

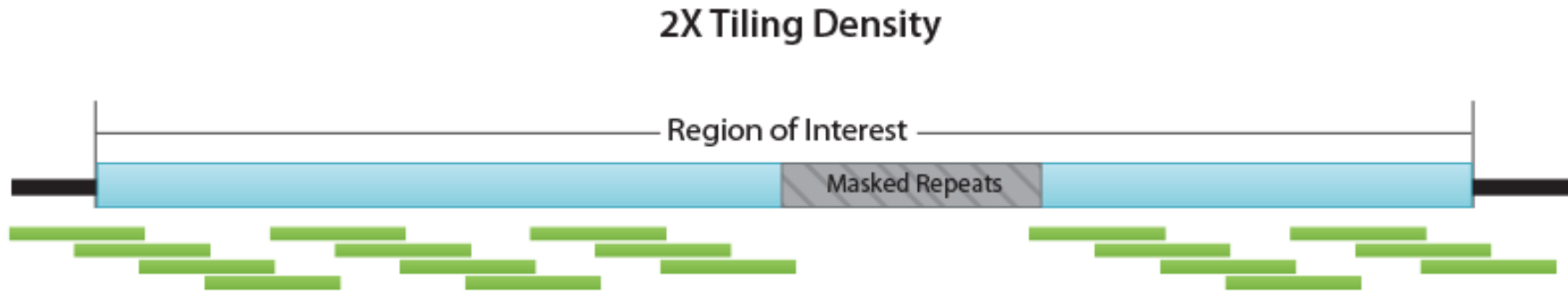


orthoCapture: Facilitating Gene Capture Probe Creation for Non-Model Organisms

M. Elise Lauterbur
Department of Ecology and Evolution
Stony Brook University
PhyloSynth Symposium 2018

How gene capture works

- Region of reference sequence
 - Guided by annotations or transcriptome → design sequence
- Overlapping 80 – 120 nt probes tiled across
-





