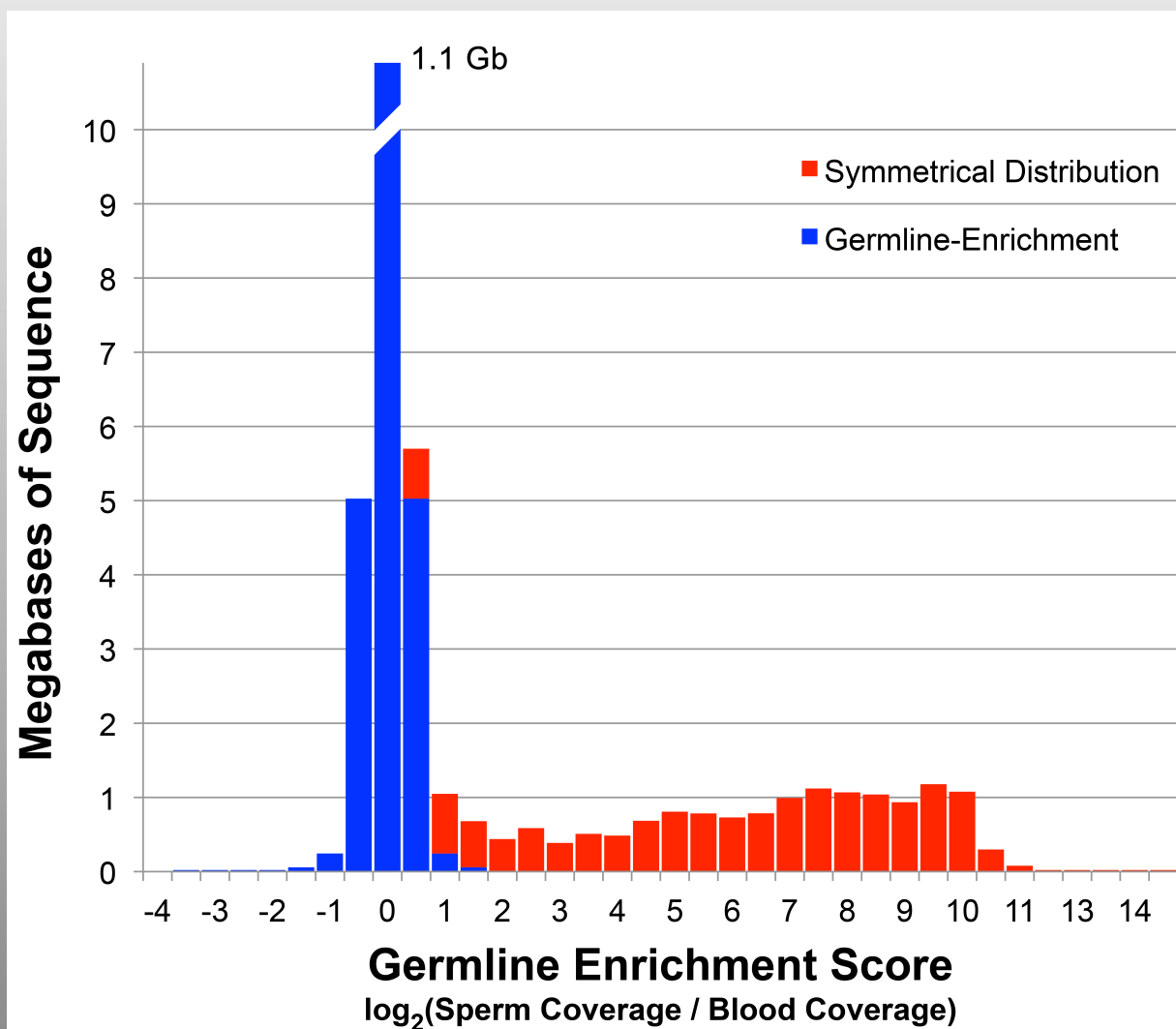
Method

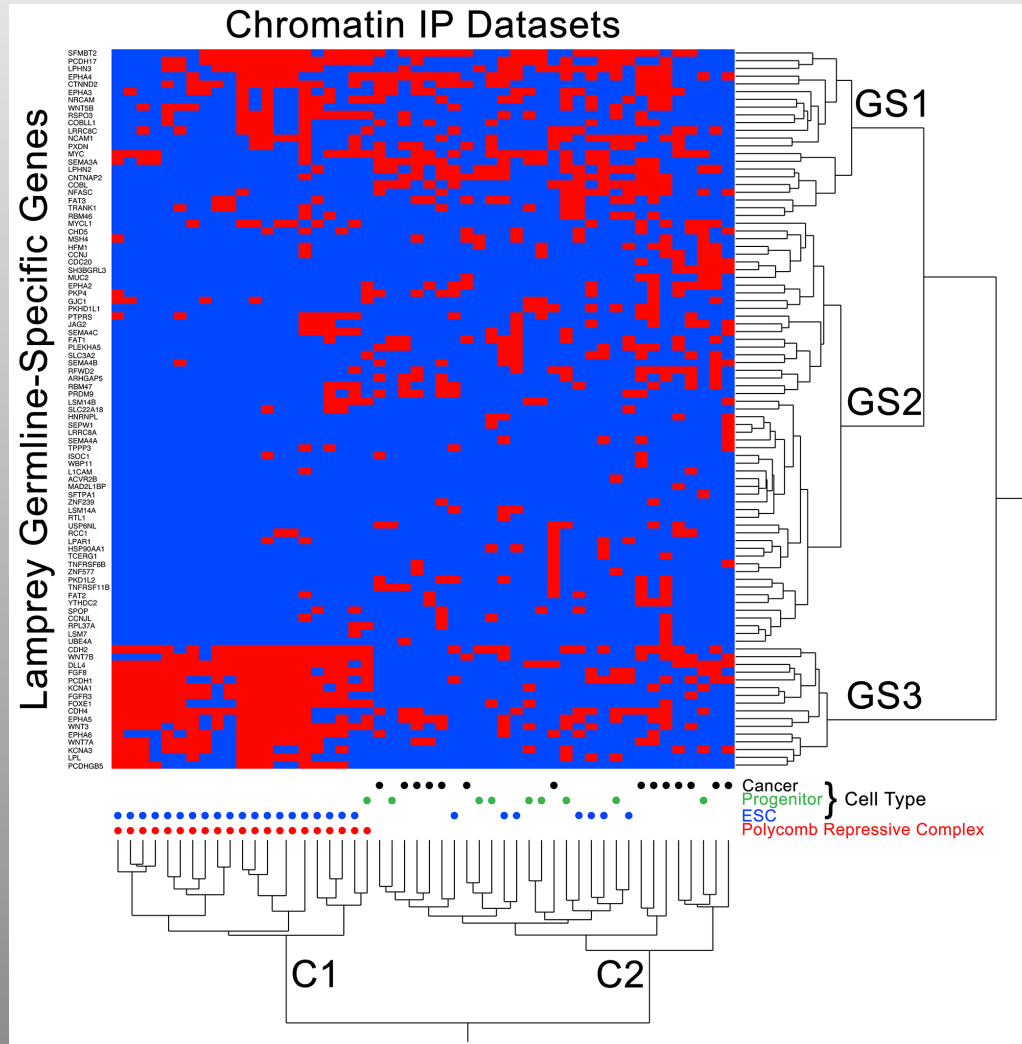
The alignment of short reads to a reference genome can be characterized by the depth of coverage computed for each genomic position as number of reads mapped over it. Fluctuations of coverage can often yield variable

