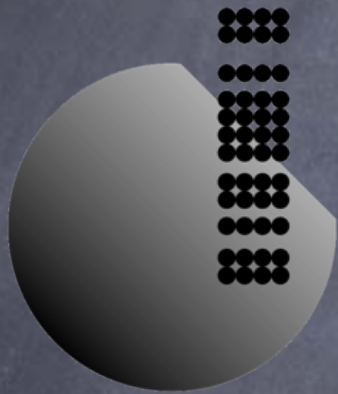


The genome sequence of *Daphnia* for the environmental sciences – Towards biotechnologies that detect chemical impact on populations



THE CENTER FOR
GENOMICS AND
BIOINFORMATICS

The *Daphnia* Genomics Consortium
&
The Joint Genome Institute

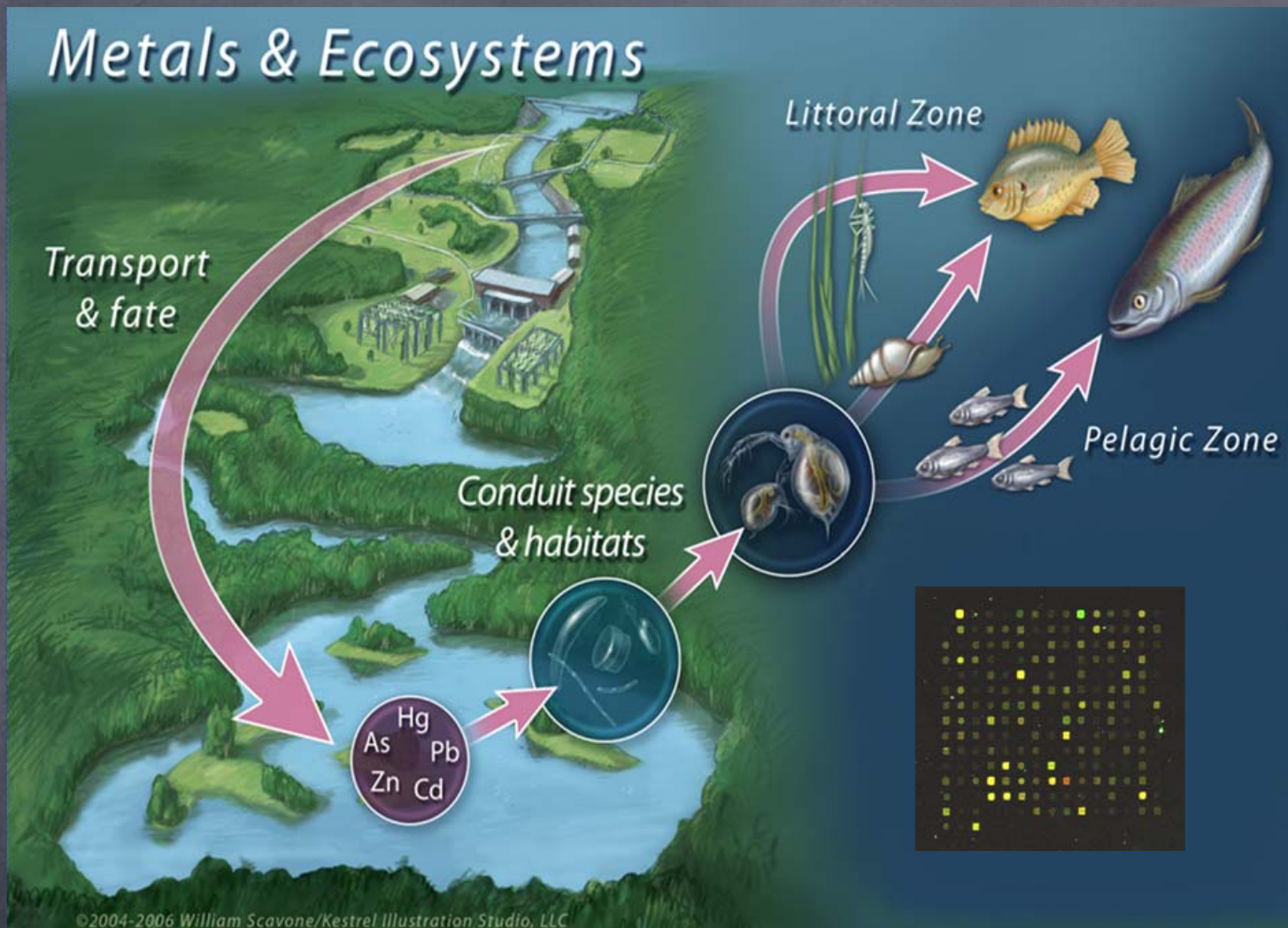


GOAL: To develop the *Daphnia* system to the same level of molecular, cell and developmental understanding as other model species, but with the added advantage of being able to interpret our observations in the context of ecological challenges.

Why *Daphnia*?

- *Daphnia* are a key component of food webs in a variety of aquatic environments.
- Their populations are easily defined by the natural boundaries of lakes and ponds.
- All life stages are easily sampled in large numbers.
- Their reproductive cycle is ideal for experimental genetics with generation times rivaling those of most other eukaryotic model systems (5-10 days).
- Reproduction by cyclical parthenogenesis is both clonal and sexual.
- They show remarkable ability and flexibility to cope with environmental stress.
- They are regarded as the premier invertebrate model for ecotoxicological research.
- Their life history includes a dormant stage that can withstand environmental stress for 100-200 years, allowing the past products of evolution to be resuscitated and evaluated against their modern descendants.