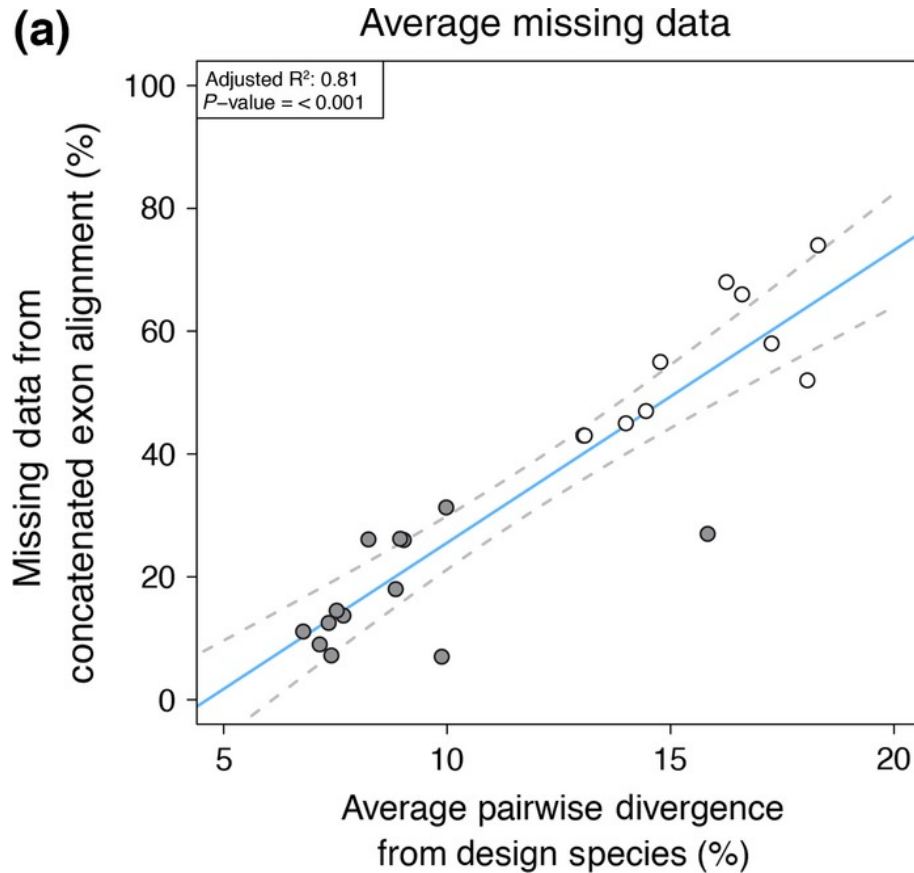
Limitations in comparative genomics

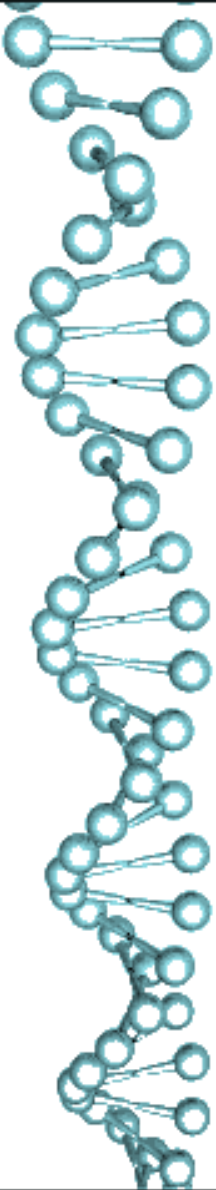
Capture efficiency **decreases** as evolutionary divergence between reference genome design sequence and species of interest **increases**

Jones and Good 2016. *Mol Ecol*;
Portik et al. 2016. *Mol Ecol Resour*;
Schott et al. 2017. *GBE*

Efficiency decrease with divergence increase

(a)