<https://github.com/timnat/DifCover>

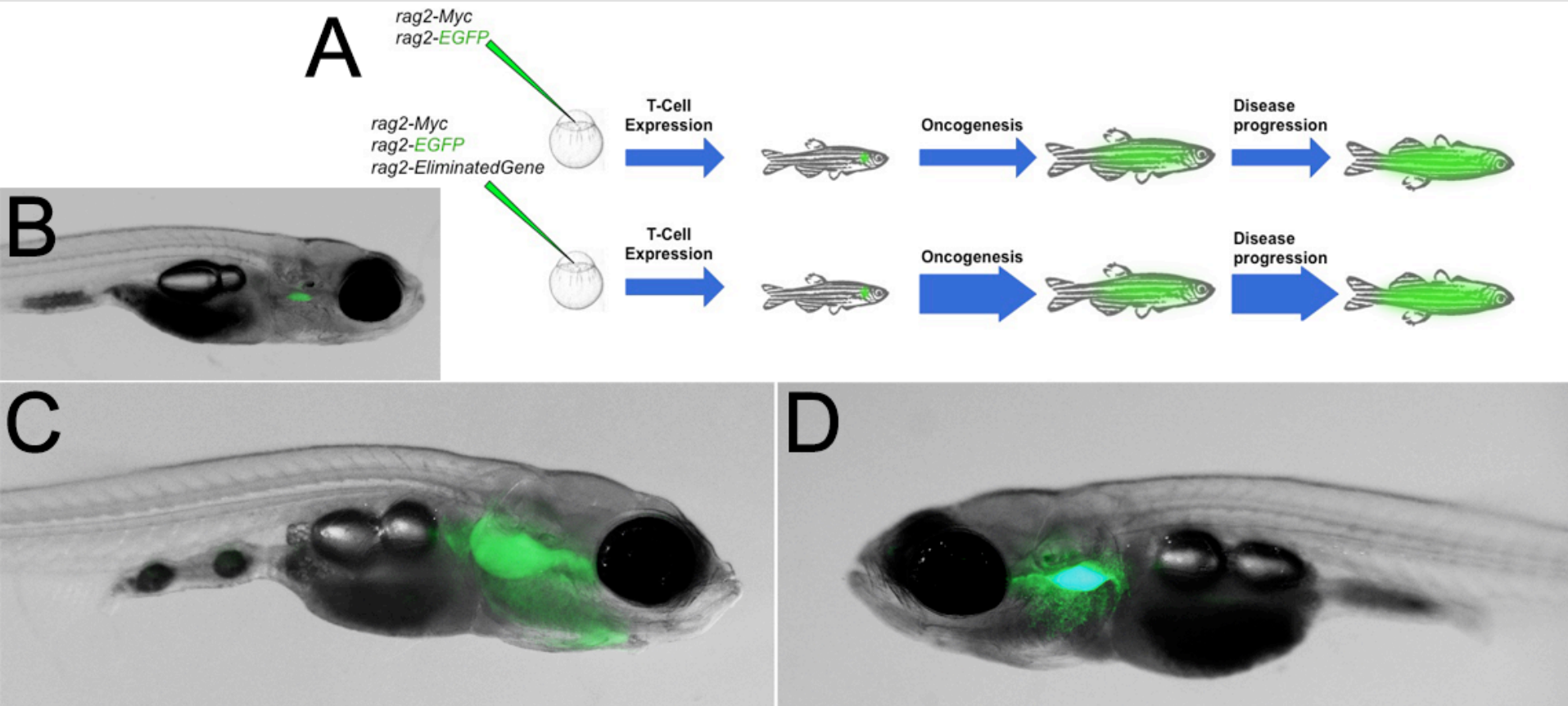
Nataliya's Poster





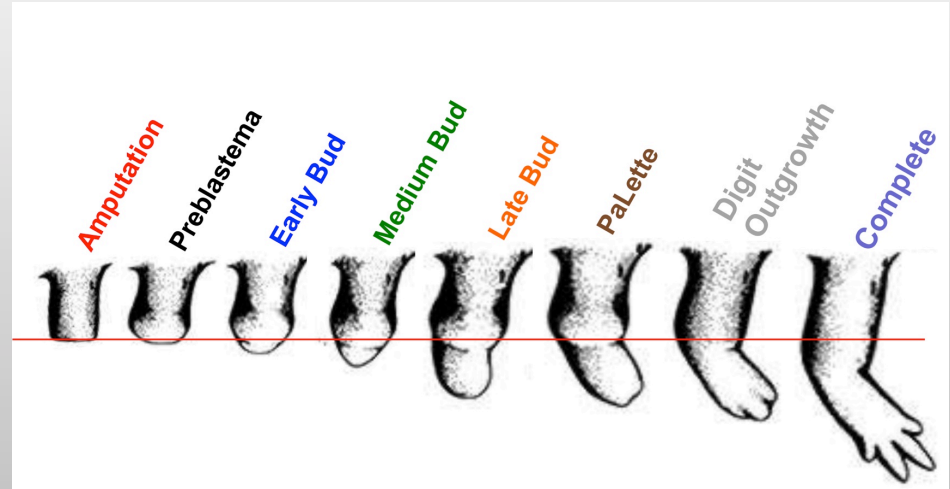
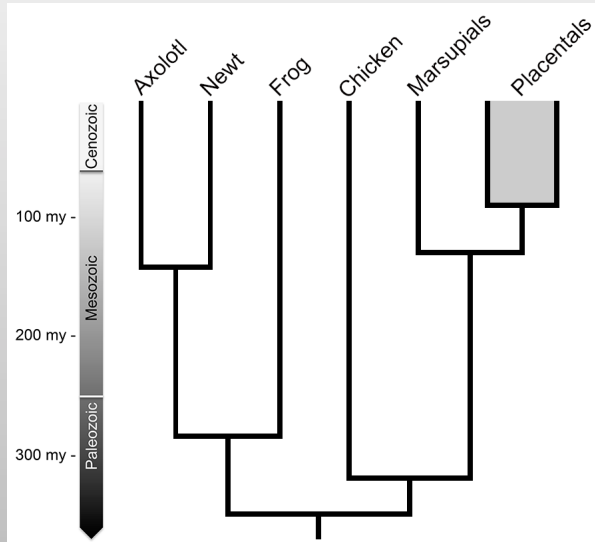