Ecotoxicological genomics and *Daphnia*



What is the molecular basis of genetic adaptation to chronic chemical plus ecological stress?

Can genomics be used to assess population susceptibility to environmental contaminants?

Progress report

- Pre-release of the draft genome sequence, assembly and gene annotations happened July 7, 2007.
- The *Daphnia* genome is tiny; its gene structures are miniaturized, but its gene inventory is huge.
- Important lineage specific gene duplications reflect *Daphnia's* unique biology and possibly adaptations to living life in water.
- When compared to genes for development and reproduction, ecologically relevant genes are less likely to share sequence conservation with those from other genomic model species.
- With the *Daphnia* genomics toolbox now *very* well equipped, rapid progress is made to uncover functions for new genes and what genes interact with the environment.
- The *Daphnia* populations in Sudbury lakes are used for an initial proof that the transcriptome (all the genetic precursors for proteins) can diagnose populations under stress.

Daphnia genome structure

	<i>Daphnia</i>	Fruitfly	Mouse	Nematode
Genome size	200 Mbp	180 Mbp	3,454 Mbp	100 Mbp
# of genes	31,000+	17,200	20,630	26,750
Gene span*	2,000/1,200	2,000/1,200	8,200/3,800	2,900/2,000
Exons/gene	5	2.5	6	6
Exon size	280 bp	580 bp	550 bp	220 bp
Intron size*	280/85	480/110	1250/870	300/175

