

# Promoter architecture and sex-specific gene expression in the microcrustacean *Daphnia pulex* revealed by profiling of 5'-mRNA ends



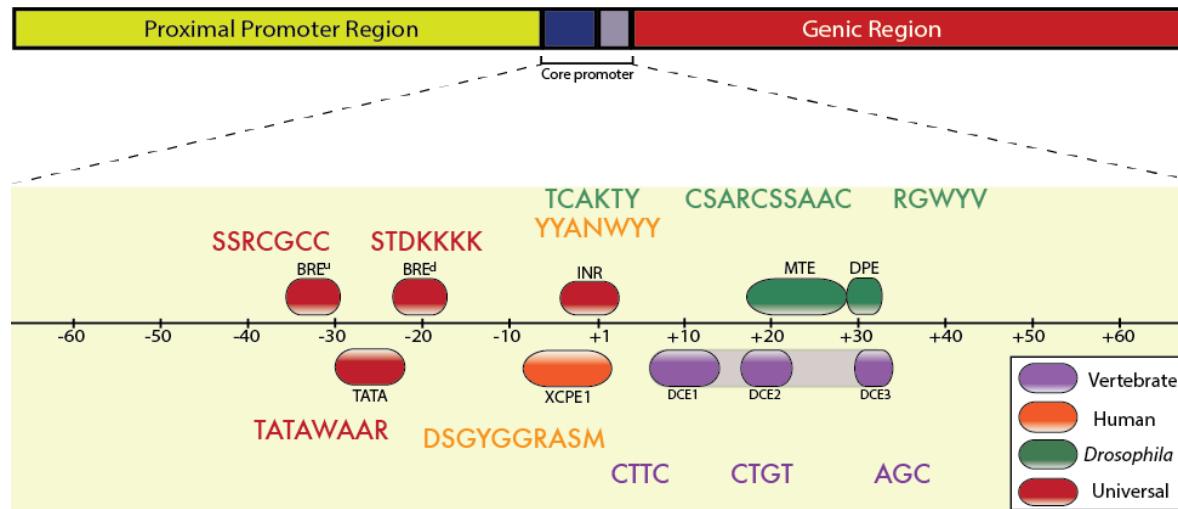
**R. Taylor Raborn**

Postdoctoral Fellow  
Department of Biology,  
Indiana University  
@rtraborn

Regulatory Genomics SIG

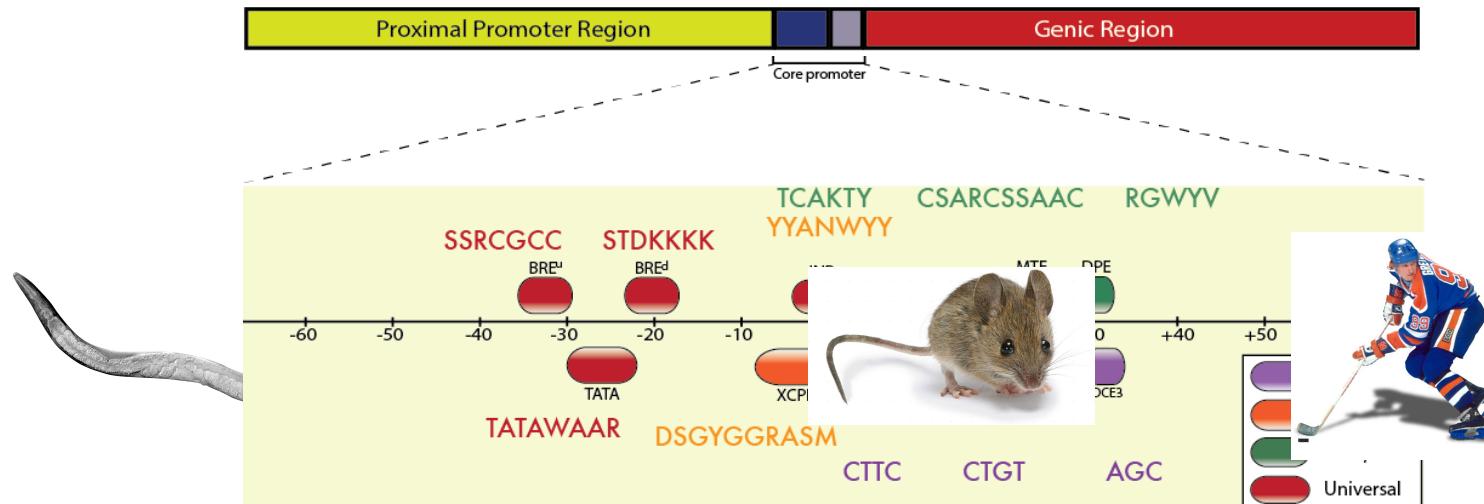


# The composition of metazoan promoters is heterogeneous



Adapted from Lenhard *et al.*, 2012

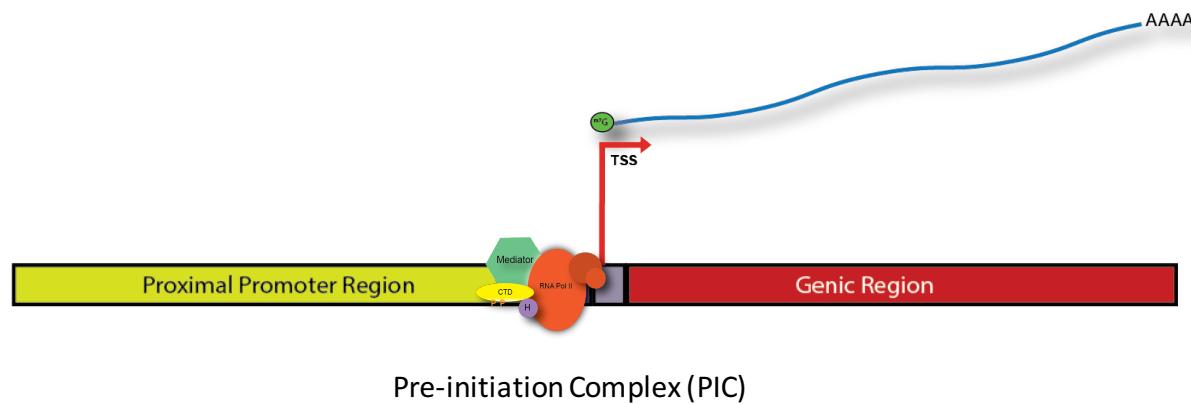
Our understanding of metazoan promoters is phylogenetically limited