Testing Eliminated Genes

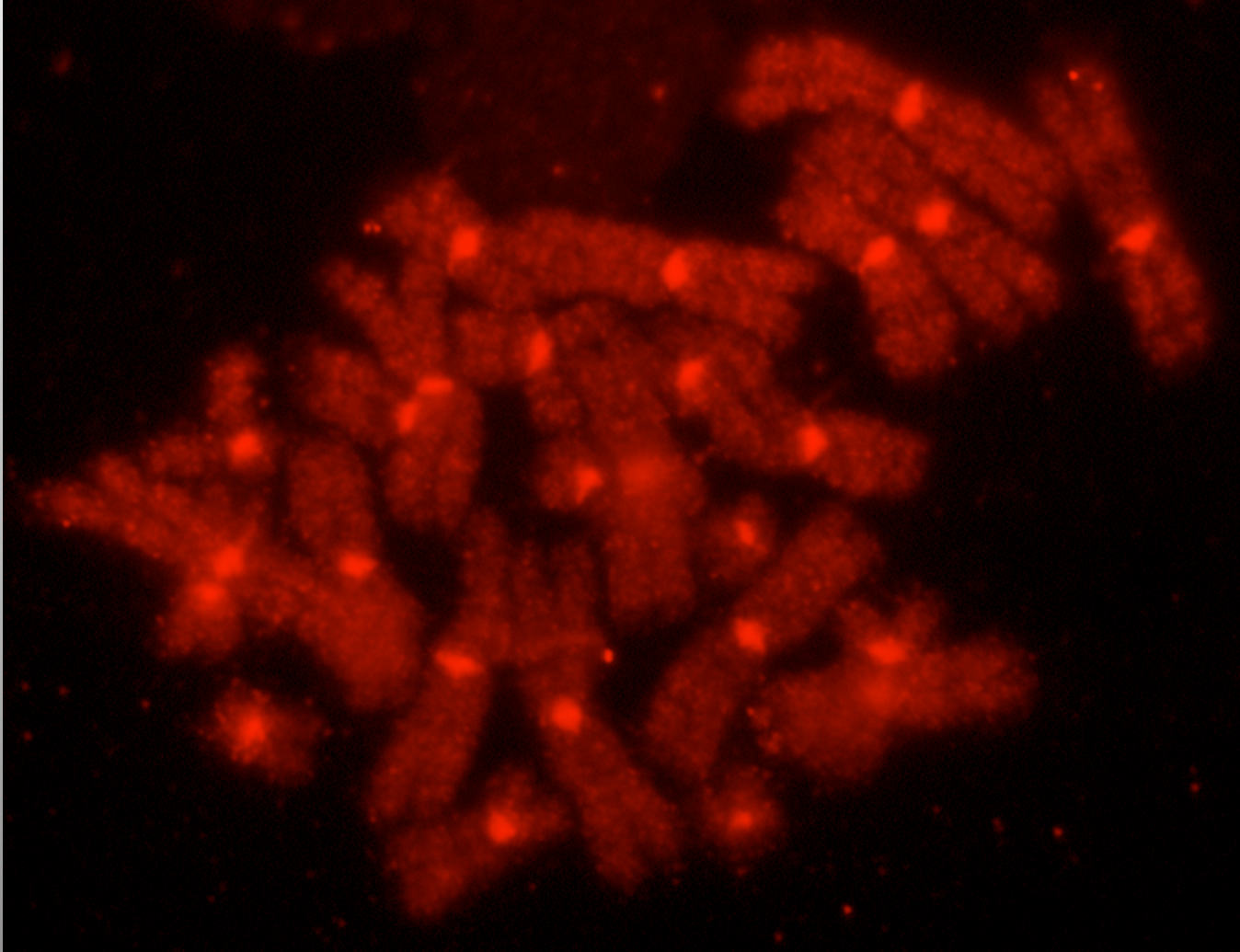