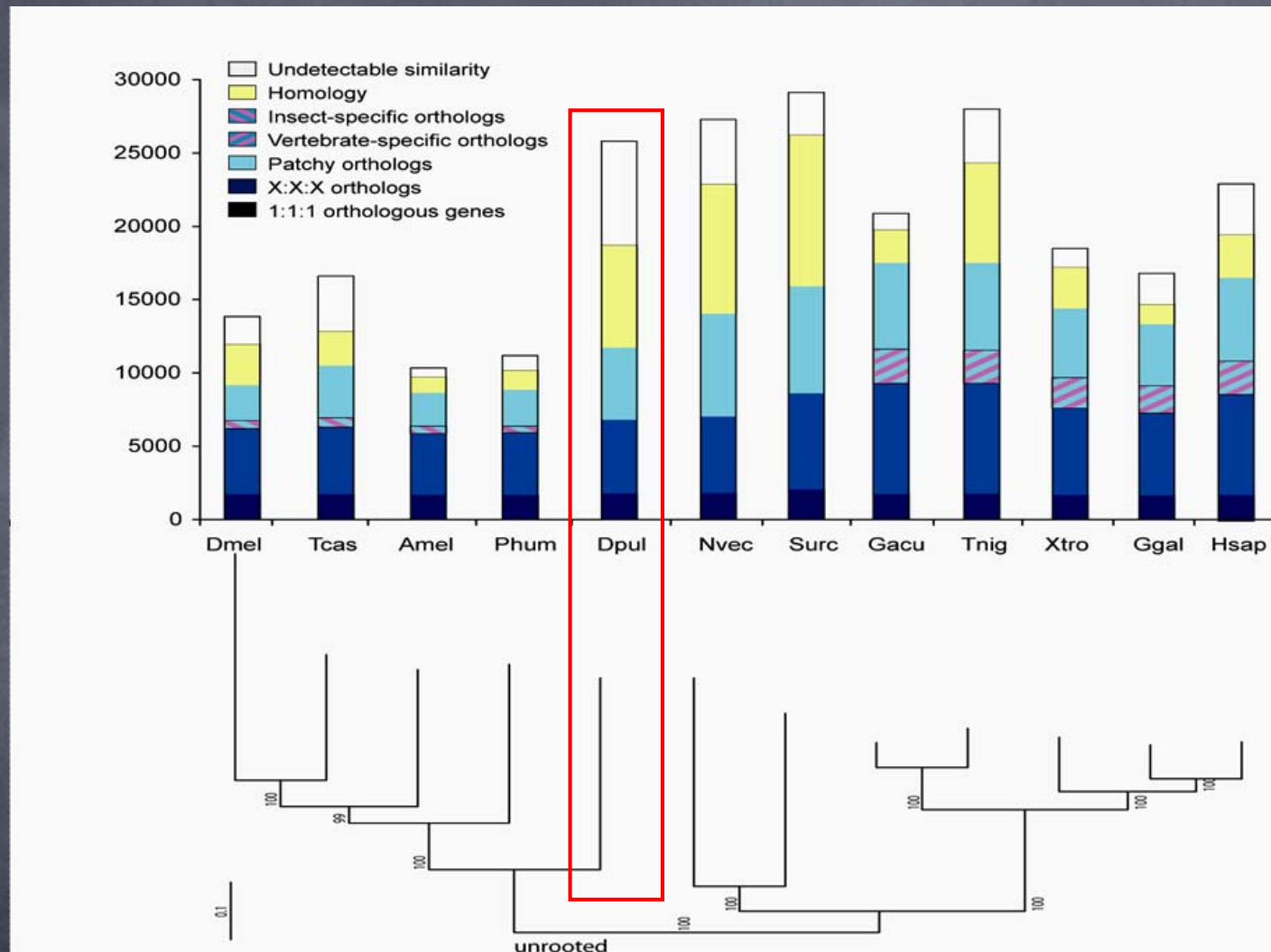
* Mean/Median in bp

Genome sequence annotation

Predicted genes	v.1.1
Total models	30,907
EST support	10,149 (33%)
Protein alignments (nr)	15,990 (52%)
Identified functional protein domains (Pfam)	13,375 (43%)
Experimental support (so far on Tiling arrays)	24,500 (79%)
Complete (5'M-3'*)	27,377 (89%)

Among a set of 716 highly conserved eukaryotic orthologs, *Daphnia* is missing only 26 – about the same as the better annotated genomes

Similarity to other genomes



□ highest fraction of genes without homology (invented genes)

Evgeny Zdobnov, EMBL

□ highest fraction of patchy homologous genes among Arthropoda because of EXTENSIVE gene duplications

Large fraction of repeated genes

	Total # repeated genes	Total # tandem repeated genes	Total # gene clusters
<i>Daphnia</i>	13,972 / 28,093 50%	5,400 / 27,000 20%	919
Fruitfly	4,497 / 13,391 34%	1,500 / 13,500 11%	168
Nematode	8,674 / 19,692 44%	3,000 / 20,000 15%	680