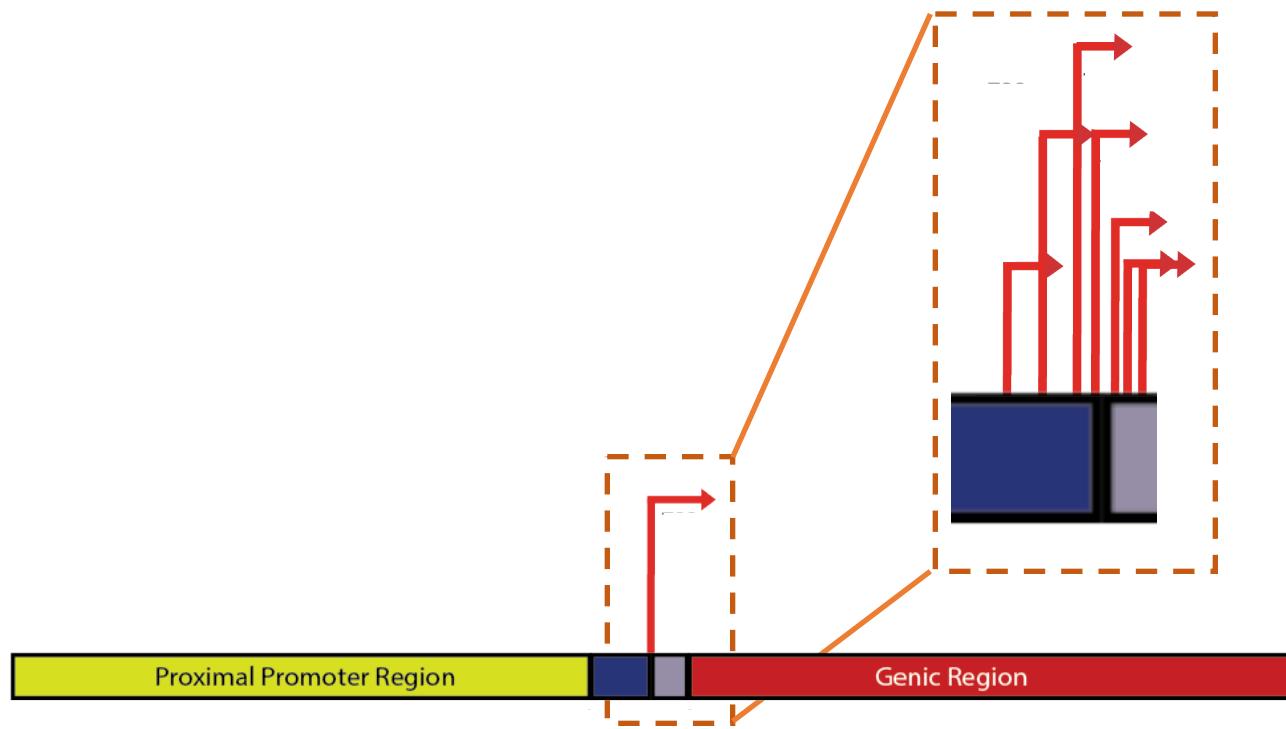


*Most of what is known about metazoan promoters comes from a small set of major model organisms.*