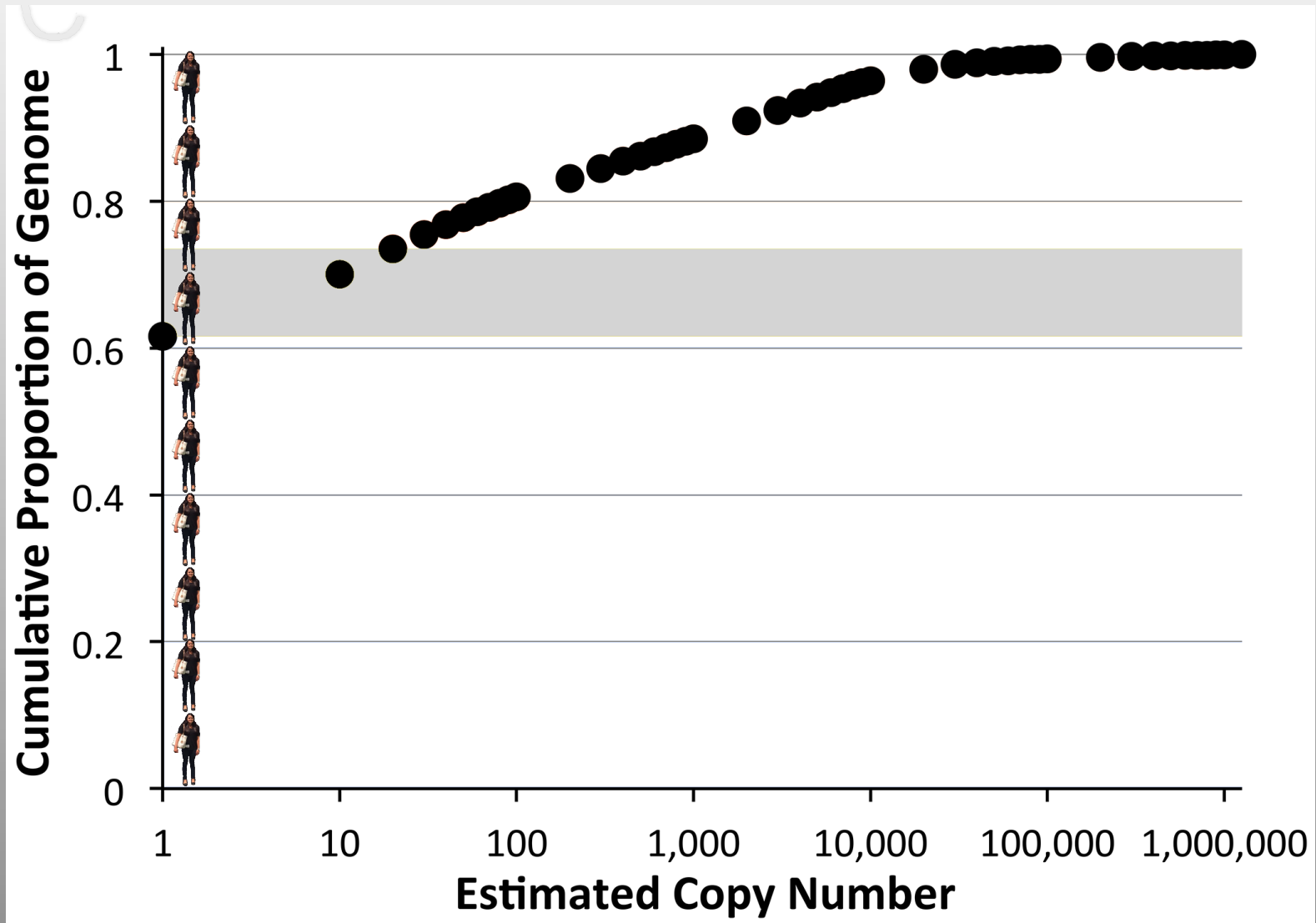
Salamanders are Cool!



