Daphnia as an emerging model for toxicological genomics

Joseph R. Shaw[1,*], Michael E. Pfrender[2], Brian D. Eads[3], Rebecca Klaper[4], Amanda Callaghan[5], Richard M. Sibly[5], Isabelle Colson[6], Bastiaan Jansen[7], Donald Gilbert[3] and John K. Colbourne[8]

[1]*The School of Public and Environmental Affairs, Bloomington, IN 47405, USA*
[2]*Department of Biology, Utah State University, 5305 Old Main Hill Road, Logan, UT 84322, USA*
[3]*Department of Biology, Indiana University, Bloomington, IN 47405, USA*
[4]*Great Lakes WATER Institute, University of Wisconsin-Milwaukee, 600 East Greenfield Ave, Milwaukee, Wisconsin 53204, USA*
[5]*School of Biological Sciences, University of Reading, PO Box 68, Reading RG6 6AJ, UK*
[6]*Zoologisches Institut, Universität Basel, Biozentrum/Pharmazentrum, Klingelbergstrasse 50, 4056 Basel, Switzerland*
[7]*Laboratory of Aquatic Ecology, Katholieke Universiteit Leuven, Ch. De Beriostraat 32, B-3000, Leuven, Belgium*
[8]*The Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN 47405, USA*

Abstract. *Daphnia* are already an established model species in toxicology. This freshwater crustacean is used commonly for environmental monitoring of pollutants around the globe and plays an important role in establishing regulatory criteria by government agencies (*e.g.*, US EPA, Environment Canada organization for Economic Cooperation and Development, Environment Agency of Japan). Consequently, daphniids represent 8% of all experimental data for aquatic animals within the toxicological databases (Denslow *et al.*, 2007). As such, their incorporation within the new field of toxicological genomics is limited only by the advancement of genomic resources. Because the development of these technologies requires the input and feedback of a large research community that extends far beyond the boundaries of any one discipline, the *Daphnia* Genomics Consortium (DGC) was formed in 2001 to: (i) provide the organizational framework to coordinate efforts at developing the *Daphnia* genomic toolbox; (ii) facilitate collaborative research and (iii) develop bioinformatics strategies for organizing the rapidly growing database. This chapter reviews the progress in establishing *Daphnia* as model species for genomic studies, with emphasis on toxicological applications. As the goals of the DGC are defined largely by extending the boundaries of current biological research in light of genomic information, this chapter first reviews *Daphnia*'s unique biological attributes that make it ideal for such an expansion of research efforts. These attributes include a long tradition of ecological, evolutionary and toxicological study, culminating in the benefits provided by emerging genomic tools.

*Corresponding author: Tel.: +812-855-1392. Fax: +812-855-7802.
E-mail: joeshaw@indiana.edu (J.R. Shaw).

ADVANCES IN EXPERIMENTAL BIOLOGY
VOLUME 02 ISSN 1872-2423
DOI: 10.1016/S1872-2423(08)00005-7

Keywords: BioMart; BLAST; cDNA; *Ceriodaphnia dubia*; *Daphnia* Genomic Consortium; *Daphnia magna*; *Daphnia pulex*; daphniid; ecology; environment; evolution; expressed sequence tag (EST); gene; gene expression; gene-expression profiling; gene inventory; Gene Ontology (GO); Generic Model Organism Database (GMOD); genetic map; genetics; genome; genomics; microarray; model system; physical map; quantitative trait loci (QTL) analysis; toxicity test; toxicogenomics; toxicology; toxicological genomics; wFleaBase.

The role *Daphnia* plays in toxicology

Biological research using daphniids

Species of the freshwater crustacean genus *Daphnia* have been the focus of steady research by naturalists and experimental biologist for centuries (Korovchinsky, 1997). Swammerdam (1669, 1758) provided their common name, water flea, while their scientific designation, *Daphnia*, was imparted one century later by Mueller (1785). The early studies focused on functional morphology, taxonomic classification and biogeography and on their unusual mode of life. Their life cycle includes parthenogenesis, environmental sex determination and the animal's ability to produce two types of eggs (Lubbock, 1857) – one which can remain dormant for decades. During the 19th century, *Daphnia* and other cladocerans were so well characterized that Richard (1895, 1896) produced 'an historical review' with a bibliography nearing 150 titles.