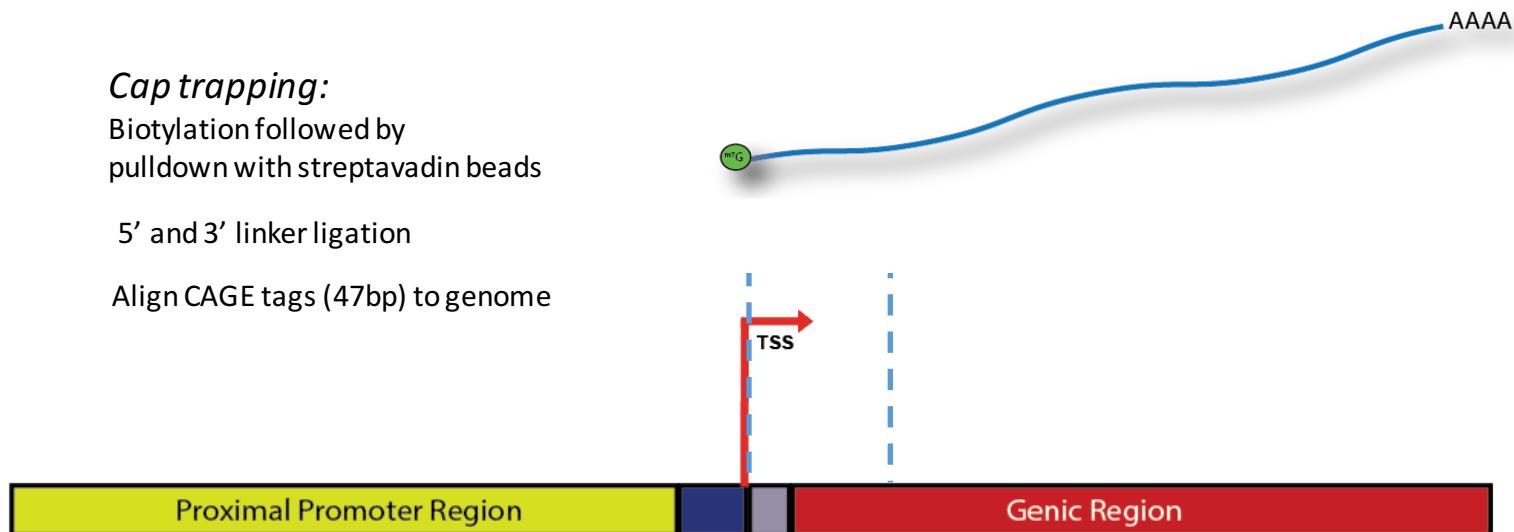
Adapted from Lenhard *et al.*, 2012

The transcription start site (TSS) provides a link between promoter and mRNA



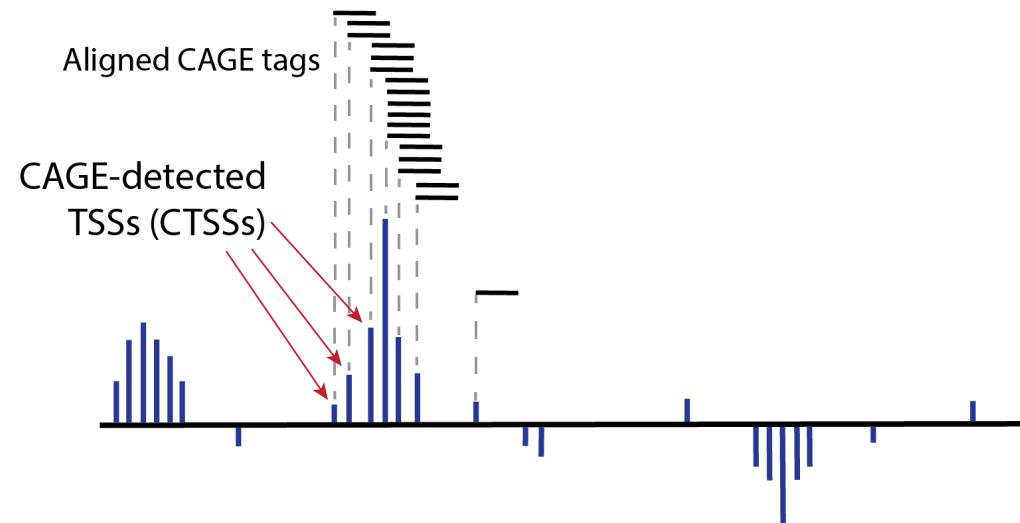


## Cap Analysis of Gene Expression (CAGE) identifies TSSs of mRNAs at large scale

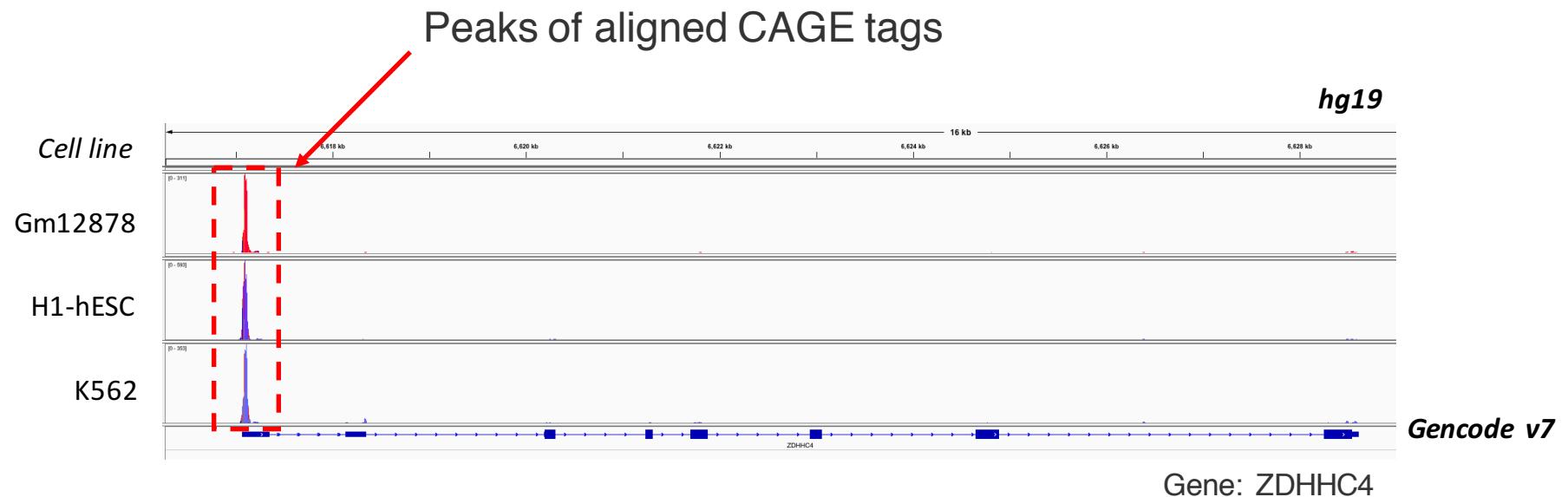


*We sought to expand the phylogenetic breadth of metazoan core promoters by applying CAGE in *D. pulex**

## Designations of CAGE-based information



# CAGE identifies the global landscape of transcription initiation



Snapshot from IGV. CAGE data from ENCODE/UCSC.

# *Daphnia pulex* is an emerging model organism for genetics

*Daphnia* spp. are globally distributed; nearly ubiquitous in freshwater habitats

Many diverse, well-characterized natural populations

A unique life cycle that includes both sexual and asexual modes of reproduction



*Daphnia*'s life cycle comprises sexual and asexual reproductive pathways

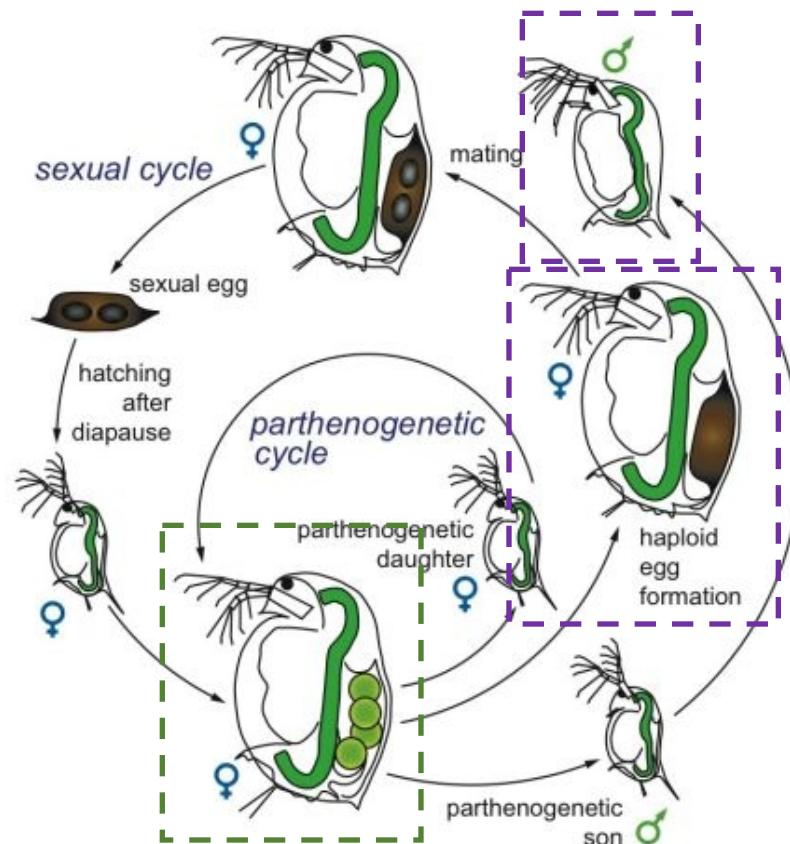


Image Credit: D. Ebert, 2005.

# *D. pulex* has a compact, gene-rich genome

Assembly Summary (“TCO”)

Criterion%	<i>D. pulex</i> JGI v1.1
Number of scaffolds	5,191
Ave. length of scaffolds (bp)	9,660
Length of N50 scaffold (bp)	318,519
Length of largest scaffold (bp)	4,193,030
Length of scaffolds (bp)	197,261,574

%Note: only results from nuclear genome are shown

Annotation Summary

Number of chromosomes	12
Estimated genome length (bp)	260,000,000
Number of annotated genes	30,907
GC content (%)	41.6
Ave. intron length (bp)	170

# Culturing and RNA isolation of *D. pulex*

We collected *Daphnia* individuals from three states using manual culturing and size separation.

Total RNA isolated from whole individuals (n=50-100) using mechanical disruption and cell lysis.

Individuals derived from a clone (PA42) originating from a well-characterized natural population in Indiana (Lynch *et al.*, 1989).



with Ken Spitze, PhD

## Preparation and analysis of CAGE libraries

CAGE libraries (DNAForm, Japan) were made using biological replicates from three states in *D. pulex*.

Sequenced using the Illumina HiSeq 2000 platform (50bp, SE)

Aligned reads to *D. pulex* assembly TCO v1.1 using bwa (Li and Durbin, 2009)

Applied CAGE analysis pipeline: clustering and identification of CTSSs, TSRs and consensus promoters using CAGEr (Haberle *et al.*, 2014).

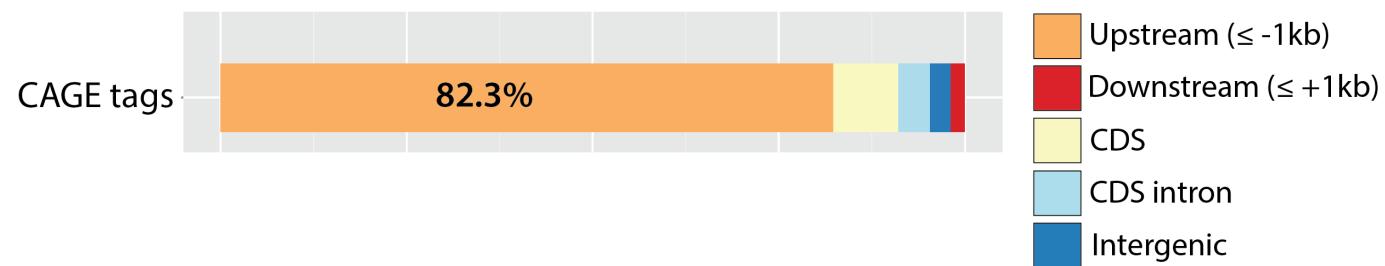
# Summary of sequenced CAGE libraries

Number	Library Name	Number of Sequenced CAGE Tags	Number of Mapped CAGE Reads
1	Asexual females-1	28,803,508	18,601,744
2	Asexual females-2	16,701,216	10,839,287
3	Asexual females-3	29,786,273	20,754,759
4	Sexual females-1	24,076,420	15,861,581
5	Sexual females-2	24,567,545	15,163,393
6	Sexual females-3	15,115,501	9,621,093
7	Males-1	18,512,317	12,412,516
8	Males-2	24,655,373	16,960,704
Total	—	182,218,153	120,215,077