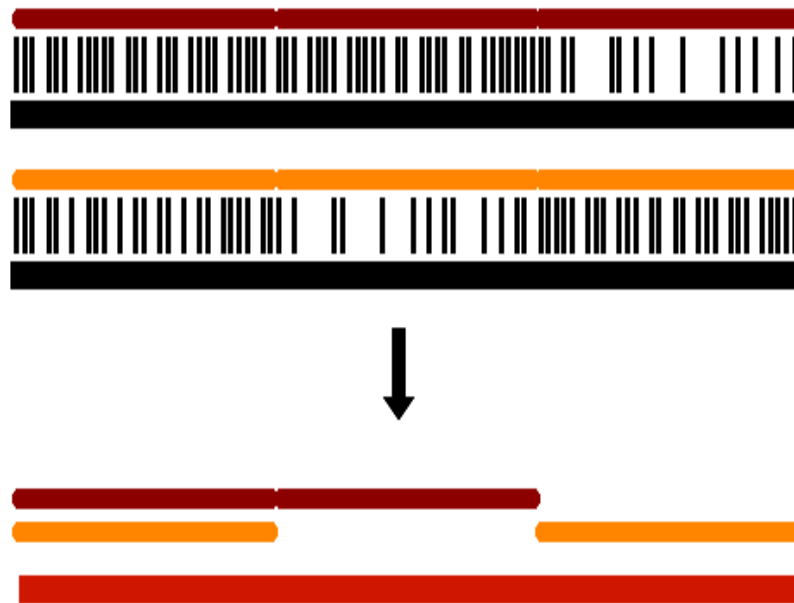


**What to do when we want to study
species without closely-related
model organisms?**

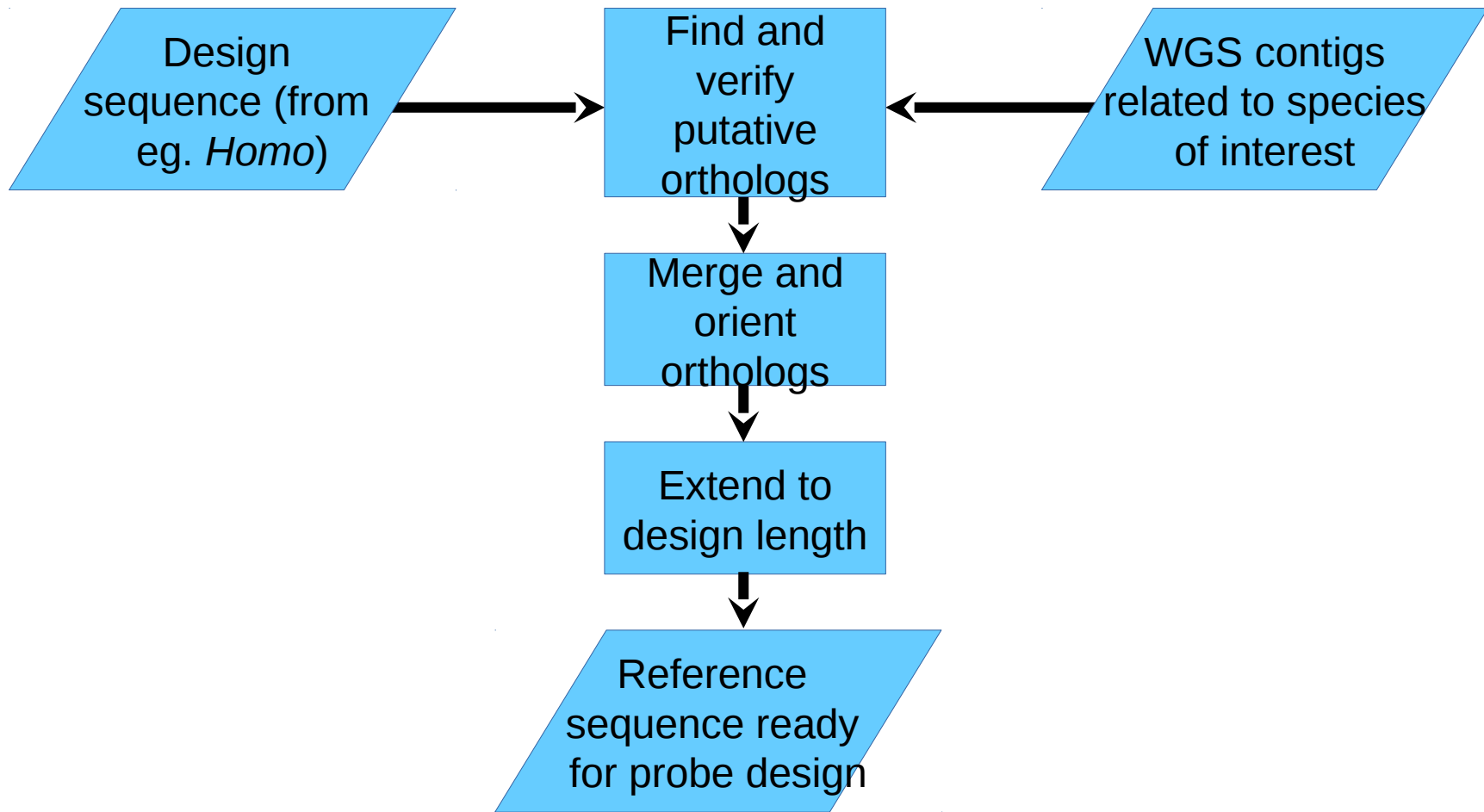
orthoCapture is...

- Given only:
 - Reference guide sequences from annotations (eg. *Homo*, *Mus*)
 - Contigs from ≥ 1 species related to species of interest
- 1.5-10x higher capture efficiency from similar method (Schott et al. 2017)
 - Extensive manual curation needed
 - Susceptible to errors from ENSEMBL and NCBI