Reasons for *Daphnia* receiving such early and sustained attention are due in part to their numerical abundance, their role in aquatic food webs and their geographical distribution (Edmondson, 1987). Daphniids are ecologically important (Carpenter *et al*., 1987) as they are often the primary grazers of algae, bacteria and protozoans and the primary forage for fish (Tessier *et al*., 2000). They inhabit a remarkable array of environments throughout the world, ranging from permanent lakes to temporary ponds, oligotrophic to eutrophic, hypersaline to freshwater and extending into the UV-rich settings of coastal dune ponds and high-alpine lakes. These radically different waters have been colonized on multiple occasions with a characteristic pattern of convergence of adaptive traits linked to specific habitats (Colbourne *et al*., 1997). This pattern has stimulated much interest into the physiological requirements needed to persist and thrive in these environments. As a result, *Daphnia* are now recognized as a sentinel species of freshwater lakes and ponds, where their decline serves as an indicator of environmental problems (Dodson and Hanazato, 1995). Work is underway to archive the

extensive literature on ecological research using *Daphnia,* which exceeds 4000 articles for the past century and the over 7,000 articles on cladocerans that have been published since 1855 (http://www.cladocera.uoguelph.ca/).

Characteristics of Daphnia *that make it useful for biological research*

Daphnia possess several characteristics that make them valuable for experimental genetic studies. These unique qualities make it possible to translate knowledge about their population structure and ecology to the study of general theories that cross biological scales and disciplines (de Bernardi and Peters, 1987). Several of these attributes revolve around their complex life cycle (Fig. 1). Most *Daphnia* species are cyclical parthenogens, therefore, capable of both clonal and sexual reproduction (Hebert, 1987). Because of clonal reproduction, their genetic background can be held constant, allowing for the maintenance of permanent intact genotypes (Hebert and Ward, 1972; Lynch and Gabriel, 1983). Then, clonal reproduction provides an effective means for comparisons of various treatments against a defined genetic background – a concept that is central to toxicological evaluations and further discussed in the

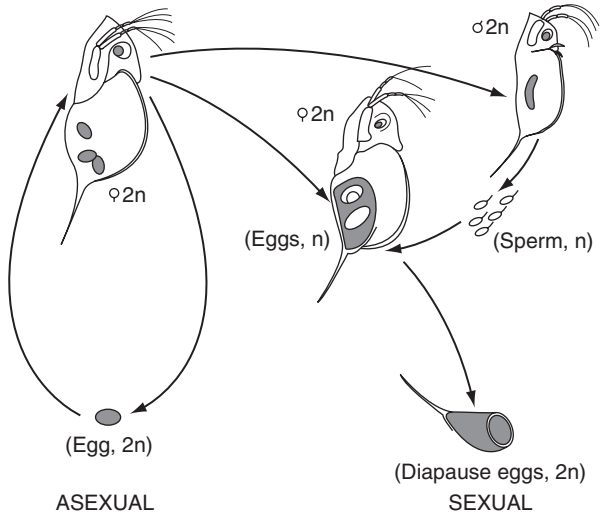

Fig. 1. Life cycle of *Daphnia. Daphnia* are cyclical parthenogens, capable of asexual and sexual reproduction. This interesting reproductive cycle provides a powerful platform for experimental genetics, which is strengthened by growing genomics resources. (From Mort (1991), used with permission from Elsevier Science Publishers LTC).