## CAGE mapping and promoter identification summary

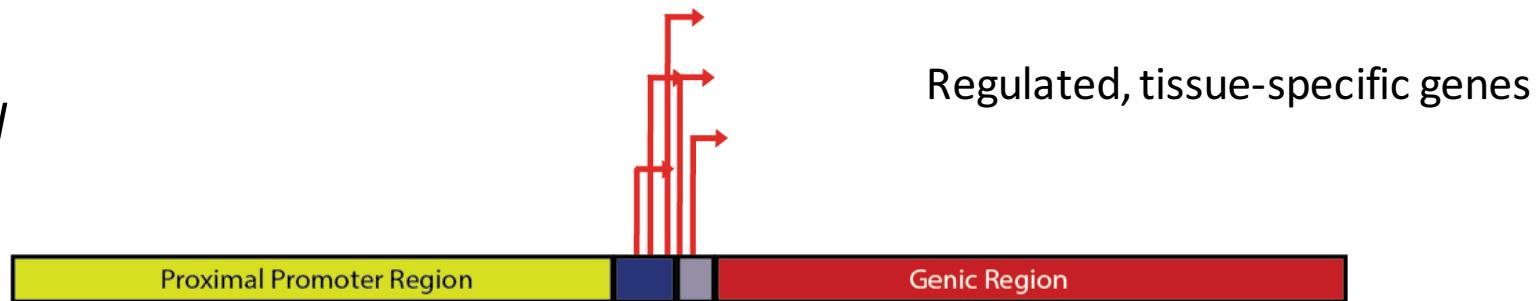
Sample Name	Number of Mapped Reads	Putative TSRs (unique)	Consensus TSRs
Asexual females	50,195,790	11,496 (316)	—
Sexual females	40,636,067	11,289 (231)	—
Adult males	29,373,220	11,558 (557)	—
Total	120,215,077	12,662	10,665
			2,332,582 CTSSs

Majorities of CAGE tags and promoters are found in 5' upstream regions

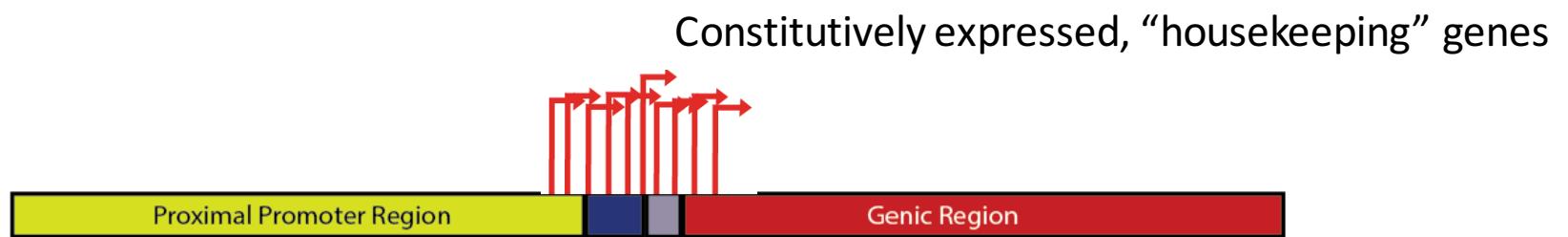


Shape of TSSs at a promoter is associated with expression properties

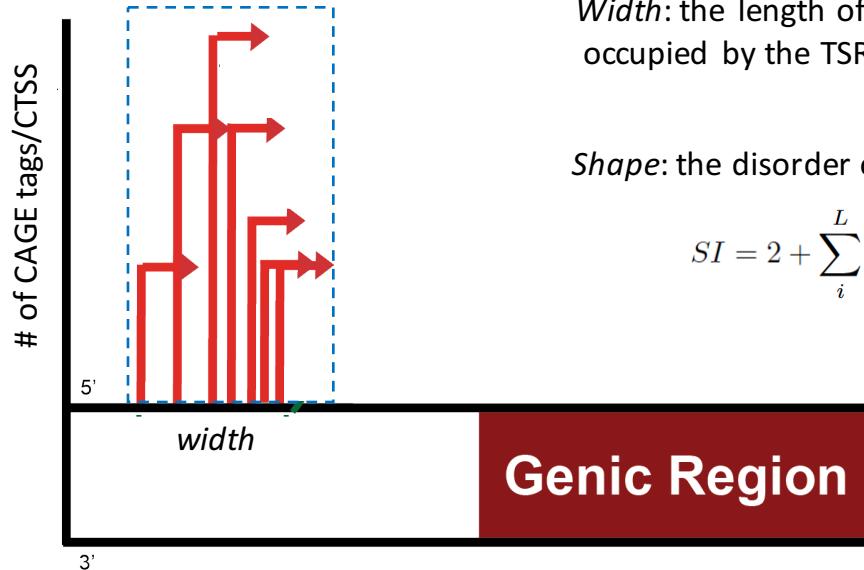
*Peaked*



*Broad*



# Shape classification of CAGE-identified promoters

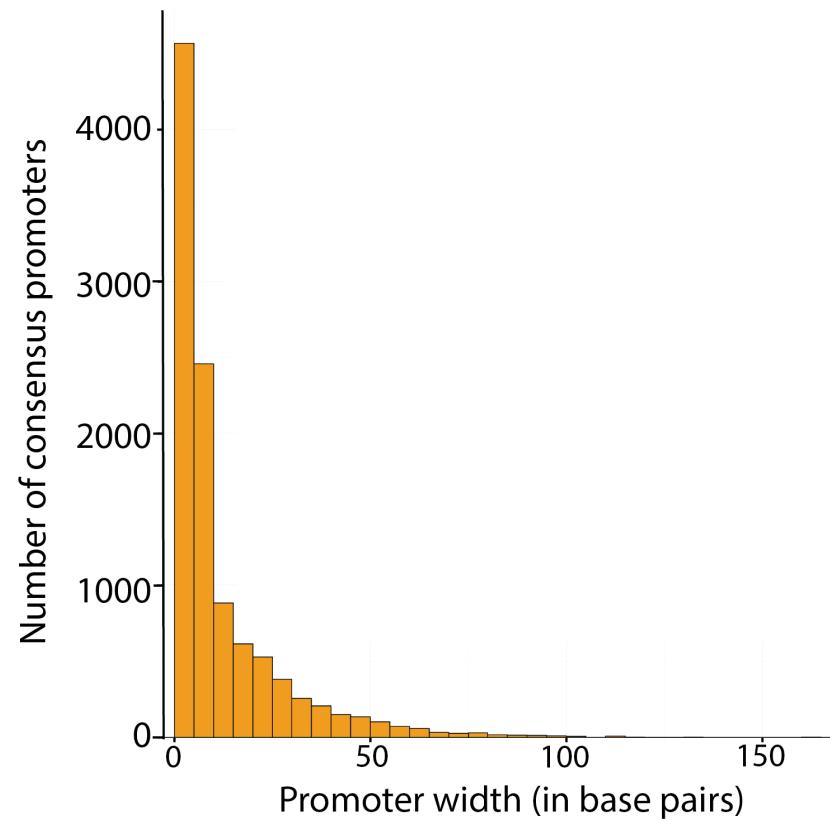


*Width*: the length of the genomic segment occupied by the TSR.