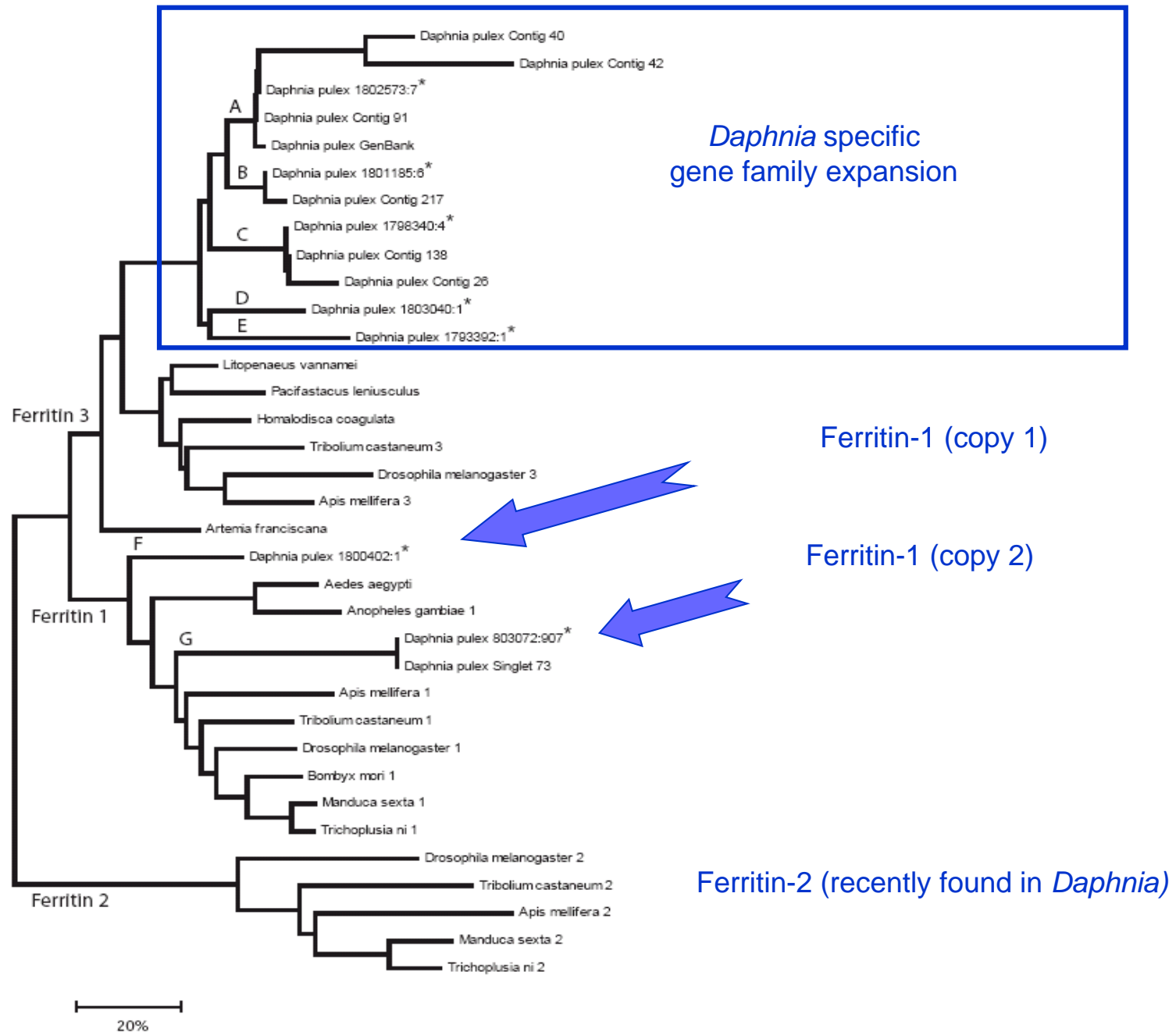


Hemoglobin 30kb gene cluster

D.pulex HB cluster



Many 'ecological' genes have expanded



Tools for experimental annotation of gene function

Sequenced cDNA libraries – Transcripts of genes expressed by *Daphnia* at different life-history stages exposed to 25 separate ecological stressors

Version 3 transcriptional microarray – 10,000 genes represented by 70 bp fragments mechanically printed on glass

Genome tiling path arrays – 4.2 Million overlapping 70 bp fragments representing all the sequenced genome chemically synthesized on glass

