

Network analysis of modENCODE gene  
co-expression networks reveals  
“selection pressure on networks” and  
“ontogeny recapitulates phylogeny”

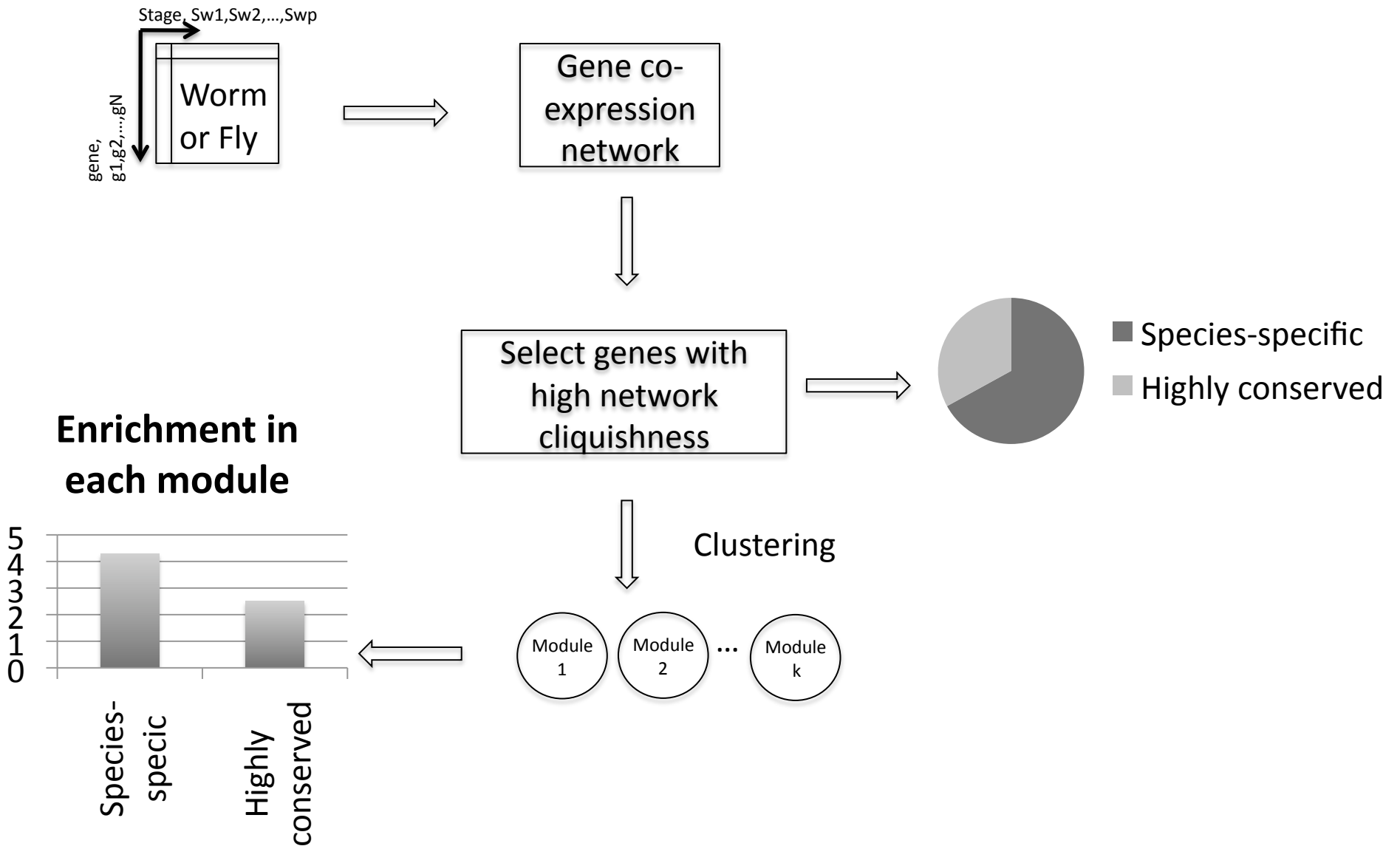
Daifeng Wang, Gang Fang, Mark Gerstein

July 17 2012

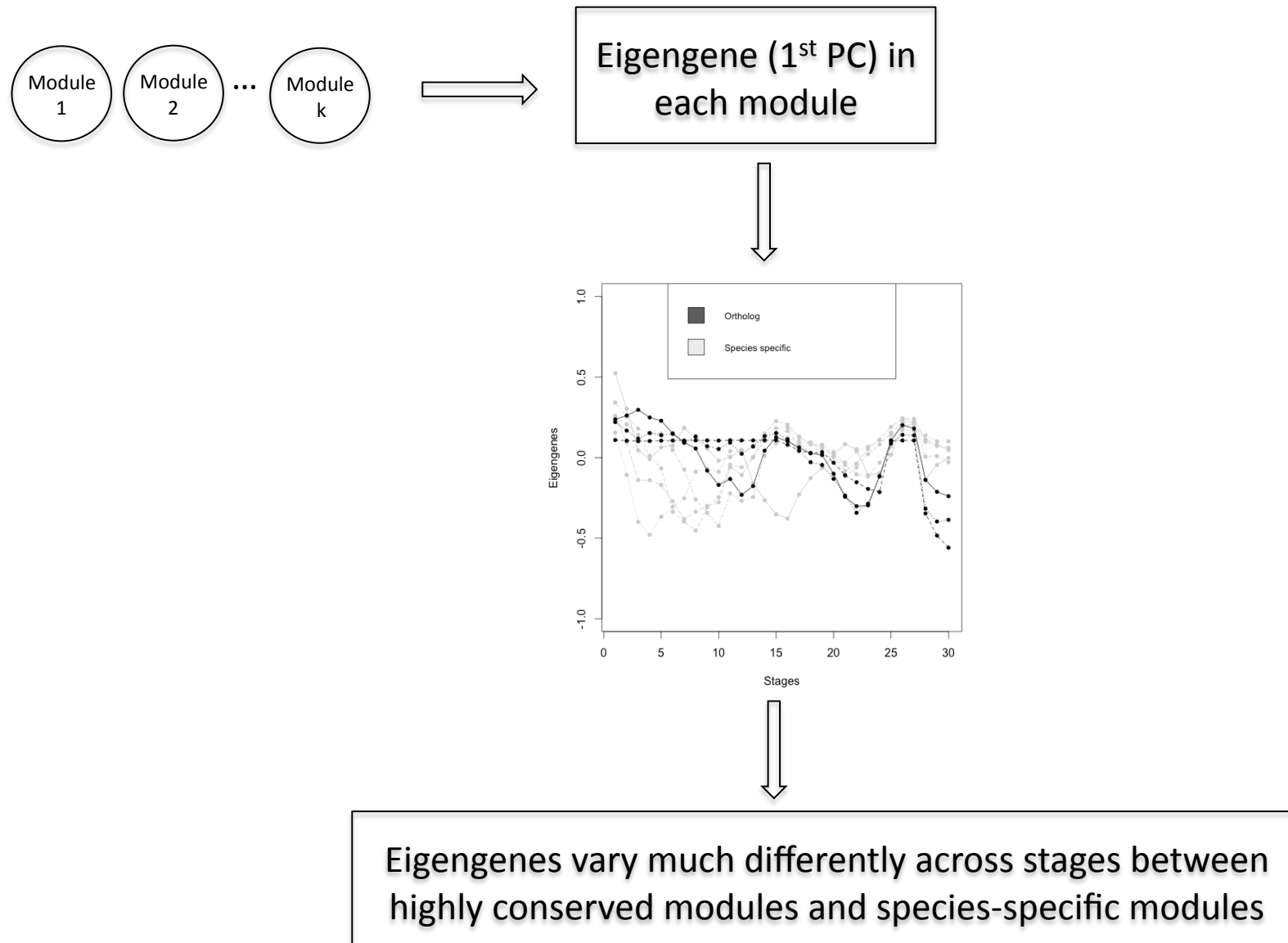
# Motivation

- Gene expression data in modENCODE
  - Species: worm vs. fly
  - Stages: embryo developmental stages
- Question 1:
  - How highly conserved genes vs. species-specific genes organize in gene co-expression network?
- Question 2:
  - How do they contribute to embryo development?