section titled *Standardization*. For *Daphnia*, sexual reproduction is environmentally induced and these cues can be transmitted in the laboratory (Olmstead and LeBlanc, 2003). During sexual reproduction, females produce sons that are genetically identical to their mothers, allowing for the development of inbred lines by selfing or genetic manipulation by out-crossing. This unusual flexibility in their breeding system makes *Daphnia* an ideal model organism for mapping and characterization of quantitative trait loci (QTL) for complex traits (discussed in section *Genetic maps and QTL analysis*). In addition, development of sexually produced diapausing eggs (*e.g.*, which are encased in ephippia) is paused by a resting stage (diapause). These embryos remain viable for decades and then provide a unique opportunity for long-term maintenance of culture stocks and for probing past populations in the field. This ability to measure evolutionary change by comparing past populations to their modern descendents is made possible by sampling buried diapausing embryos in lake or pond sediment, which also contains a chemical record of the changing environment (Hairston *et al.*, 1999; Pollard *et al.*, 2003).

Other practical attributes that make *Daphnia* a good model system for experimental investigation include the ease by which they are manipulated and maintained in the laboratory (Peters, 1987). Generation times are approximately one week in culture (20°C), which rivals that of most other model eukaryotes and makes it possible to track response throughout their ontogeny. They are maintained in relatively simple defined media (Elendt and Bias, 1990; Kilham *et al.*, 1998; US EPA, 2002) and are fed simple diets that include controlled concentrations of algae and/or bacteria.

Daphnia *in molecular ecology and evolutionary biology*

Many early empirical investigations that employed *Daphnia* were driven by renewed interests in Mendelian genetics, which permitted a modern understanding of evolutionary theory based on natural selection and quantitative theory of population growth (Edmondson, 1987). These early studies attempted to characterize heritable variation in such traits as male production and life history parameters, among others (Banta, 1939). At this time, there was also considerable interest in characterizing the physiochemical conditions associated with daphniid distribution/occurrence (Edmondson, 1987). This problem was studied by chemical ecologists (and toxicologist; see section *Traditional use as toxicity test species*) that attempted to define the chemical limits tolerated by *Daphnia* and by geneticist that treated variability in chemical limits as a heritable

trait (Warren, 1900). It proved difficult, however, to identify the genetic mutations underlying these heritable traits adequately without current molecular tools (Hebert, 1987). As a result, the intersection of evolutionary theory, ecological understanding and genetic experimentation with *Daphnia* was not realized until the advent of molecular techniques in the 1970s. These techniques allowed for rapid and more precise identification of genetic variation in natural populations and resulted in a 'resurgence of *Daphnia* genetics' (De Meester, 1996; Hebert, 1974a, 1974b, 1974c; Lynch, 1983). In fact, the increase in focus attributable to the advent of molecular biology has established *Daphnia* as one of the preeminent models in ecological and evolutionary genetics (Lynch and Spitze, 1994; Mort, 1991; Schierwater *et al.*, 1994). With these techniques, the molecular phylogeny of *Daphnia* has been defined for several geographical units: North America (Colbourne and Hebert, 1996), Europe (Schwenk *et al.*, 2000), South America (Adamowicz *et al.*, 2004) and Australia (Colbourne *et al.*, 2006). This characterization of the animal's ecology, population genetics and phylogenetic history provides the necessary foundation for comparative studies on the evolutionary origin of species groups and of phenotypic diversity in *Daphnia*, including those induced by stress (*i.e.*, predation, UV, thermal, etc.). Cumulatively, this knowledge allows interpretation of molecular modifications, individual phenotypes and population-level responses in context of environmental change and represents a key advantage offered by this emerging genomic model.