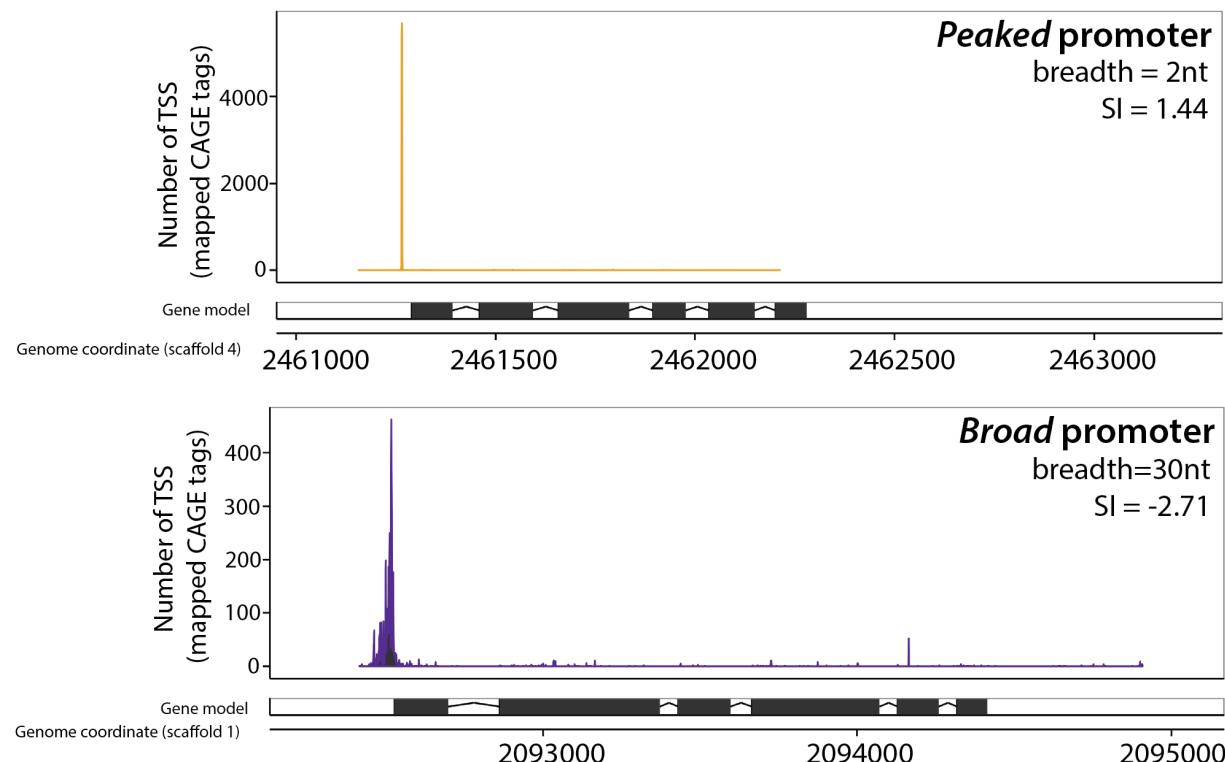
*Shape*: the disorder of TSSs within a TSR.

$$SI = 2 + \sum_i^L p_i \log_2 p_i$$

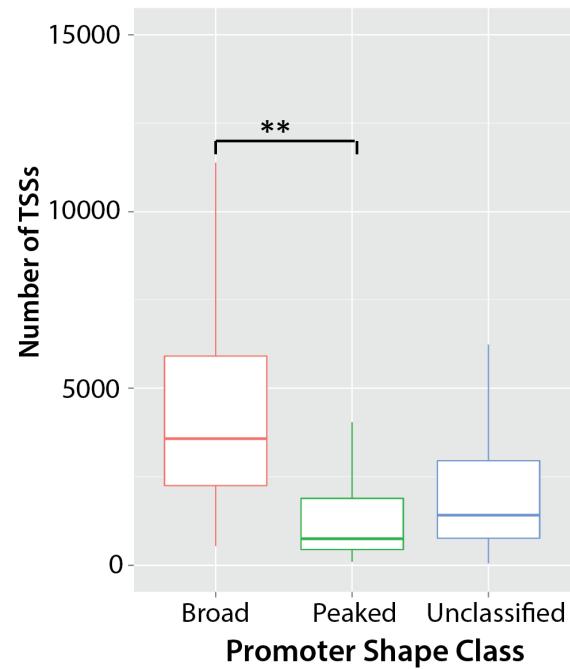
CAGE-identified promoters in *D. pulex* exhibit a range of shape properties



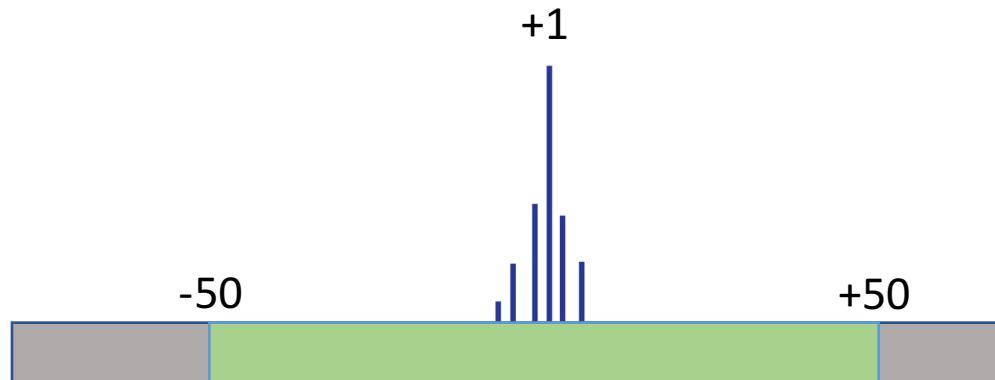
CAGE-identified promoters in *D. pulex* exhibit a range of shape properties



Broad promoters are more highly expressed than peaked promoters



## *De novo* discovery of *D. pulex* core promoter motifs



- 1) Retrieved sequences [-50, +50] surrounding CAGE peaks
- 2) Identified over-represented motifs (HOMER)
- 3) Performed cross validation to statistically verify all identified motifs
- 4) Compared motifs to panel of core promoter elements in *D. melanogaster*.

Heinz *et al.*, 2010 (HOMER)

Raborn *et al.*, 2016. Submitted

With Yao Xiangyu

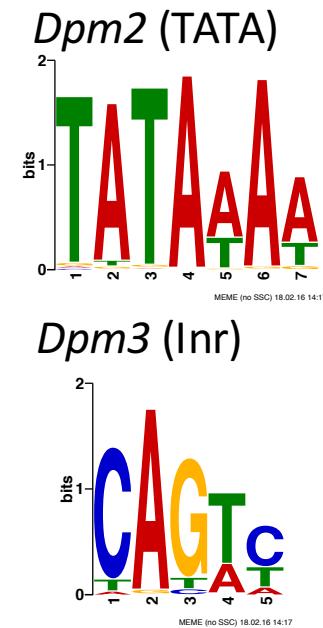
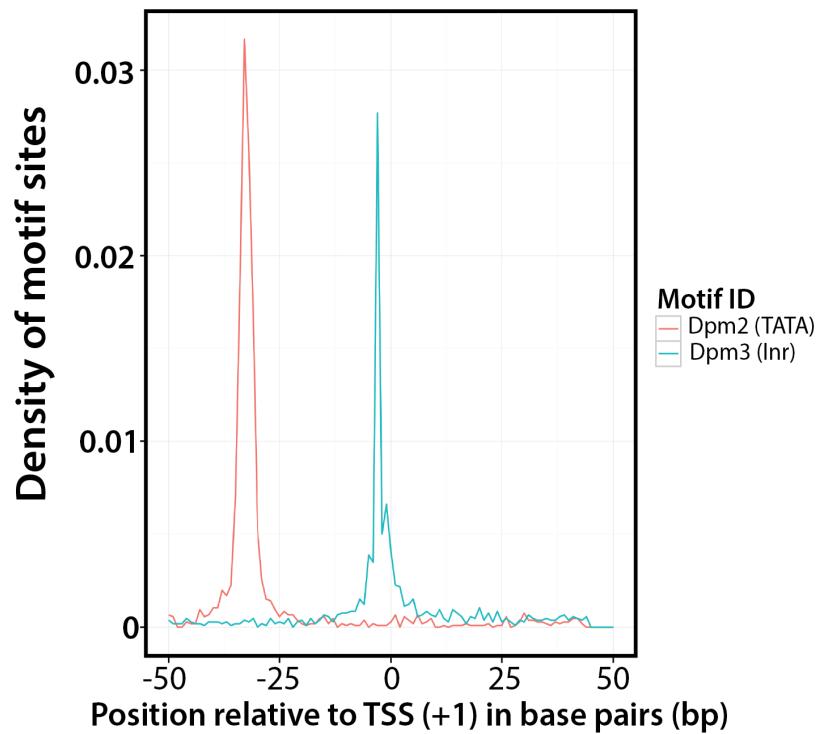
# Validated *D. pulex* core promoter motifs

