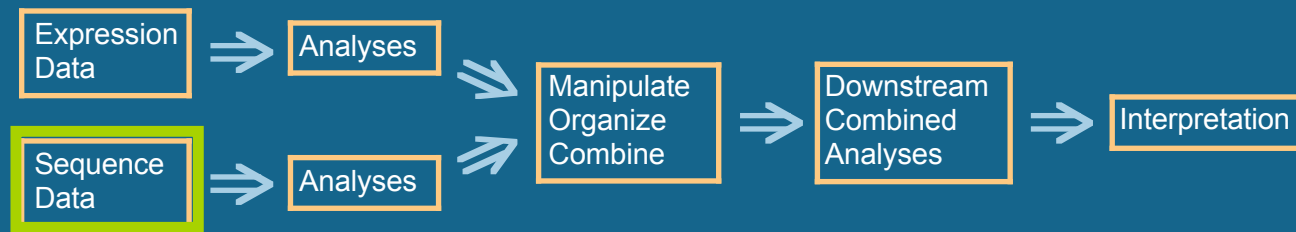


Adaptive gene expression divergence

1. Is expression divergence associated with adaptive cis-regulatory sequence evolution?
2. Are genes with expression divergence also fine-tuning protein function through adaptive evolution?
3. Are genes that change expression over short time scales clustered into distinct ontology groups?

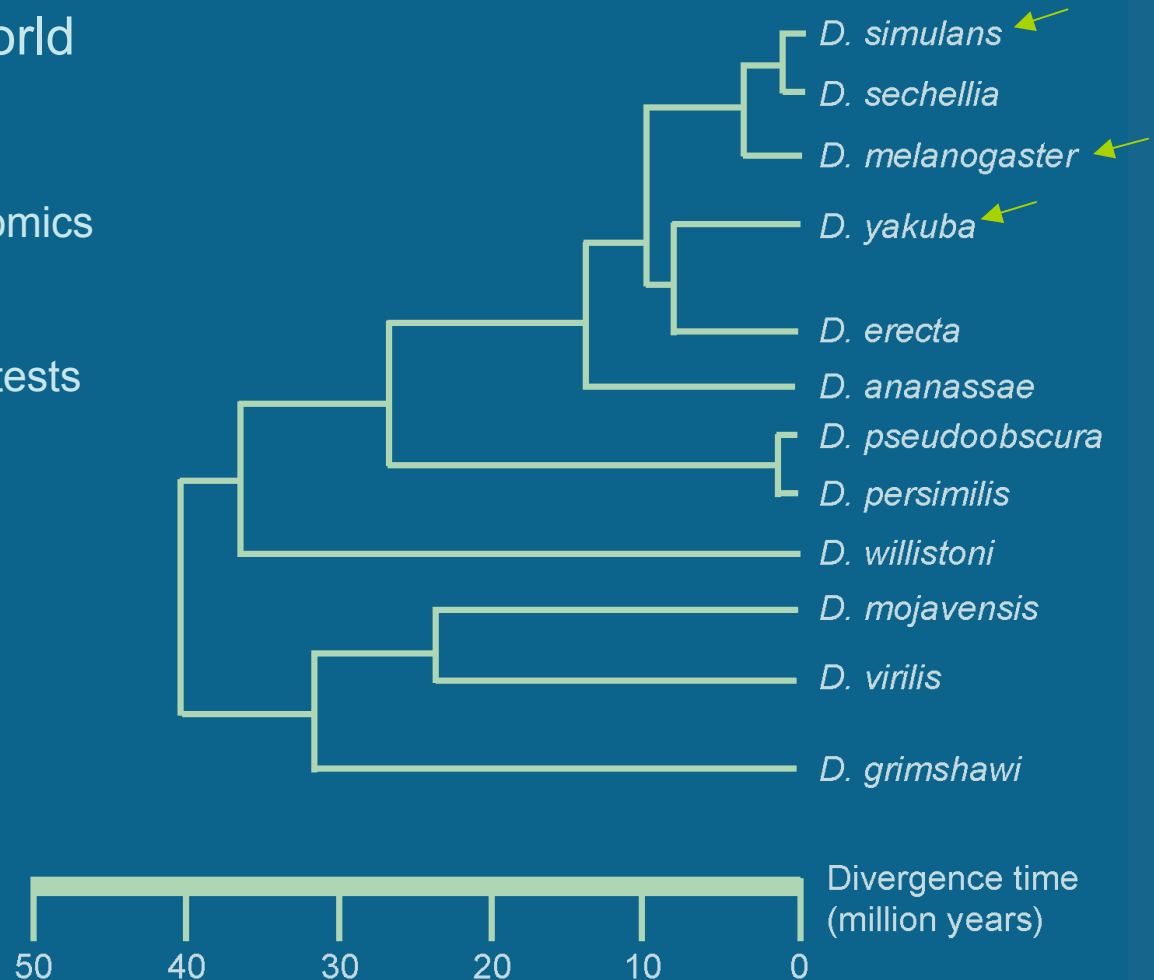
Combine genomic sequence data with expression array data.