Repetitive DNA

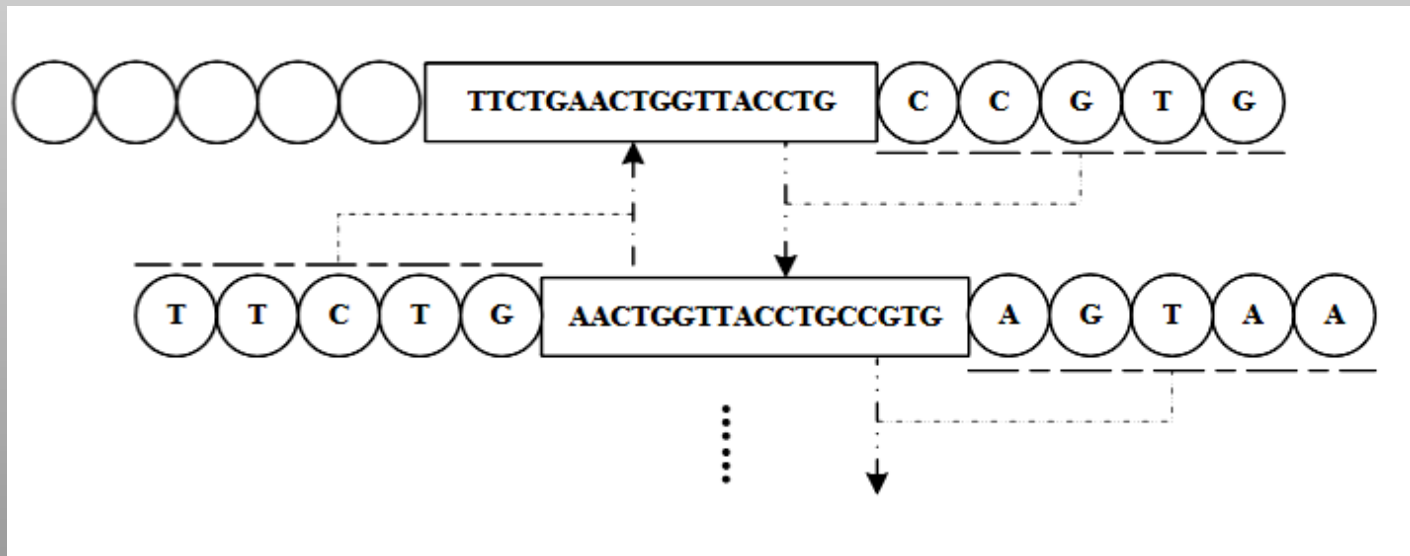


Single-copy vs. Repetitive

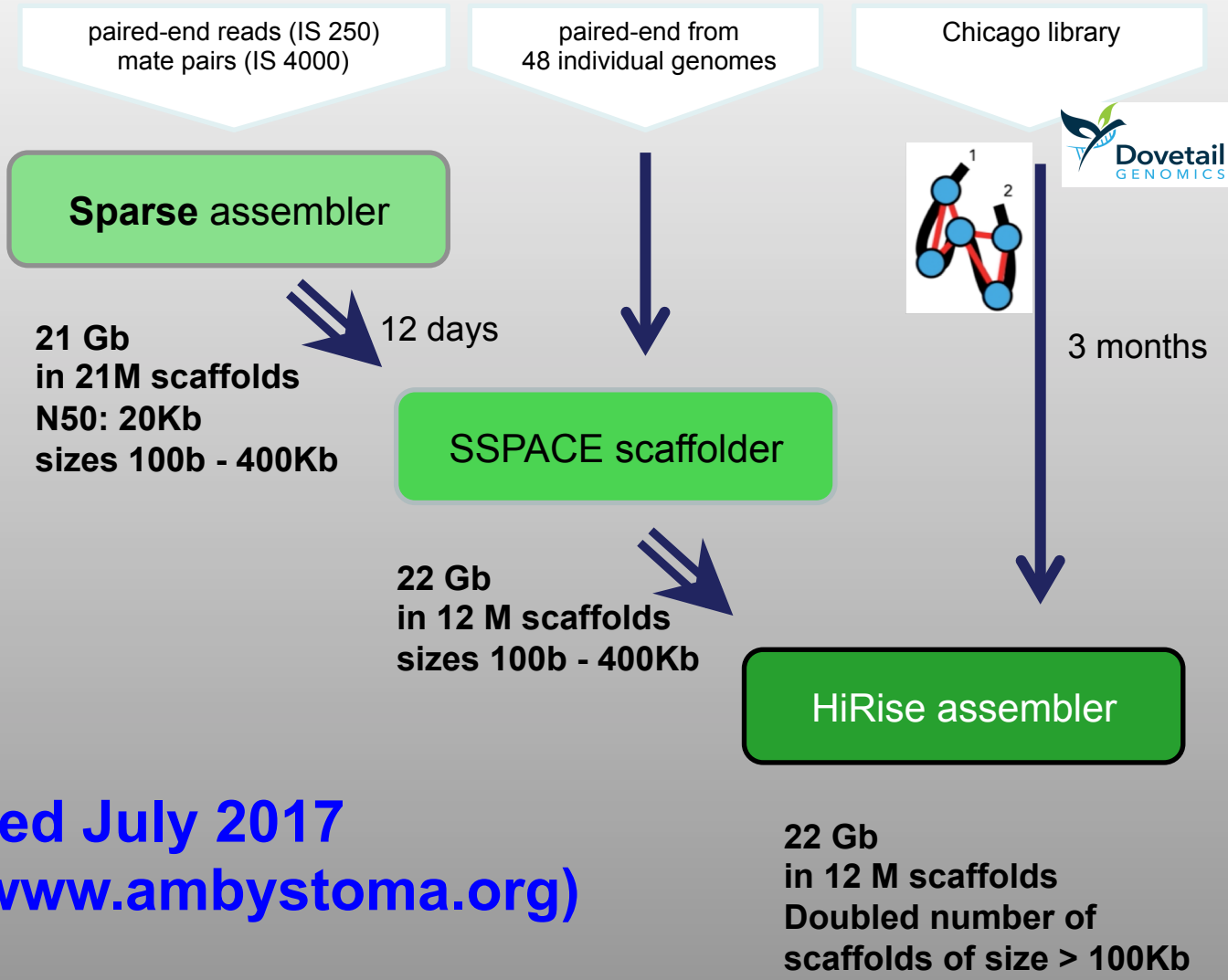


Assembling Moderate-Coverage (1.2T bases) Sequence Data

SPARSE de Bruijn graph to surmount memory limitations



Ye C, et al. *Exploiting sparseness in de novo genome assembly*, *BMC Bioinformatics*, 2012 13 (Suppl. 6):S1



Released July 2017
(<http://www.ambystoma.org>)