* Version 4 transcriptional microarray – 12 experiments performed at a time on a single glass slide, interrogating all genes represented by 135,000 probes

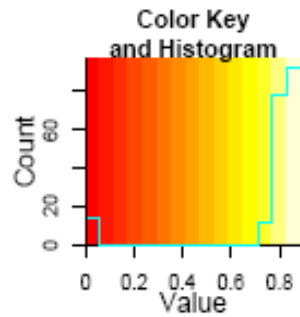
* Environmental screening tool (January, 2008)

Daphnia cDNA resource

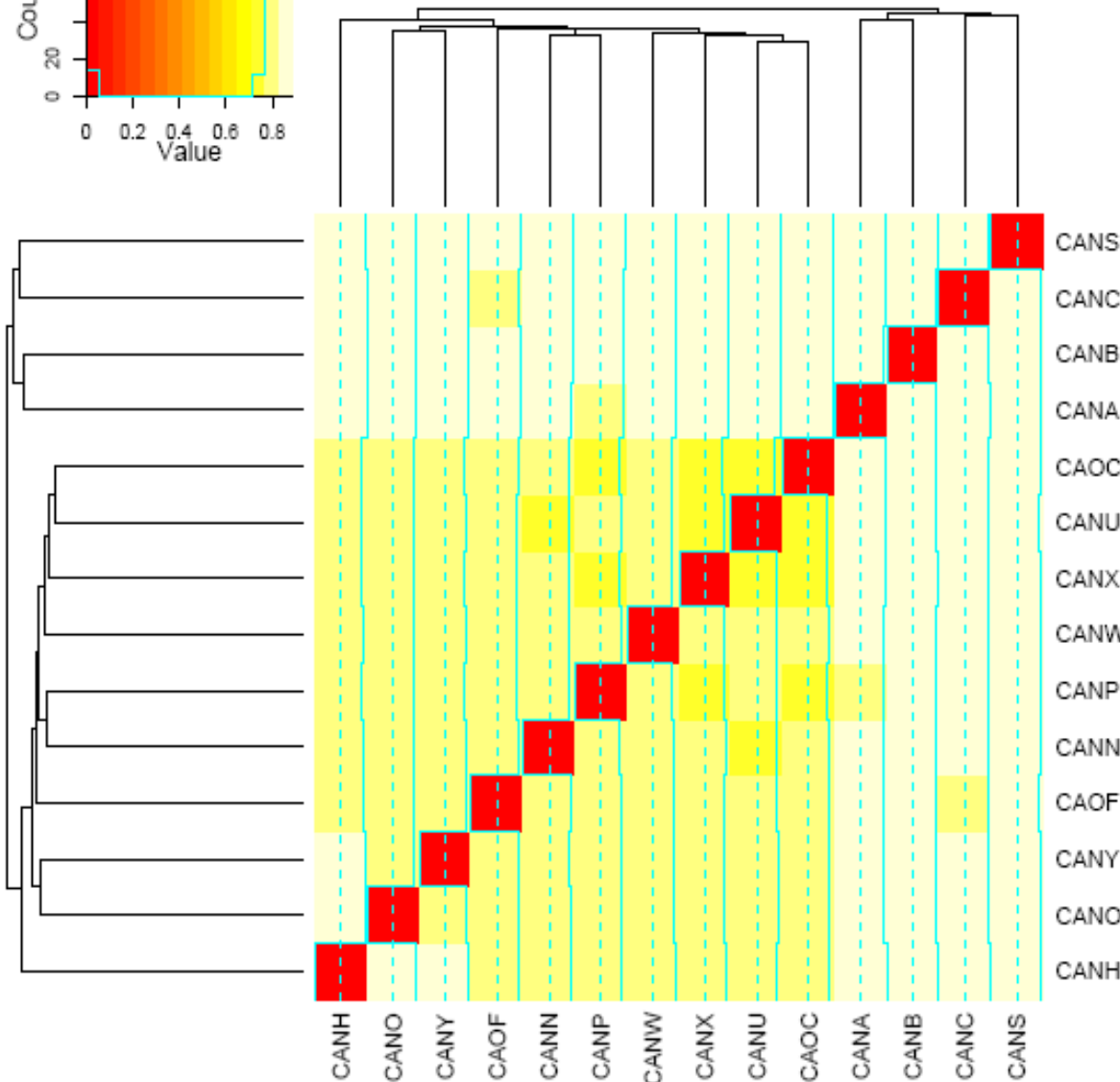
Condition	# Sequenced	# Nuclear	# Clusters	# Clones	% Clone
Non-normalized cDNA libraries	Clones	ESTs			Diversity
4 Unchallenged, mixed	3,568	3,789	1,683	2,732	65
2 Hypoxia	6,144	8,922	2,563	4,856	53
2 UV exposure	2,688	3,070	1,201	1,676	74
4 Metals (Cd, Ar, Zn, mixed)	15,744	24,420	6,474	13,048	50
3 Predator kairomone	8,832	11,717	3,759	6,398	63
2 Food Starvation	4,608	5,510	1,784	2,669	67
2 Microcystis fed	768	675	314	383	82
2 Methyl Farnesoate hormone	4,224	5,246	1,493	2,831	63
TOTAL	46,576	63,349		34,593	