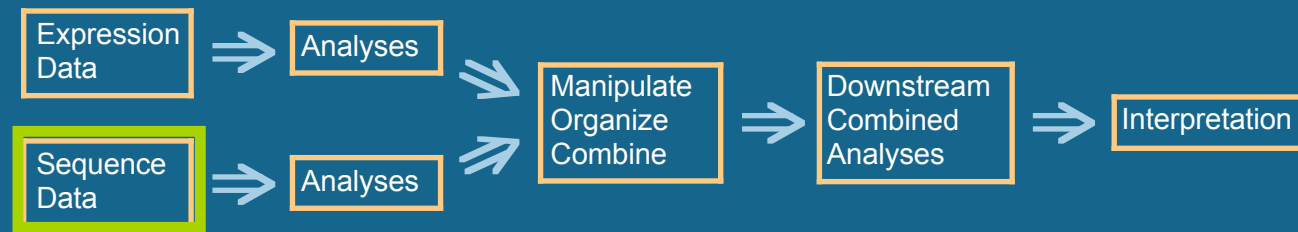


melanogaster-centric world

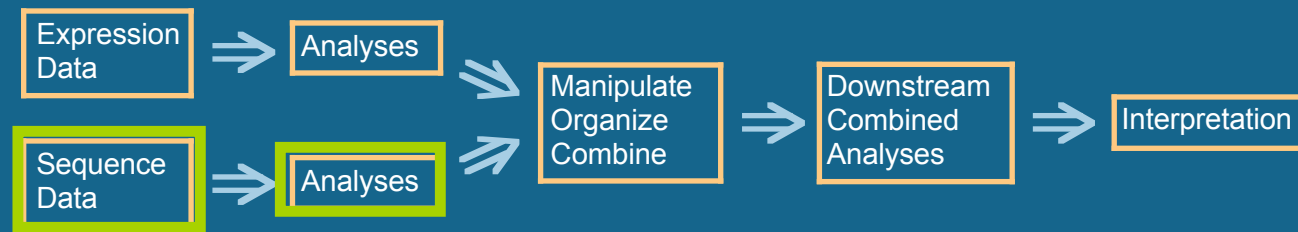
- Rich annotation ~14,000 genes
- Functional genetics/genomics
- Genome variation data
 - 6 lines *D. simulans*
 - molecular pop gen tests to infer forces of evolution