Motif ID	Motif logo	Occ. (%)	Enrichment	Best match (JASPAR)	E-value	
Dpm1		9.48	1e <sup>-17</sup>	MA0154.1_EBF1	2.67x10 <sup>-5</sup>	
Dpm2		22.62	1e <sup>-308</sup>	MA0108.1_TBP	6.19x10 <sup>-9</sup>	TATA
Dpm3		12.04	1e <sup>-283</sup>	MA0092.1_Hand1_T	2.36x10 <sup>-3</sup>	Initiator
Dpm4		11.95	1e <sup>-84</sup>	MA0139.1_CTCF	4.12x10 <sup>-5</sup>	
Dpm5		15.28	1e <sup>-147</sup>	MA0365.1_RFX1	2.12x10 <sup>-6</sup>	Ohler8
Dpm6		6.86	1e <sup>-45</sup>	MA0266.1_ABF2	5.51x10 <sup>-5</sup>	
Dpm7		4.11	1e <sup>-33</sup>	MA0009.1_T	1.74x10 <sup>-4</sup>	
Dpm8		4.63	1e <sup>-33</sup>	MA0326.1_MAC1	2.38x10 <sup>-4</sup>	

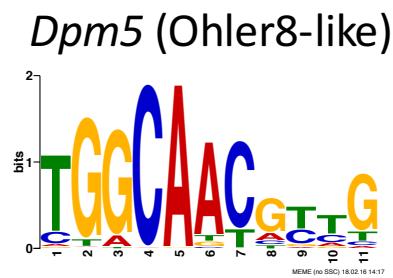
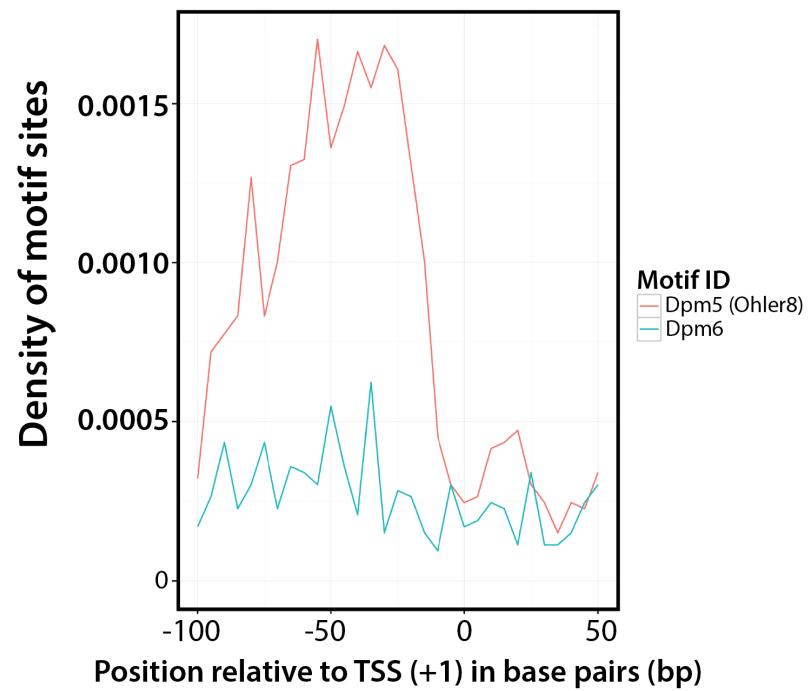
Ohler *et al.* 2002.

Raborn *et al.*, 2016. Submitted

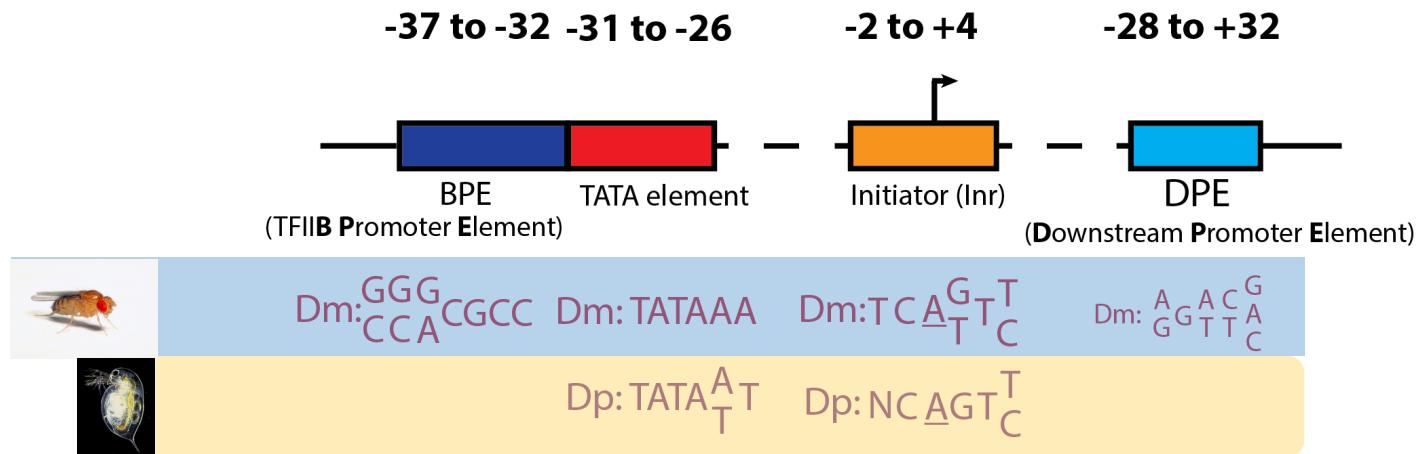
TATA and Inr exhibit expected enrichments relative to +1