Current studies across diverse disciplines

While experimental characterization of *Daphnia* has emerged from the fields of ecology, evolutionary biology and toxicology (discussed in section *Traditional use as toxicity test species*), the enhanced ability to dissect environmental influences on biological responses at multiple levels (*e.g.*, gene, cell organism and population) has proven attractive to other disciplines. Recent studies utilizing *Daphnia* have also focused on physiology (Campbell *et al.*, 2004; Glover and Wood, 2005), developmental regulation (Shiga *et al.*, 2006), innate immunity and host–parasite interactions (Ebert *et al.*, 2004; Jensen *et al.*, 2006; Little and Ebert, 2000; Little *et al.*, 2003), ageing (Dudycha, 2001, 2003; Yampolsky and Galimov, 2005) and epidemiology (Chiavelli *et al.*, 2001; Hall *et al.*, 2006). *Daphnia* are also providing the investigative platform in studies that cross disciplines and address questions fundamental to biology, such as, the causes and consequences of recombination (Nielsen, 2006; Paland and Lynch, 2006). Such expansion has occurred because the biology of

Daphnia offered unique research advantages that will only increase as genomic resources for *Daphnia* grow.

Traditional use as toxicity test species

The unique biological attributes of *Daphnia* that make it a well-suited model system for ecology and evolutionary biology also make it a useful model system for toxicology. The purpose of the following section is to highlight how *Daphnia* have been used in toxicology. In this section we do not provide detailed test methodologies/protocols, but rather illustrate the potential for expanded application of *Daphnia* as a toxicological model with the rapid growth of genomics.

Early studies

Daphnia have a long history of use in toxicological evaluations. Some of the first toxicity studies were conducted by the geneticist, Ernest Warren (1900) who defined the range of sodium chloride tolerated by *Daphnia magna* (Straus). These studies were not only the first to quantitatively establish the link between concentration, duration of exposure and organism response, which eventually was disseminated as Haber's rule (Haber, 1924; Miller et al., 2000), but they also represent one of the first intersections between toxicological principle and ecological understanding. Warren was breeding *Daphnia* to study contemporary theories of heredity, but realized that '[his] results seem to present certain features of considerable interest and of wide biological significance; they illustrate how closely the organism is knit to its external conditions of life.' This statement is one of the first to define aquatic toxicology inquiry, as most of the early studies investigated the physical and chemical conditions that governed daphniid distribution (*e.g.*, pH, Klugh and Miller, 1926; magnesium, Hutchinson, 1932).

The rapid growth in the production of synthetic compounds (*e.g.*, drugs, pesticides, munitions, etc.) during the early 1900s resulted in the rapid evolution of modern toxicology (Eaton and Klaassen, 1996). It was the early growth of the pharmaceutical industry and the advent of food and drug laws during this period that was responsible for promoting the use of *Daphnia* in bioassays (Viehoever, 1931, 1936). Toxicity tests with *Daphnia* (primarily *D. magna*) were being used to define mechanisms of action and for testing the safety and efficacy of drugs. As Arno Viehoever (1937) noted, the concept of '*D. magna* as the biological reagent – has been established in both scientific circles and in the public press (*Time*, 1937).' This assertion resulted in *Daphnia* being coined the

'diminutive drug detective' and receiving praise from Dr Mayo, Rochester, MN (Viehoever, 1937). Contaminant testing with *Daphnia* grew in a similar and parallel fashion, but without regulatory directives, was far less prevalent. It is worth noting that mercury (Breukelman, 1932) and DDT (Anderson, 1944, 1945) garnered some of the focus of these early studies. Ironically, it was environmental tragedies decades later with these very compounds that helped provide the impetus for current environmental regulations. Today *Daphnia* are one of the most used models for environmental toxicological evaluations.

Standardization

The need for defined culture conditions and standard methodologies with respect to *Daphnia* experimental biology was quickly realized, especially in light of the early emphasis on defining the limits of *Daphnia*'s external environments. Likewise, the goals of pharmaceutical and toxicological testing, which were guided by regulatory directives, favoured a reduction of test variables to increase reproducibility and facilitate data comparison (Davis, 1977; Duodoroff and Katz, 1950; Viehoever, 1937). This move towards standardization highlights a fundamental difference between *Daphnia* studies in ecology and evolutionary biology and those in toxicology. The emphasis in toxicology has traditionally focused on defining the media/chemical condition (*e.g.*, toxin, stressor, etc.) as opposed to defining the biological condition. In fact, regulatory compliance testing often employs a single clone distributed among laboratories to limit biological variability, whereas the emphasis in other fields often centres on investigating biological variability among genotypes, populations and species.