Silent sites:
 mel-sim 5%
 mel-yak 20%



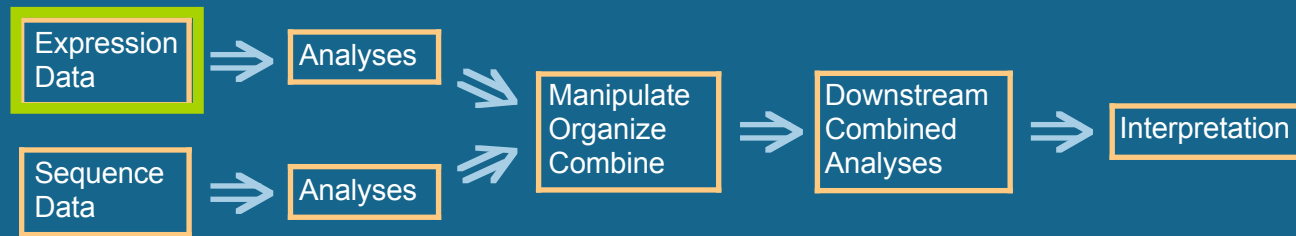


- *D. melanogaster* & *D. yakuba* - 8x coverage
- 6 *D. simulans* genomes - Light-shotgun - 1x coverage
 - Missing data - no *de novo* assembly
 - Syntenic alignment on *D. melanogaster* backbone
 - Recent duplications cause read pile-up
 - Lineage-specific duplications are eliminated
 - Inversions remain unmapped
 - Coordinate space preserved
 - Annotation - closely related, overlay mel features
 - Exons/Introns, UTRs, putative regulatory regions
 - Validation
 - Gene finding not a matter of typical gene prediction & best reciprocal blast hit
 - Gene models - initiation, splice, termination
 - ~1% imperfect gene models
 - New assembly releases - start over



For coding sequence, UTRs, and other putative regulatory regions:

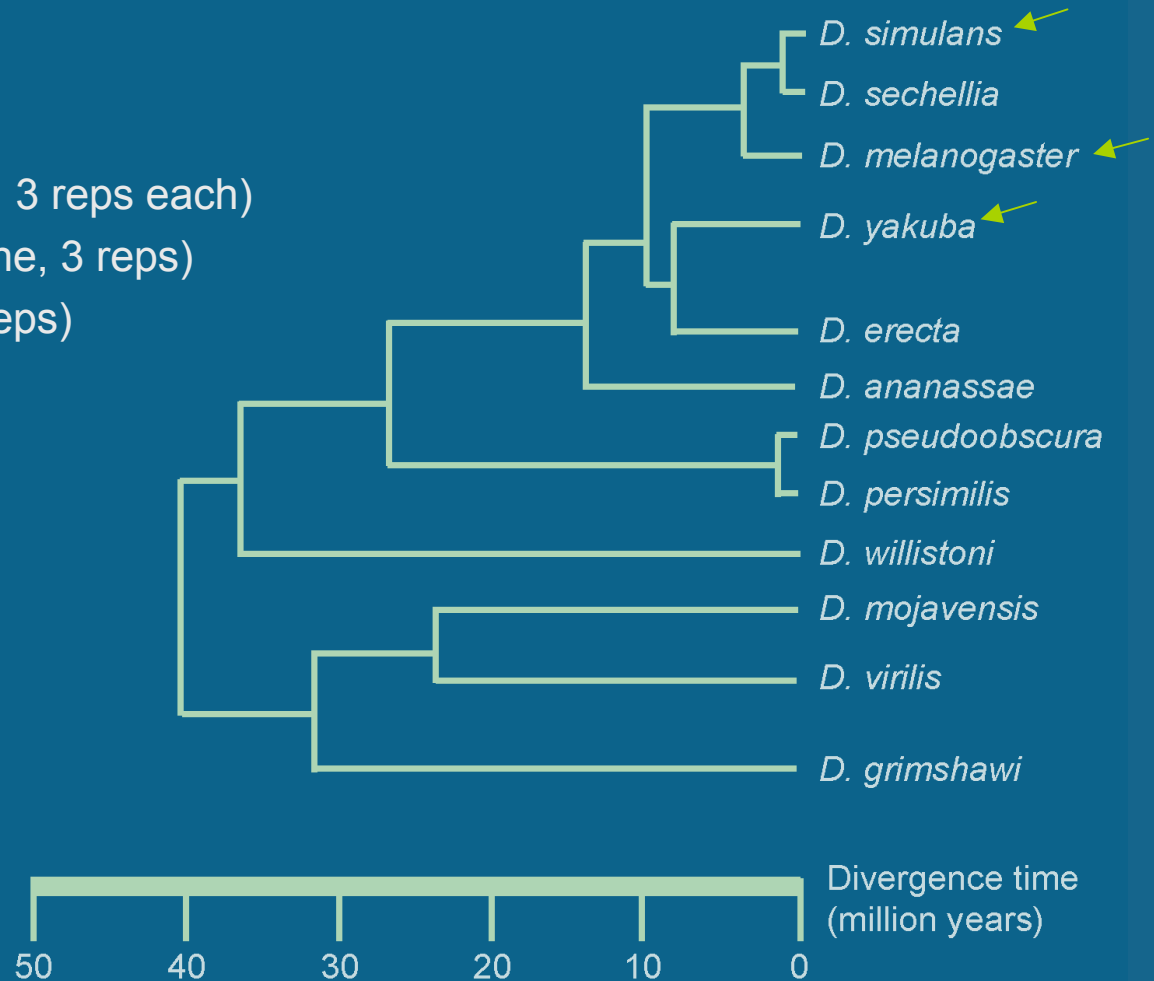
- Polymorphism
 - Analyze on site by site or codon by codon basis, not all pairwise differences because weights sites by coverage
 - Script to estimate π , batch
- Divergence
 - Codeml - Bioperl module
 - Baseml - ugly wrapper
- Contrast polymorphism:divergence in region of interest to “neutrally” evolving sites
 - No existing software for polarized and/or batch mode McDonald-Kreitman tests
 - Non-coding regions - compare to silent sites
 - Boneheaded perl scripts