Dpm5 (Ohler8-like) is enriched between -40 and -60

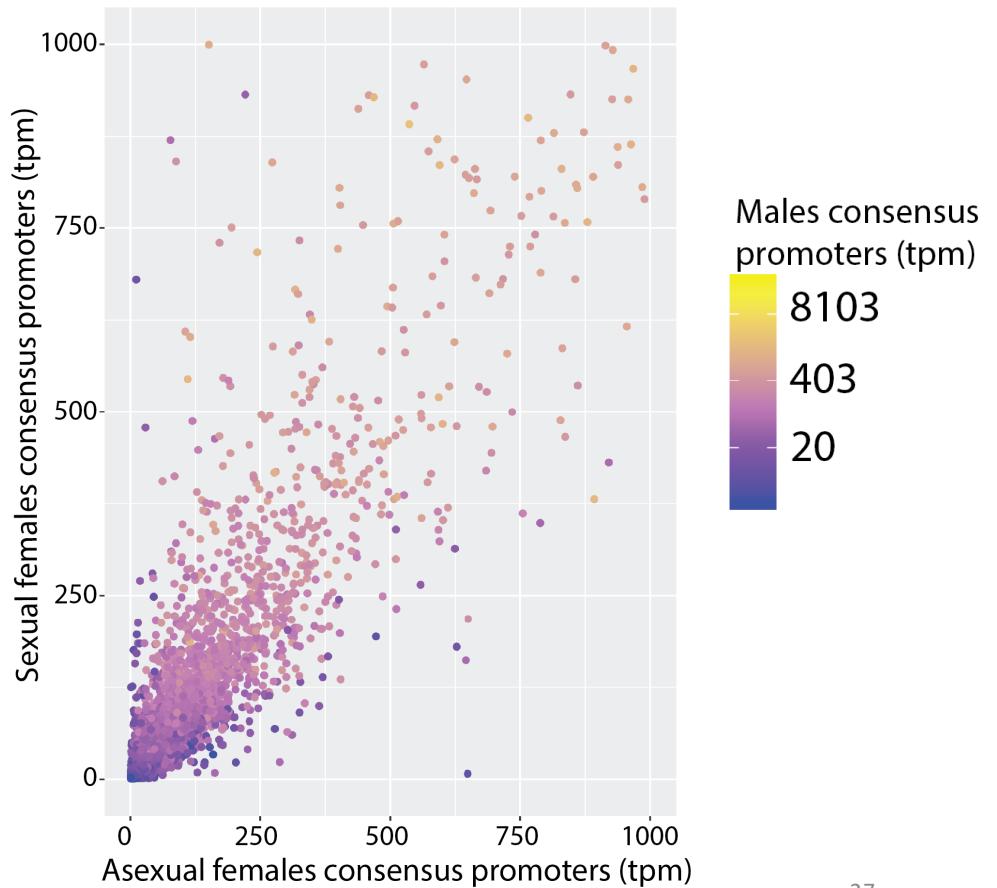
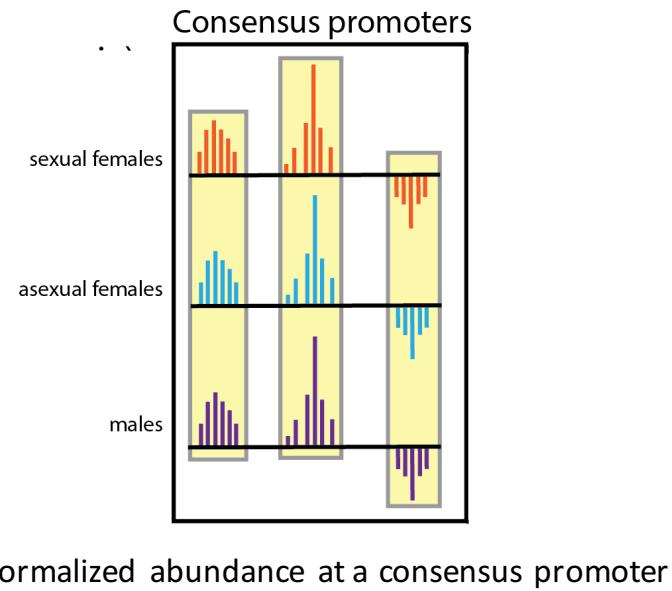


# Current model of *Daphnia* core promoter architecture

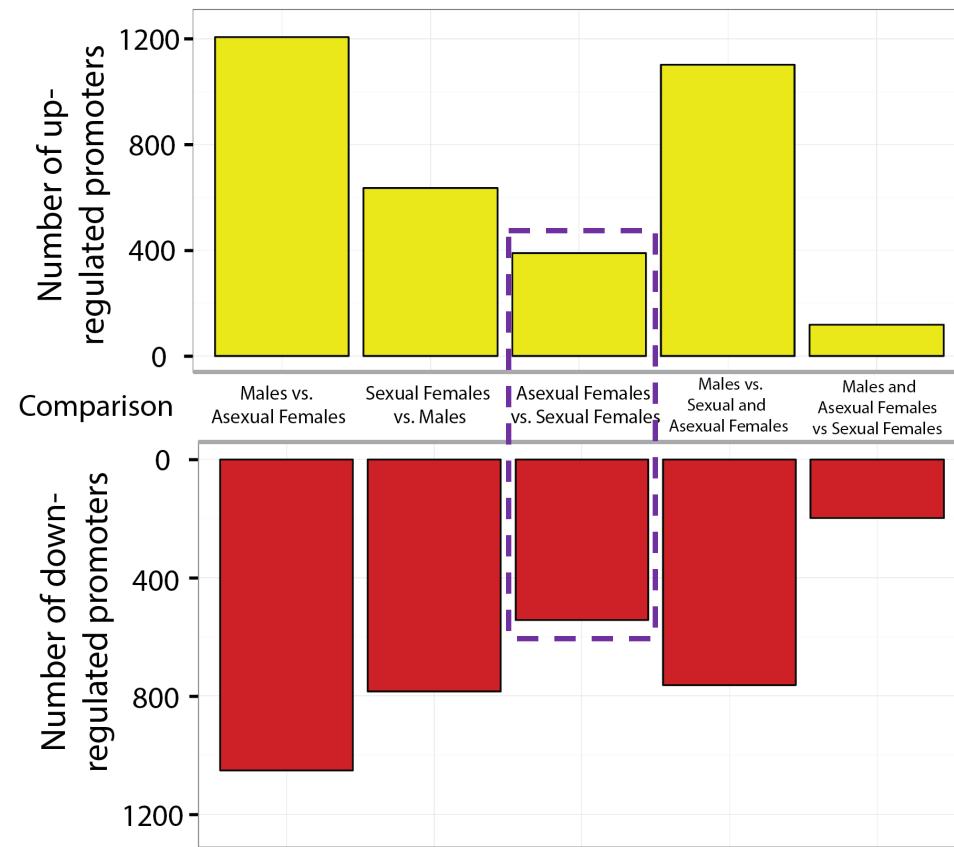


Adapted from Kadonaga, 2012.  
Raborn *et al.*, 2016. Submitted

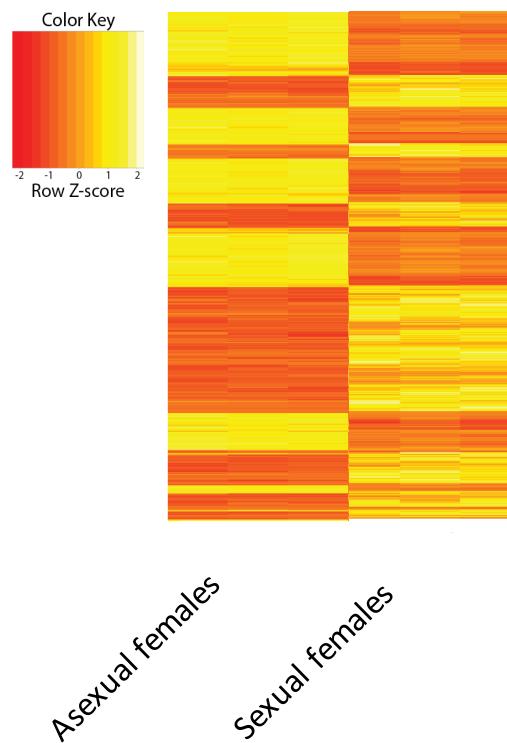
## Differential promoter activity by developmental state



We observe differential promoter activity by developmental state



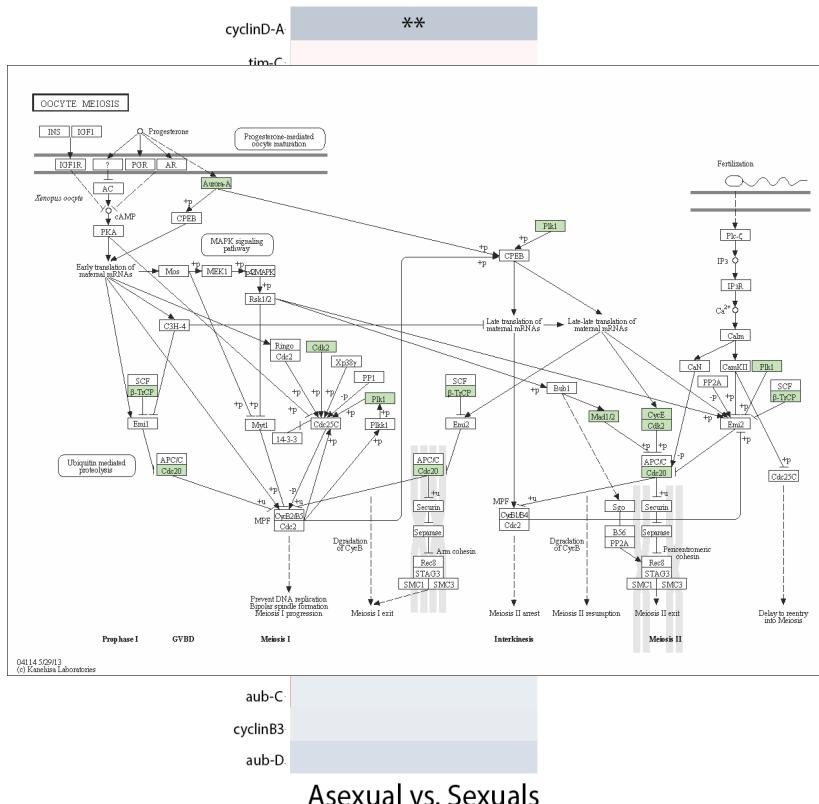
Meiosis and cell cycle genes are associated with differential promoter activity in sexual vs. asexual *Daphnia* females