# Question 1: Selection pressure on networks



# Question 2: “ontogeny recapitulates phylogeny”



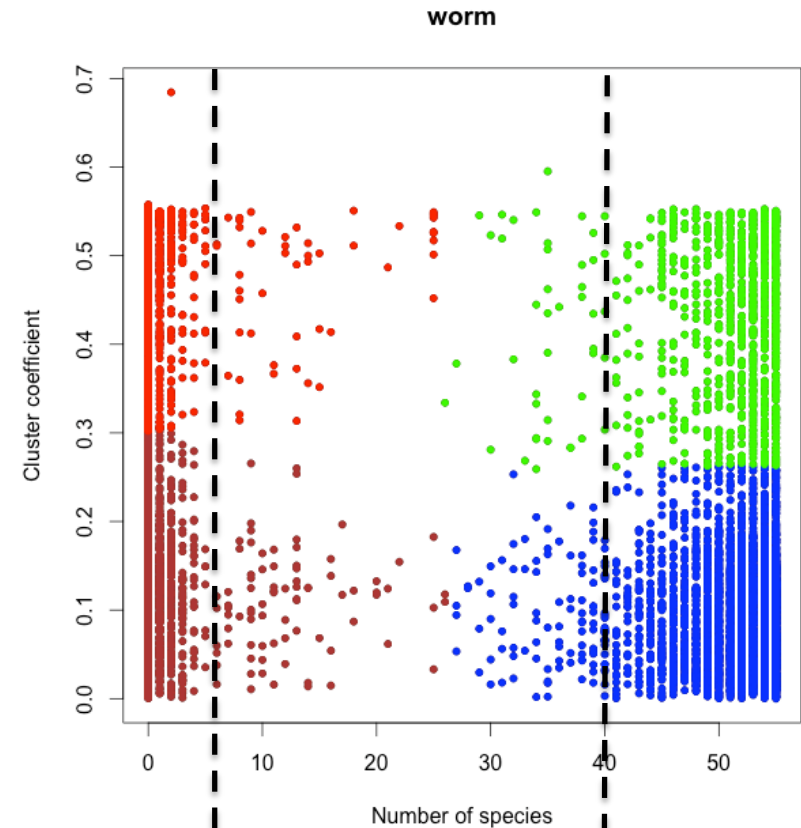
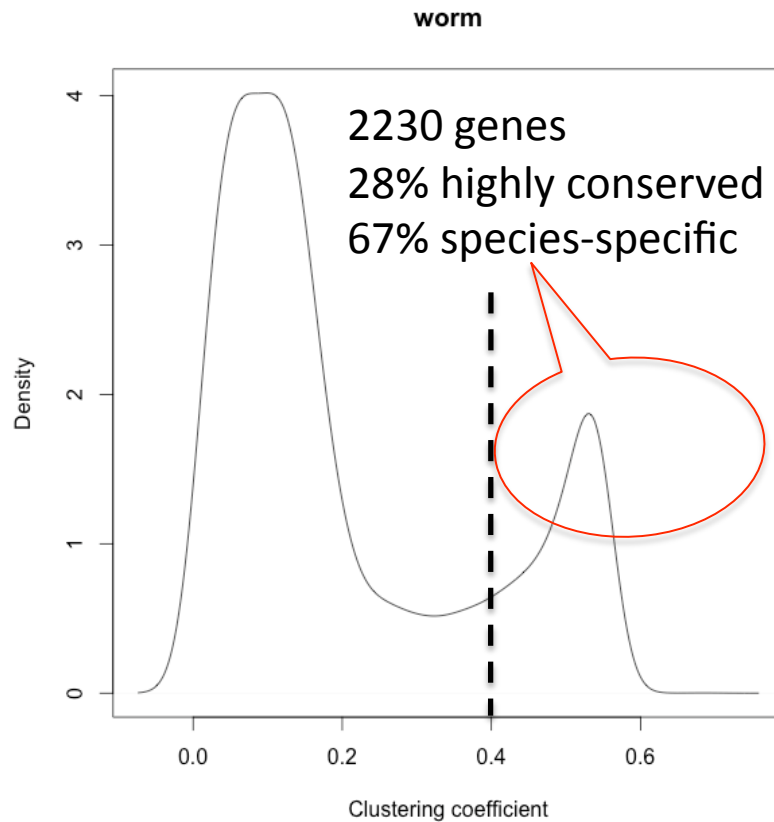
# modENCODE datasets