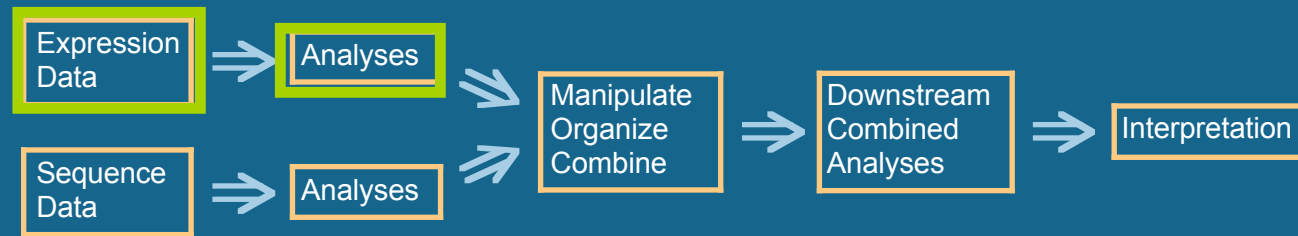


Expression Data:

Flies - adult males

- *D. simulans* (10 lines, 3 reps each)
- *D. melanogaster* (1 line, 3 reps)
- *D. yakuba* (1 line, 3 reps)

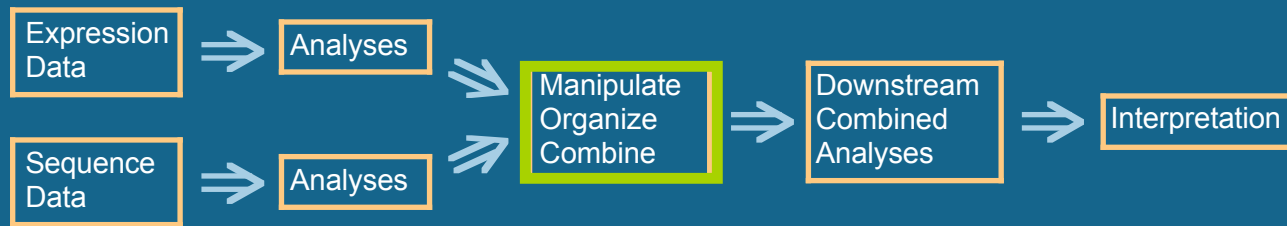




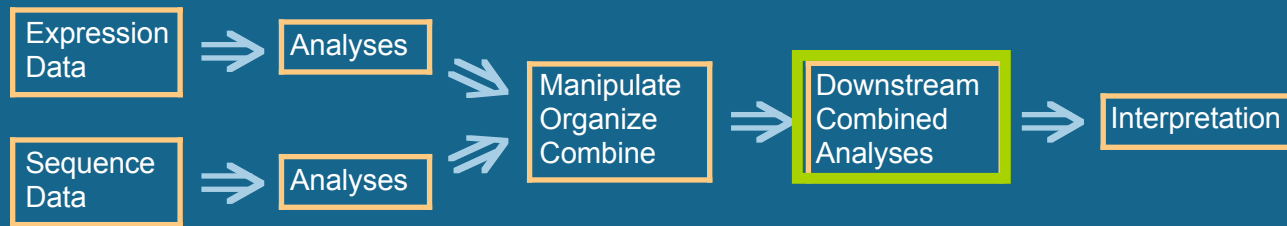
D. melanogaster expression micro-array

- Old assembly - validate probe targets
- Probe sequence divergence between mel and simulans/yakuba
 - confounds expression and sequence evolution
 - mask mismatched probes
- Analysis - gcrma through Bioconductor
- Reconstruct ancestral expression level
- Detect expression divergence

NO pipeline

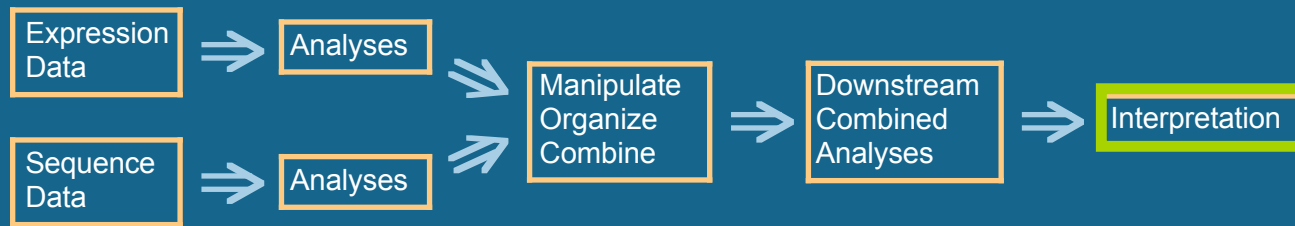


- **Database**
 - Query
 - Version control
- **Data for each element:**
 - Gene name
 - Flybase ID
 - Ontology information
 - Sequence data
 - sequences
 - heterozygosity - coverage, number of sites
 - divergence
 - Expression data
 - probe set
 - mean, se, num probes
 - each probe - expression value



Resampling tests:

- Are genes that are evolving adaptively overrepresented among those gene with significant expression divergence?
 - Are genes in particular ontology categories showing more expression divergence than random selection of genes?
1. Have n genes with significant expression divergence, calc prop of those genes evolving adaptively
 2. Random draw of n genes from all data
 3. Calc prop genes in random sample evolving adaptively
 4. Repeat 2&3 10,000 times to get empirical distribution



- Changes in gene expression are associated with recurrent adaptive evolution of putative 3' regulatory regions of genes.
 - miRNAs that regulate higher prop of genes with changes in expn do not evolve significantly faster ($\rho = 0.21$, $p = 0.1073$)
- Genes with changes in expression show evidence of recent adaptive evolution in coding regions - protein fine tuning?
- Major changes in expression in certain GO categories motivate deeper investigation of population genetic mechanisms acting on these genes.
 - Regulation of transcription & chromatin remodeling



Lineage-specific adaptive evolution & functional consequences

1. Identify bursts of evolution in historically conserved genomic regions
2. Test for mechanisms of evolution - adaptive vs. neutral
3. Investigate functional differences - hybrid incompatibilities?