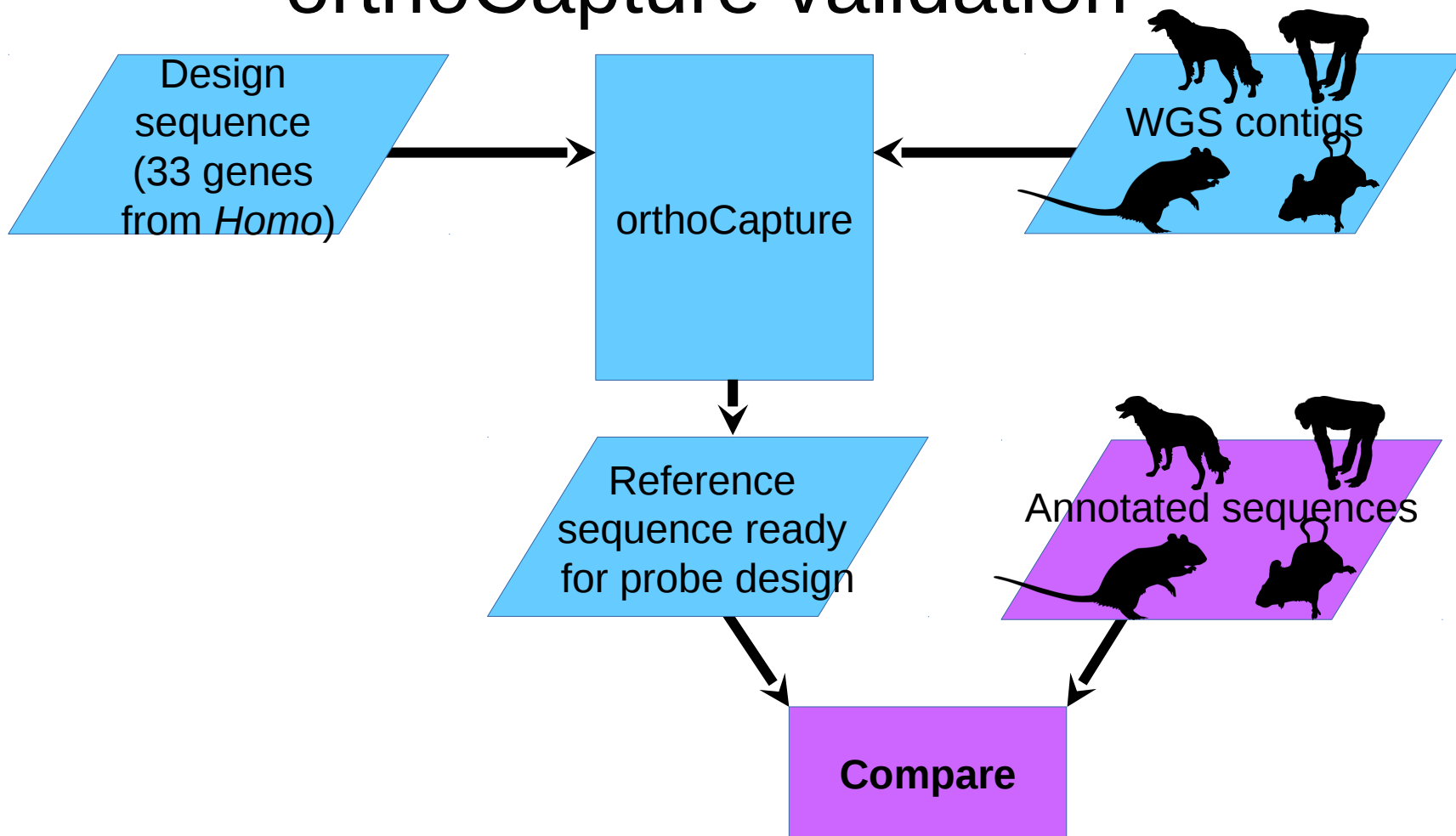
Two Probe Species



orthoCapture steps