Species	# of Genes*	Developmental Stages**
Worm (C. ele.)	10450	33 stages: "N2_EE_50.0", "N2_EE_50.30", "N2_EE_50.60", "N2_EE_50.90", "N2_EE_50.120", "N2_EE_50.150", "N2_EE_50.180", "N2_EE_50.210", "N2_EE_50.240", "N2_EE_50.300", "N2_EE_50.330", "N2_EE_50.360", "N2_EE_50.390", "N2_EE_50.420", "N2_EE_50.450", "N2_EE_50.480", "N2_EE_50.510", "N2_EE_50.540", "N2_EE_50.570", "N2_EE_50.600", "N2_EE_50.630", "N2_EE_50.660", "N2_EE_50.690", "N2_EE_50.720", "L1_N2_L1.1", "L2_N2_L2.4", "L3_N2_L3.1", "L4_L4b", "YA_N2_Yad.1", "AdultSPE9", "DauerEntryDAF2_DauerEntryDAF2.2_DauerEntryDAF2.1.1_DauerEntryDAF2.4.1", "DauerDAF2_DauerDAF2.2_DauerDAF2.2.1_DauerDAF2.5.1", "DauerExitDAF2.2_DauerExitDAF2.3.1_DauerExitDAF2.6.1"
Fly (D. dmel.)	11235	30 stages: "em0.2hr", "em2.4hr", "em4.6hr", "em6.8hr", "em8.10hr", "em10.12hr", "em12.14hr", "em14.16hr", "em16.18hr", "em18.20hr", "em20.22hr", "em22.24hr", "L1", "L2", "L3_12hr", "L3_PS1.2", "L3_PS3.6", "L3_PS7.9", "WPP", "WPP_12hr", "WPP_24hr", "WPP_2days", "WPP_3days", "WPP_4days", "AdF_Ecl_1day", "AdF_Ecl_5days", "AdF_Ecl_30days", "AdM_Ecl_1day", "AdM_Ecl_5days", "AdM_Ecl_30days"