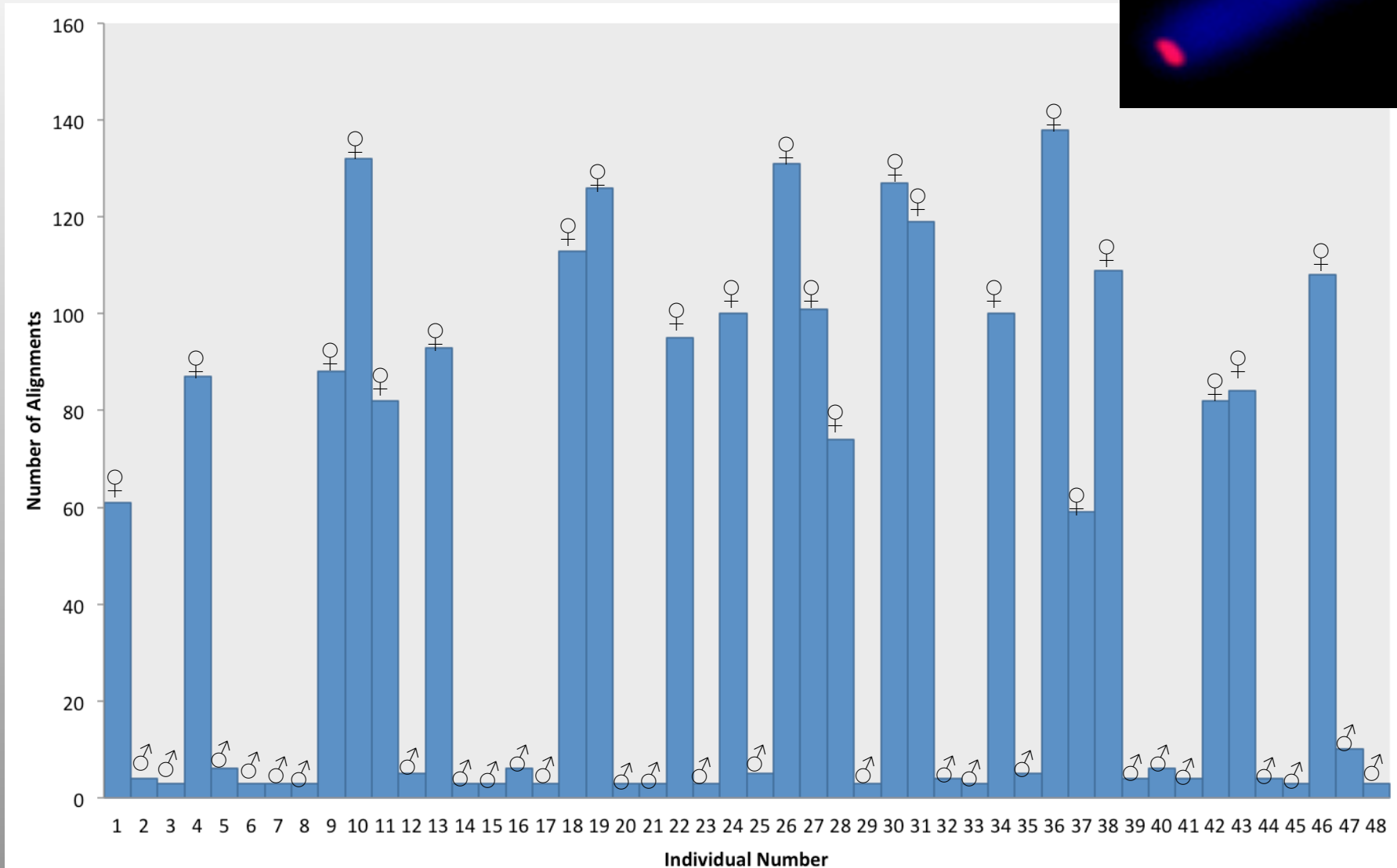
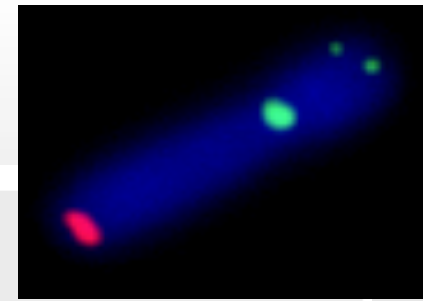
22 Gb
 in 12 M scaffolds
 Doubled number of
 scaffolds of size > 100Kb