orthoCapture validation

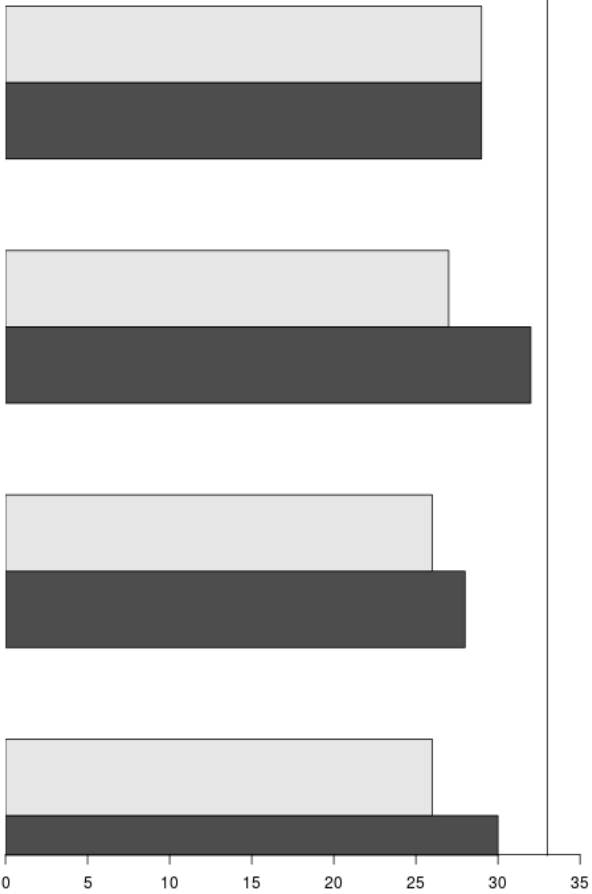
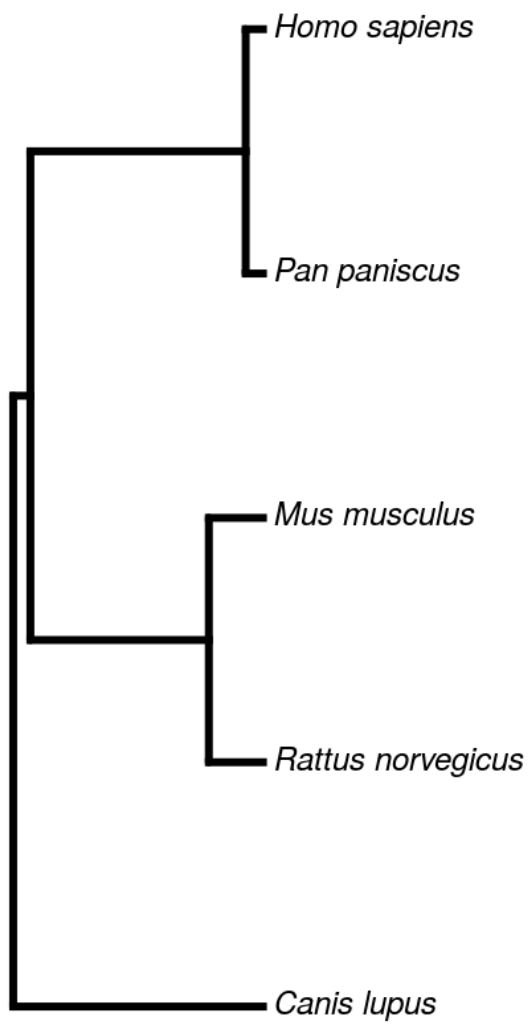