\*genes without invalid samples

\*\* Stages selected from the finalized modENCODE RNA-seq datasets in June 2012

# Worm: network cliquishness vs. preservation

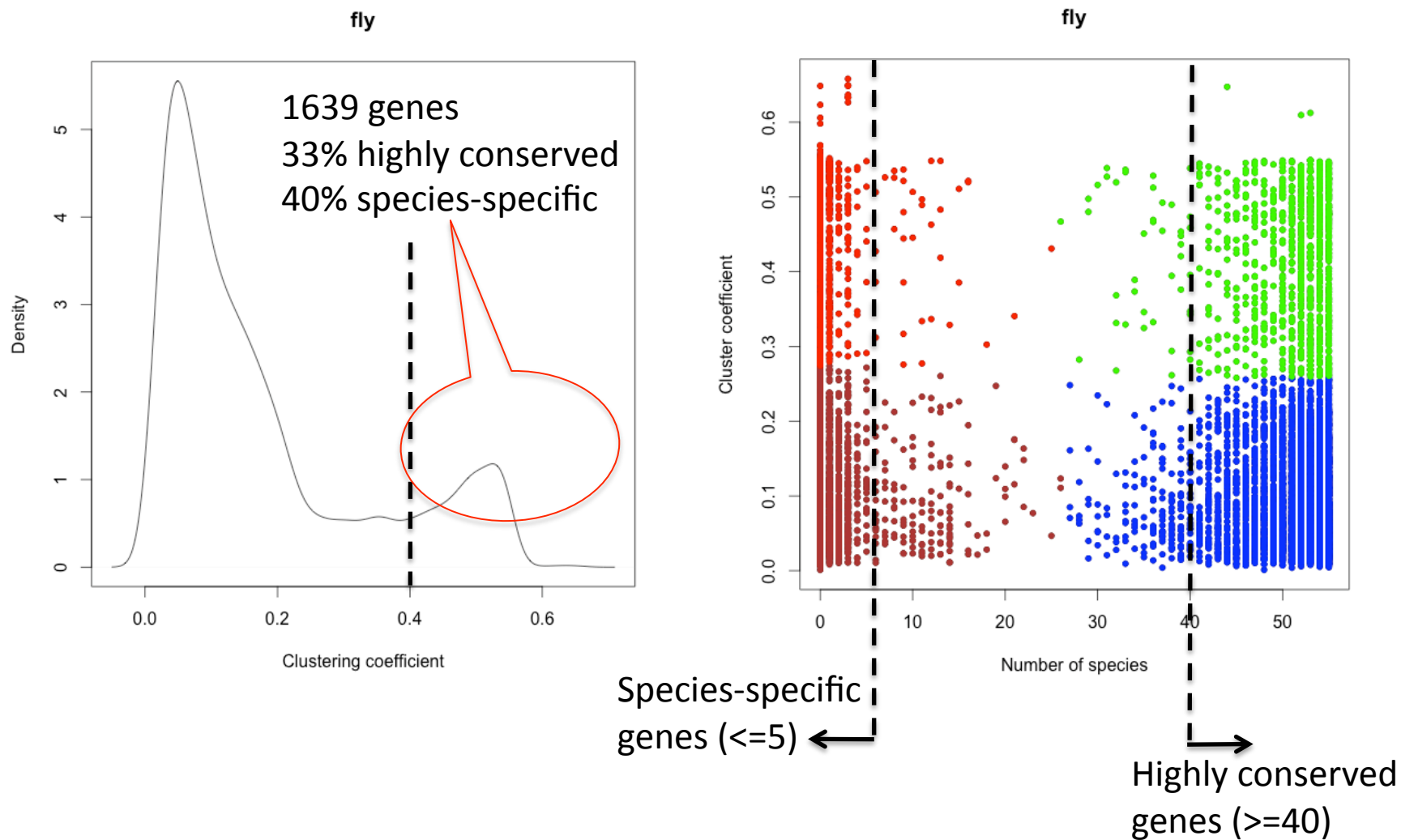