**KEGG pathways enriched in asexual females (vs. meiotic females)**

Pathway Name (KEGG ID)	Sig. Genes (Expected)	Sig. Genes (Observed)	Odds Ratio	p-value
Cell cycle (04110)	2.99	12	5.04	1.57x10 <sup>-5</sup>
Spliceosome (03040)	2.80	11	4.89	4.37x10 <sup>-5</sup>
Viral carcinogenesis (05203)	3.92	11	3.39	8.29x10 <sup>-4</sup>
Prion diseases (05220)	0.530	4	10.0	1.17x10 <sup>-3</sup>
Alcoholism (05034)	3.9	10	3.06	2.77x10 <sup>-3</sup>
p53 signaling pathway (04115)	1.16	5	5.27	3.80x10 <sup>-3</sup>
RNA transport (03013)	3.67	9	3.06	5.94x10 <sup>-3</sup>
Progesterone-mediated oocyte maturation (04914)	3.14	8	3.02	6.88x10 <sup>-3</sup>
Oocyte meiosis (04114)	3.82	9	2.79	6.88x10 <sup>-3</sup>
Systemic lupus erythematosus (04114)	2.55	7	3.26	7.50x10 <sup>-3</sup>

# Differential CAGE-estimated expression of meiotic genes



We investigated the activity of promoters associated with meiosis genes (Schurko *et al.*, 2009).

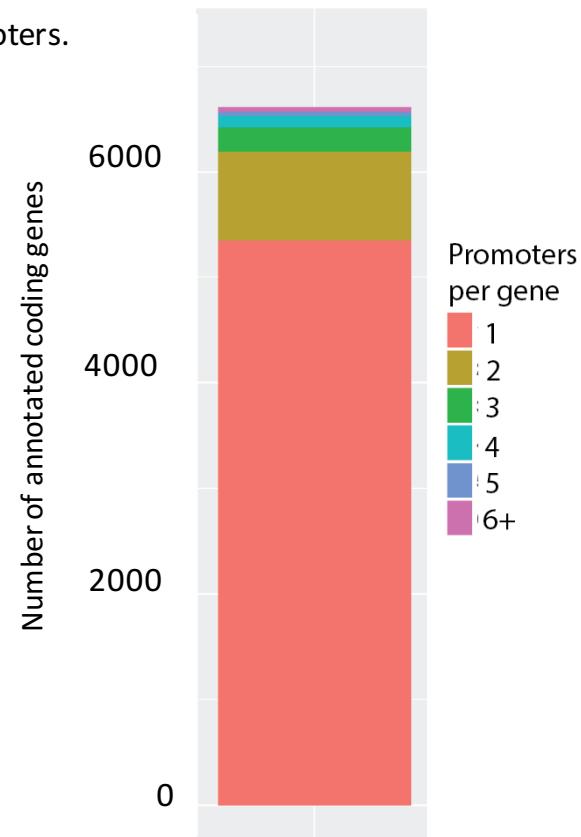
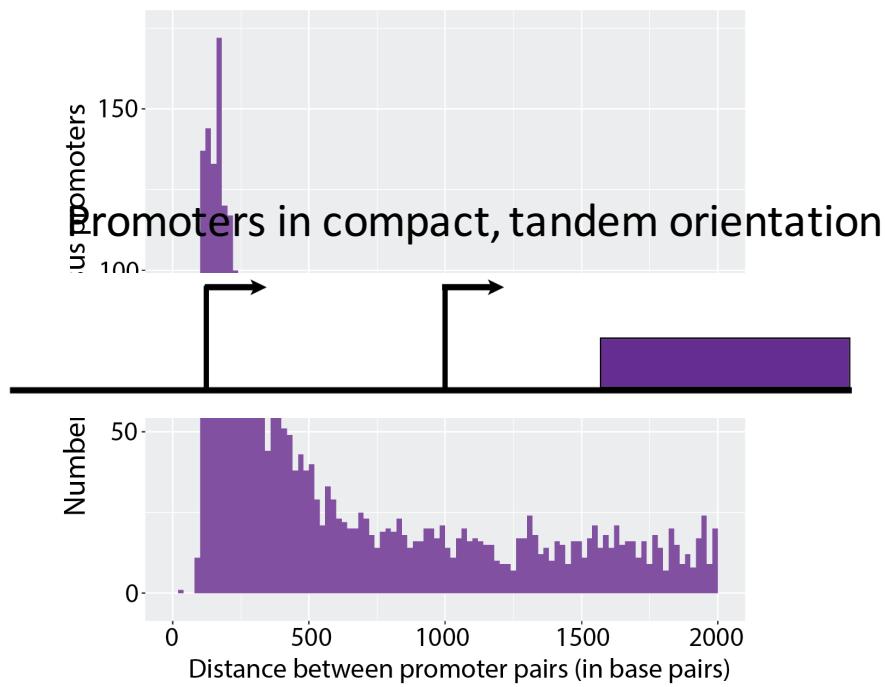
Unexpectedly, we find differential *up-regulation* in the activities of meiosis gene promoters

*cyclinA, cyclinB-D, cyclinD-A, cdk2, cdc20, plk1*

Negative regulation of meiosis I progression.

# *D. pulex* exhibits widespread alternative promoter usage

Nearly 1/5<sup>th</sup> (19.1%) of genes measured have multiple promoters.



## Summary

- We identified more than 12,000 *D. pulex* promoters using CAGE
- *D. pulex* promoters exhibit shape properties consistent with other metazoans
- Discovered eight distinct core promoter motifs in *D. pulex*, including TATA and Initiator (Inr)
- Identified patterns of sex-specific gene expression
- Observed widespread promoter usage is widespread in *D. pulex*
- A high proportion of promoters are in head-to-head orientation

# Acknowledgements

## Contributors

**Volker Brendel**

Yao Xiangyu

**Michael Lynch**

**Ken Spitze**

Kimberly Young

## Critical Comments

Peter Cherbas

Nathan Keith

Dave Tang

Sen Xu

## Funding

NIH R01 (M. Lynch) "Population genomics of *Daphnia pulex*"  
(Award 1R01GM101672-01A1)

The Brendel Group is supported by NSF.

Sequencing services: Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by two NIH S10 Instrumentation Grants.



**INDIANA UNIVERSITY  
BLOOMINGTON**



Please check out our pre-print of this work on bioRxiv



New Results

**Promoter architecture and sex-specific gene expression in the microcrustacean  
*Daphnia pulex* revealed by large-scale profiling of 5'-mRNA ends**

R. Taylor Raborn, Ken Spitzke, Volker P Brendel, Michael Lynch  
doi: <http://dx.doi.org/10.1101/047894>

<http://dx.doi.org/10.1101/047894>