Using the Assembly



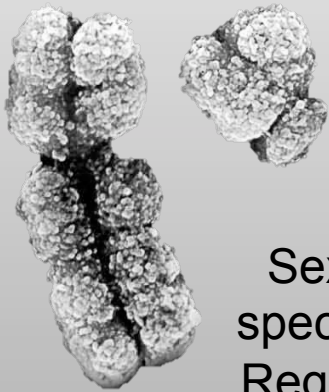
White / Edn3





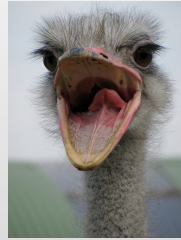


Human



Sex-specific
Region
~56.4Mb

56.4Mb/3.2Gb
 $\approx 1/55^{\text{th}}$ of genome



Ratite

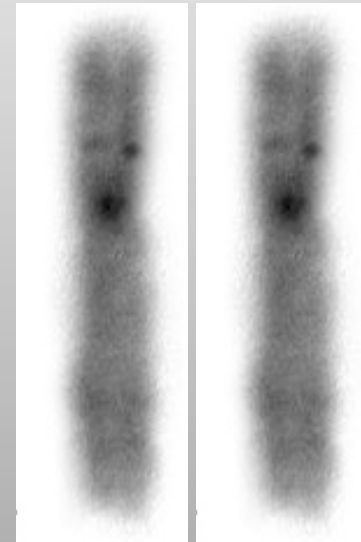


Sex-specific
Region
~30 Mb

30.0Mb/2.16Gb
 $\approx 1/70^{\text{th}}$ of genome



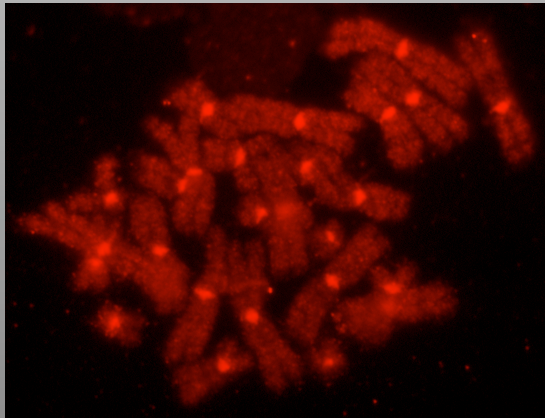
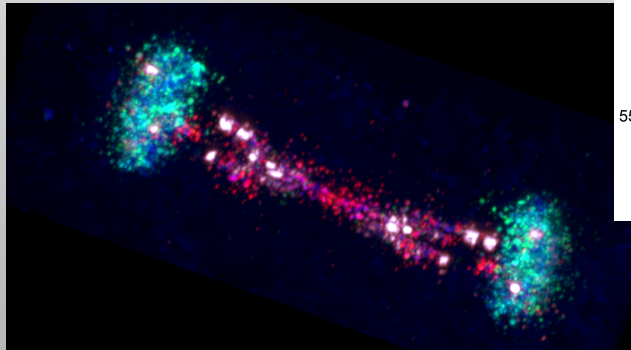
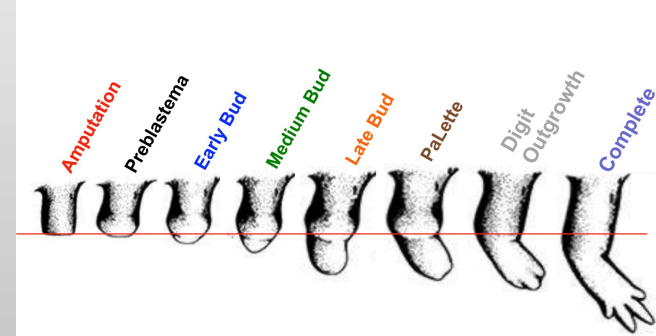
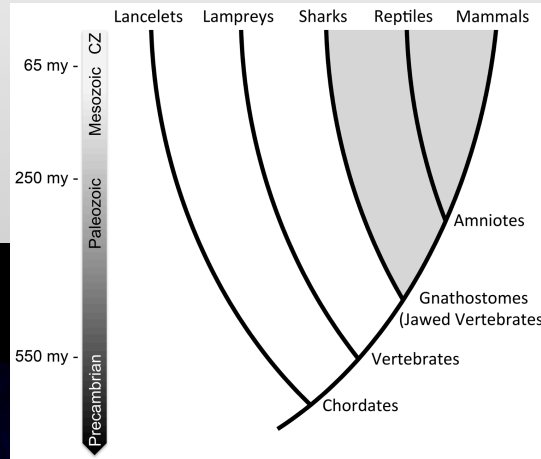
Axolotl



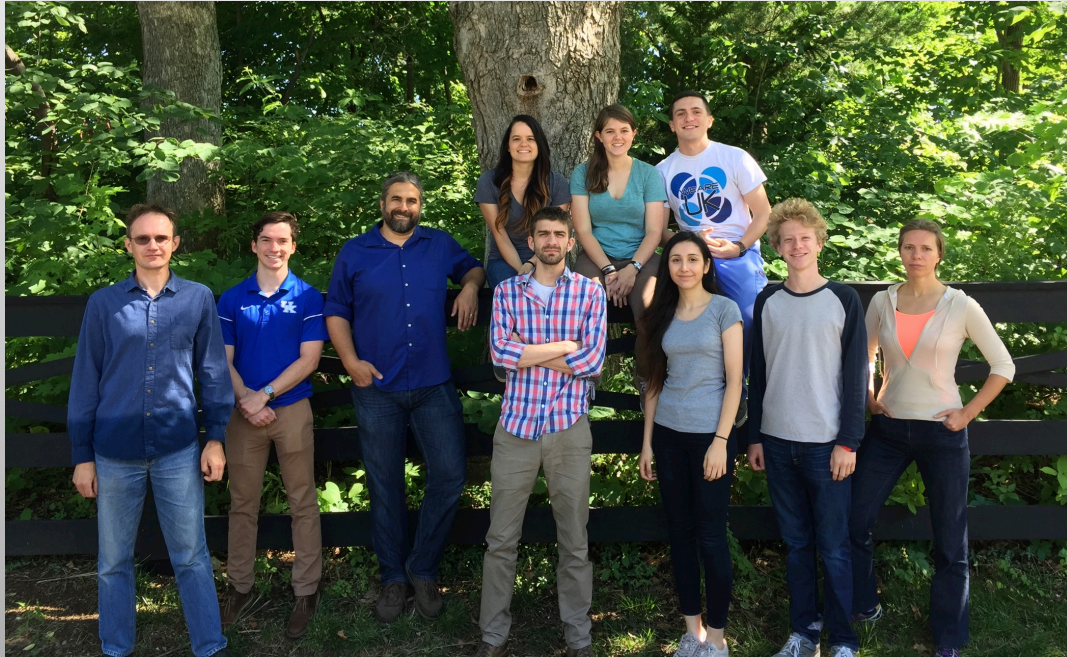
Sex-specific
Region
~300 Kb

300Kb/32Gb
 $\approx 1/100,000^{\text{th}}$ of genome

Salamanders and Lampreys) are Cool!



Acknowledgments



Lamprey Genome Consortium

Krumlauf Lab: *Robb Krumlauf, Leanne Wiedemann, Malcolm Cook, Hugo Parker; Chris Amemiya, Greg Elgar, Chengxi Li, Marianne Bronner, Tatjana Sauka-Spengler, GLFC, Evan Eichler, Francesca Antonacci, Ralph Lampman, Jon Hess ...*

Salamander Genome Consortium

Randal Voss Lab

Funding:



National Institute of
General Medical Sciences

F32GM087919

R01GM104123

R24OD010435

DOD W911NF-11-1-0475