Species-specific  
genes ( $\leq 5$ )\* ←

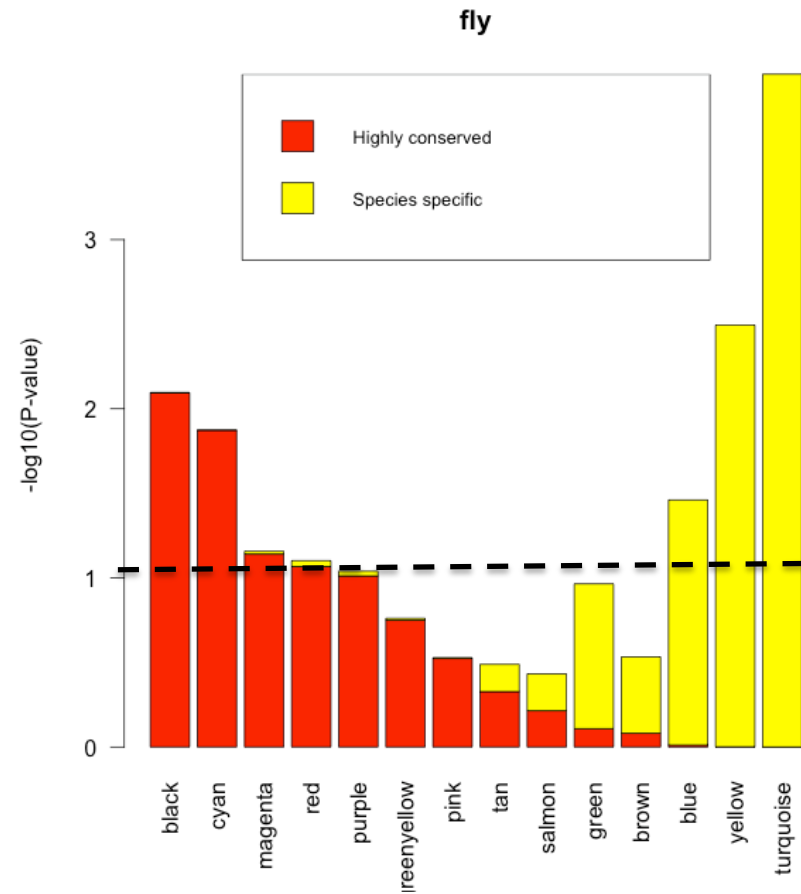
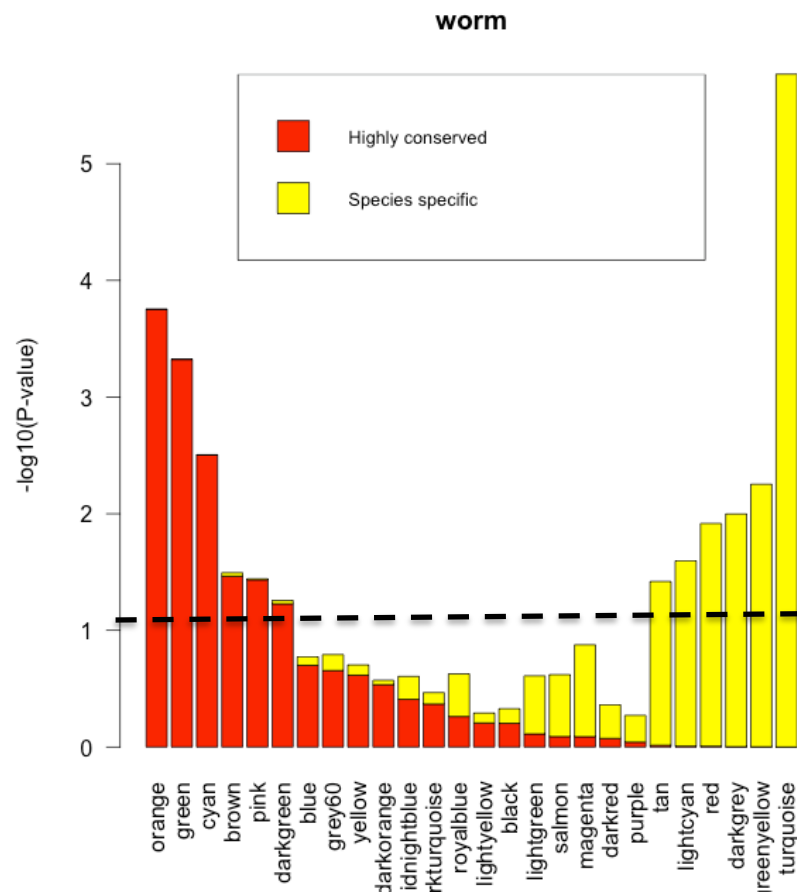
→  
Highly conserved  
genes ( $\geq 40$ )\*