Condition	# Sequenced	# Nuclear	# Clusters	# Clones	% Clone
Normalized cDNA libraries	Clones	ESTs			Diversity
1 Unchallenged, mixed	5,376	8,962	3,413	4,762	72
2 Females	3,840	5,524	2,350	2,942	90
1 Males	4,224	5,425	2,168	2,883	83
5 Metals (Ni, Cu, low mix, high mix, MMA)	19,968	29,415	13,073	16,090	88
1 Acid stress	3,840	6,626	2,870	3,514	86
1 Salinity stress	3,840	6,121	2,645	3,275	85
2 Nanoparticle	8,448	11,385	4,918	6,081	85
1 Bacterial infection	3,456	5,639	2,553	2,935	90
1 Microcystis fed	3,072	4,734	2,052	2,522	86
1 Calcium starvation	3,840	5,309	2,278	2,887	84
TOTAL	59,904	89,140		47,891	

cDNA sampling from non-normalized libraries



Non-Normalized



- Mixed stage, UV
- Juvenile, Control
- Juvenile, Starvation
- Adult, Starvation
- Mixed stage, Mixed metals**
- Mixed stage, Arsenic**
- Mixed stage, Zinc**
- Mixed stage, Cadmium**
- Juvenile, Hypoxia
- Juvenile, Fish
- Adult, Chaoborus
- Adult male, Methyl farnesoate
- Adult, Hypoxia
- Mixed stage, Normalized control

Metal libraries

51 genes found up-regulated in metal vs. non-metal libraries

GeneID	Description	#Metal (10,465)	#Non-Metal (12,262)
237348	Metallothionein	28	3
219769	Metal ion binding; copper, zinc superoxide dismutase (SOD)	76	48
213312	Zn-binding, cysteine-rich protein	6	0
231915	Lipid binding, transporter activity	49	26
219778	Lipid binding, transporter activity	31	11
204011	Metallopeptidase (zinc ion binding)	11	2
222819	Zeelin	5	0
239785	Insect cuticle protein	16	1
219065	Insect cuticle protein	9	0
231476	14-3-3 protein	7	1
224629	Putative defense protein	10	2
230574	Protein with unknown function	8	1
258289	Hypothetical protein	6	0