Homo reference

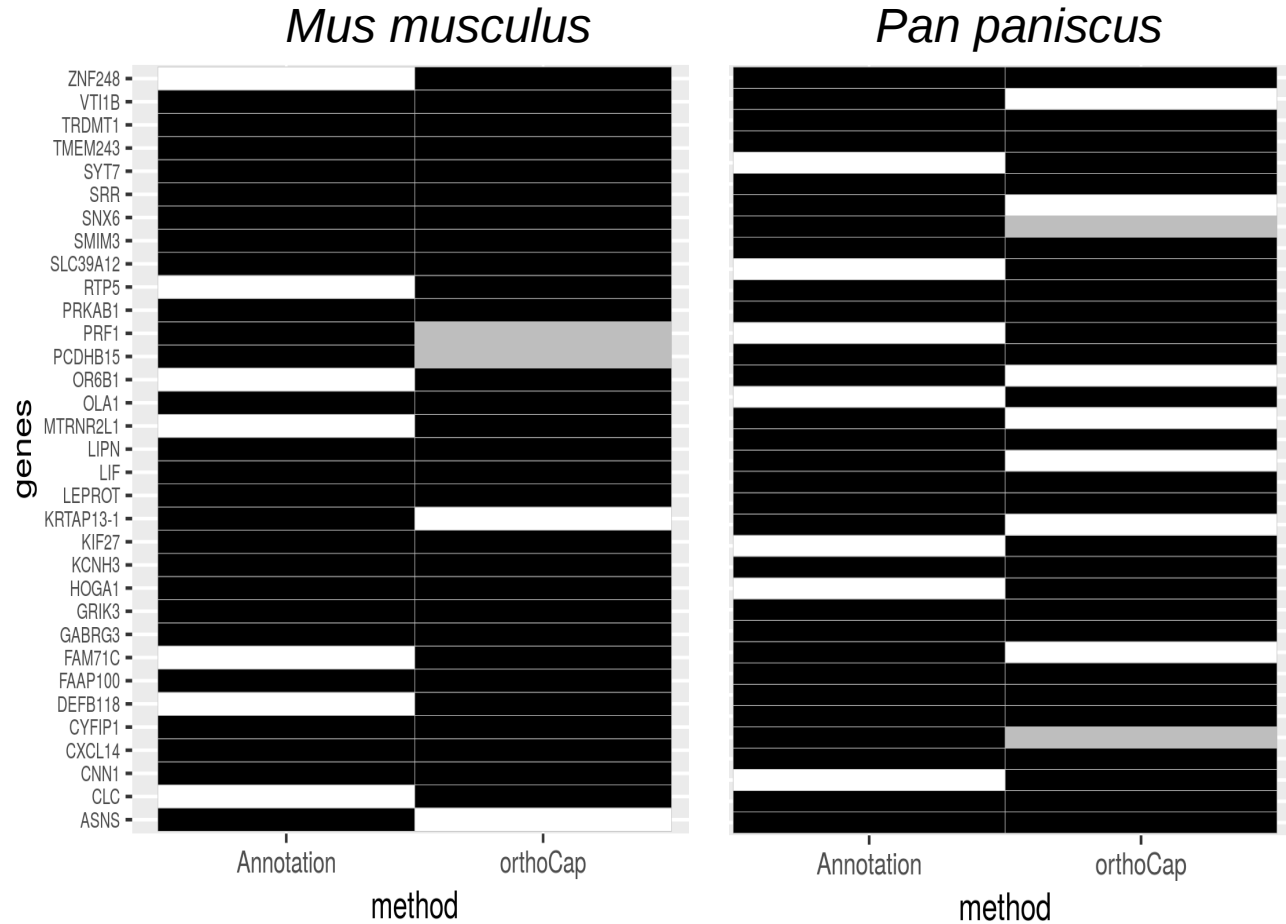


0 5 10 15 20 25 30 35

Number of genes retrieved

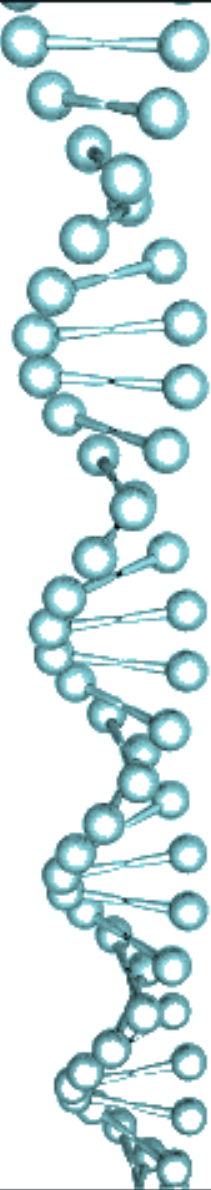


Genes retrieved



| | Genes missing | Genes mis-identified |
|-----------|---------------|----------------------|
| Ann. | 21-24% | - |
| ortho Cap | 6 – 21% | 6% |

Quality of exons retrieved



| Species | Number orthoCapture longer | Number annotation longer | Number equal length | Average percent identity |
|-------------------------|----------------------------|--------------------------|---------------------|--------------------------|
| <i>Canis familiaris</i> | 8 | 9 | 3 | 97 |
| <i>Mus musculus</i> | 8 | 5 | 10 | 100 |
| <i>Rattus rattus</i> | 6 | 3 | 9 | 100 |
| <i>Pan paniscus</i> | 8 | 9 | 3 | 97 |



Advantages

- Number of genes retrieved comparable to number available in well-annotated genome
- Can also retrieve non-coding sequences

Limitations

- Trade-off between probe diversity and # loci/cost
- Exact impact of WGS quality and divergence still being determined



orthoCapture

- Gene capture for species without reference genome
- **Increase** probe diversity → **increase** capture efficiency
- Available shortly at:
<https://github.com/lauterbur/orthoCapture>

Thank you



- Dr. Liliana Dávalos
- The Dávalos lab
- Stony Brook University HPC Seawulf and LI-Red
- Indiana University HPC Mason and Carbonate
- Funders:
 - American Society of Naturalists
 - King/Miller Travel Scholarship
 - Sigma Xi
 - NSF DEB-1442142
 - NSF DGE-1633299
 - PhyloSynth Symposium