\* From ensembl

# Fly: network cliquishness vs. preservation



# Enrichments of modules via clustering genes with high cliquishness



***50% modules enriched with either highly conserved or species-specific genes***



# Discussion 1

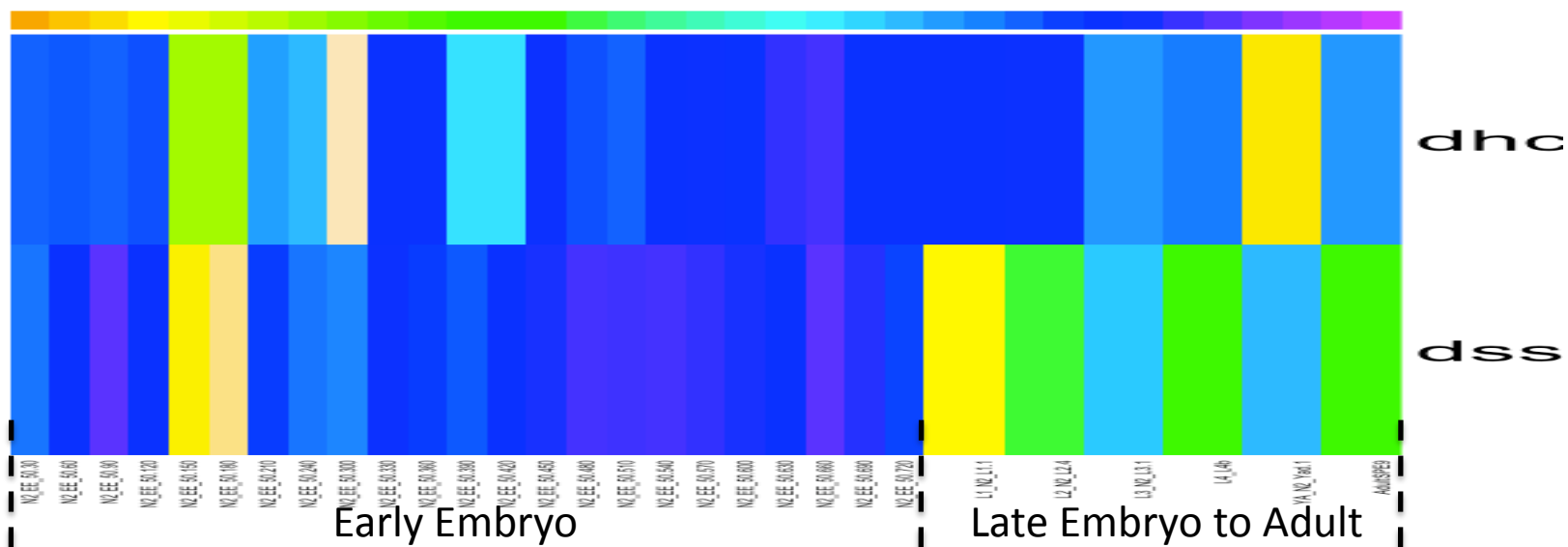
- Highly conserved modules
  - evolutionary pressure in on biological networks, not on individual genes
  - important biological functional coupling of the genes maintaining gene sequence conservation during evolution
- Species-specific modules
  - associated with phylotypes
- *To compare gene regulation across stages between those modules....*