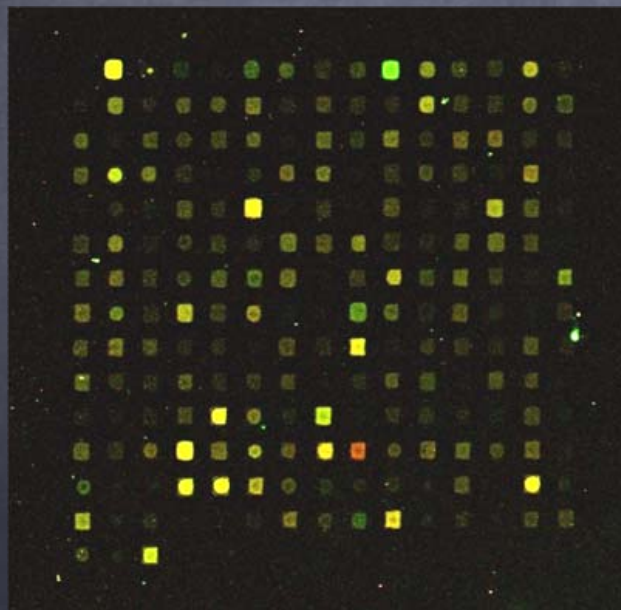
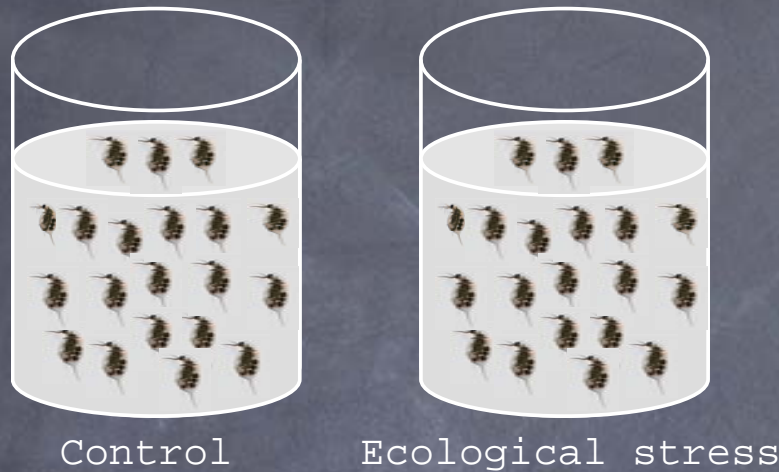
Novel genes are overabundant for biotic stresses

Model Organism Homology

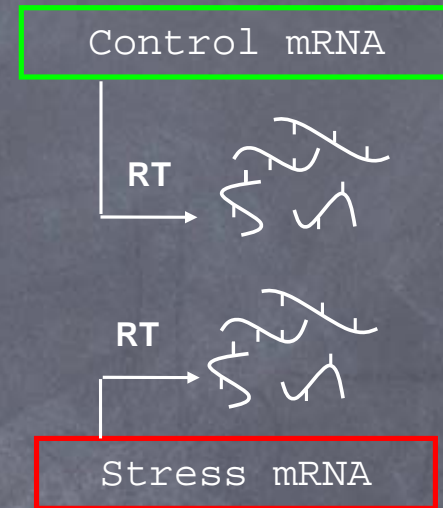
	Any	None	Significance
Biotic	14%	19%	***
Inorganic	21%	23%	ns
Metal	23%	24%	ns
Normal	42%	34%	***

*** $\chi^2 = 172$, d.f. = 3, $p = 4e^{-37}$

Microarray experiments to classify gene responses to stress



Ver. 3 array - 10,000 elements
Representing unique genes



Control labeled target



Stressed labeled target



Microarray experiments to classify genes as ecologically relevant from transcriptional profiles

Biotic Stress

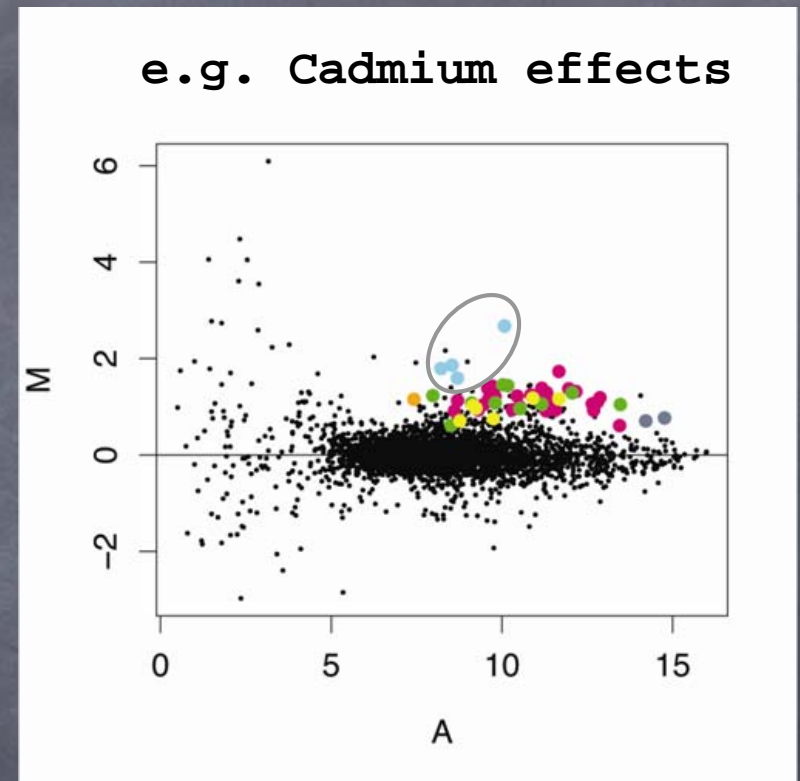
- Exposure to *Chaoborus*
- Exposure to fish under light and dark conditions
- Exposure to parasites
- Exposure to Juvenile Hormone
- Starvation of Phosphorus in algal food source

Abiotic stress

- Exposure to pH 6.0
- Exposure to limited calcium
- Exposure to extreme temperatures
- Exposure to high salinity
- Exposure to UV
- Exposure to low oxygen
- Toxic exposure to nanoparticles

Metal stress

- Toxic exposure to cadmium
- Toxic exposure to arsenic
- Toxic exposure to zinc
- Toxic exposure to copper
- Toxic exposure to nickel



- Metallothioneins
- Cuticle proteins
- Hemoglobins
- Chitinases
- Opsins
- Glutathione-S-transferase

Counts of the regulated genes thus far

Biotic Stress	Up-regulated	Down-regulated
<i>Chaoborous</i> exposure	89	106
Fish exposure light	42	38
Fish exposure dark	1	1
Abiotic stress		
Calcium starvation	5	125
Temp 10 vs 20 Deg	105	159
UV treatment	552	620
Hypoxia	903	698
Development		
Males versus females	49	124

Summary

- *Daphnia's* genome is notably more compact compared to other sequenced taxa; within a genome comparable in size to flies, it contains twice the number of genes. The expanded gene families are largely responsible for its unusual nature.
- The expanded number of genes is partly due to an overabundance of tandem duplicates (TD). The majority of recognizable TD genes are linked to *Daphnia's* response to stressors. Functional investigations using microarrays are validating this finding.
- Even though this initial genome annotation seems reliable, over 35% of the predicted genes have no sequence similarity to proteomes of other model species. This class of orphan genes is overrepresented in cDNA libraries and microarray studies of *Daphnia* responding to ecological challenges.
- These early observations from cDNA sequencing and microarrays highlight the extreme environment condition-dependency of gene regulation. Work is underway to assess the potential in using gene regulation for diagnosing environmental stress responses of natural populations.

Acknowledgements

The Center for Genomics and
Bioinformatics, Indiana University

Don Gilbert, Qunfeng Dong, Justen
Andrews, Brian Eads, Jeong-Hyeon Choi,
Zuojian Tang, Rupali Patwardhan, Zach
Smith, Jacqueline Lopez, Zhao Lai

Joint Genome Institute, U.S. Department
of Energy

Jeff Boore, Igor Grigoriev, Harris
Shapiro, Andrea Aerts, Erika Lindquist

Center for Environmental Health
Sciences, Dartmouth College

Joseph Shaw, Jennifer Davey, Joshua
Hamilton

Hubbard Center for Genome Studies,
Univ. of New Hampshire

Darren Bauer, Kelley Thomas

Utah State University

Michael Pfrender

NimbleGen Systems

John Manak, Roland Green, Courtney
Skalitzky

Others - *Daphnia* Genomics Consortium

Eugene Koonin, Alexander Souvorov
(NCBI), Peer Bork (EMBL), Evgeny
Zdobnov (Univ. Geneva), Shin-ichi
Tokishita, Toshihiro Ohta, Hideo
Yamagata (Tokyo Univ. Phar. Life Sci.)