# Measurement of gene regulation variation

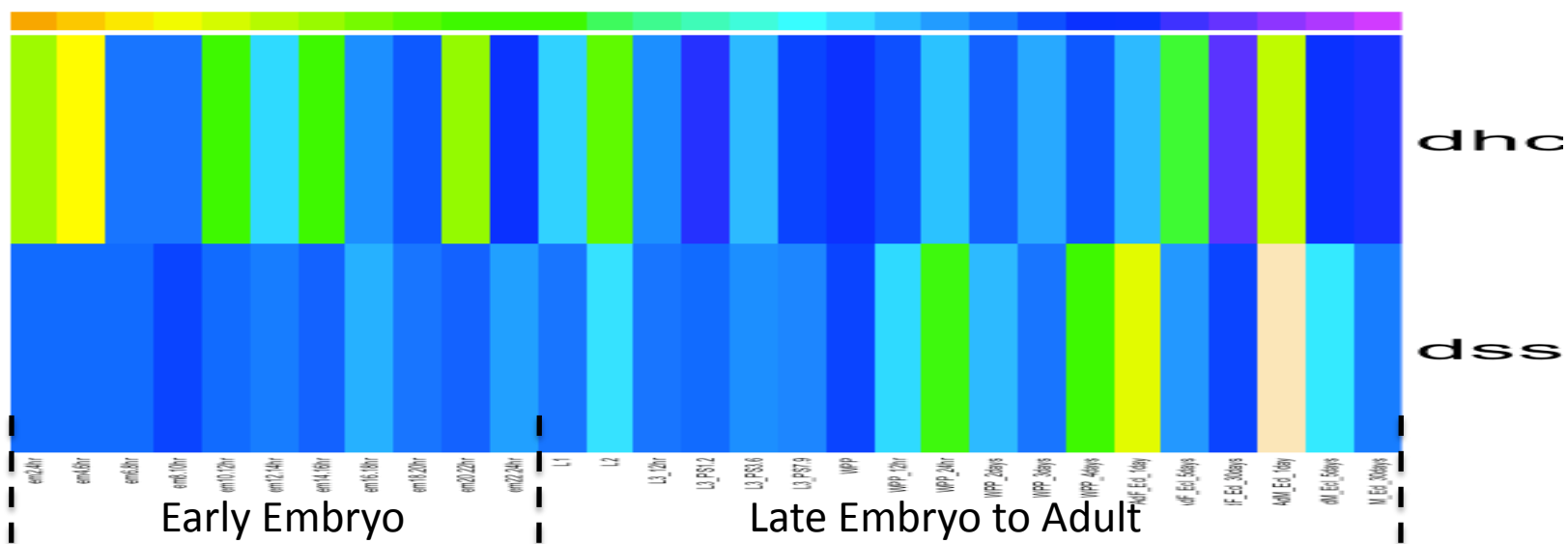
- Eigengene in each module
  - 1<sup>st</sup> PC, capture the most co-varying temporal direction in module
  - $E_i(t_j)$ , the expression value of eigengene at stage  $t_j$  of module  $i$
- Highly conserved modules
  - $d_{hc}(t_j) = \text{Mean}(\text{abs}(E_i(t_j) - E_i(t_{j-1})))$  for all highly conserved  $i$
- Species-specific modules
  - $d_{ss}(t_j) = \text{Mean}(\text{abs}(E_i(t_j) - E_i(t_{j-1})))$  for all species-specific  $i$

# $d_{hc}$ vs. $d_{ss}$

Worm



Fly



# Discussion 2

- Early embryo developmental stages
  - highly conserved > species-specific in variation
  - highly conserved genes are the major regulation targets and play more significant roles
- Late embryo to adult stages
  - species-specific > highly conserved in variation
  - species-specific genes mainly function during the phenotype developmental stages
- Support “ontogeny recapitulates phylogeny”