Current protocols are standardized with respect to pretest animal maintenance/care, age of test organisms, media, food, duration, ambient light and light–dark cycle, temperature and monitored endpoints. Standardized methods are also specific for species and, as previously mentioned, sometimes for clones. Without question, *D. magna* is the most common species used in toxicology followed by *Daphnia pulex* (Leydig) and *Ceriodaphnia dubia* (Richards; *e.g.*, Table 1 in Shaw *et al.*, 2006). *D. magna* is one of the largest daphniids and, as in many arenas, size has been a major factor in its widespread acceptance (Viehoever, 1937). Some studies, however, have indicated that *D. magna* is more tolerant than other species (Koivisto *et al.*, 1992; Koivisto, 1995; Shaw *et al.*, 2006), which may be problematic in regions where it is not widely distributed (*e.g.*, North America). The use of *C. dubia* is often

Table 1. The utility box for toxicological genomic investigations using *Daphnia*.

	Species	Utility	Notes	References
Arrayed and archived libraries				
Nonnormalized cDNA	*D. pulex*, *D. magna*	CGA, GD, GFR	Standard plus 15 experimental conditions	*Daphnia* Genomics Consortium (2007), Colbourne *et al.* (2007), Watanabe *et al.* (2005)
Normalized, full-length enriched cDNA	*D. pulex*	CGA, GD	Additional 15 experimental conditions	*Daphnia* Genomics Consortium (2007)
Genomic DNA – small inserts (avg. 8 kb)	*D. pulex*	VGS	End sequenced from 5′ & 3′	*Daphnia* Genomics Consortium (2007)
Genomic DNA – large inserts (Fosmids)	*D. pulex*	VGS, VGA	End sequenced from 5′ & 3′	*Daphnia* Genomics Consortium (2007)
Genomic DNA – large insert (Cosmids)	*D. pulicaria*	VGS, VGA	Gene capture by PCR screening	*Daphnia* Genomics Consortium (2007)
Genomic DNA – large insert (BACs)	*D. pulex*	VGA	Fingerprinted by RFLP	*Daphnia* Genomics Consortium (2007)
Molecular markers				
Microsatellites (SSRs)	*D. pulex* complex, *D. longispina* complex, *D. magna*	GPM	Determined from GSS of genomic libraries, ESTs and from genome sequence	Colbourne *et al.* (2004), Fox (2004), Brede *et al.* (2006), Cristescu *et al.* (2006)

Single nucleotide polymorphisms (SNPs)	*D. pulex*	GPM	From genome sequence and ESTs	*Daphnia* Genomics Consortium (2007)
Data				
Genetic map	*D. pulex*	GPM	185 SSRs in 12 linkage groups	Cristescu *et al.* (2006)
Draft genome sequence assembly	*D. pulex*	GPM, VGS, VGA	9080 scaffolds totaling 227 Mb	*Daphnia* Genomics Consortium (2007)
Draft genome sequence annotation	*D. pulex*	CGA, GD	In progress	*Daphnia* Genomics Consortium (2007)
Expressed sequence tags (ESTs)	*D. pulex*, *D. magna*	CGA, GD	Over 200,000 available	Colbourne *et al.* (2007), Watanabe *et al.* (2005), *Daphnia* Genomics Consortium (2007)
Infrastructure and other resources				
wFleaBase	Genus *Daphnia*		Genome database	Colbourne *et al.* (2005)
Microarrays – cDNA	*D. pulex* complex, *D. magna*	GFR	Best for comparing closely related species	Eads *et al.* (2007), Shaw *et al.* (2007), Eads *et al.* (2008), Poynton *et al.* (2007), Soetaert *et al.* (2006, 2007a, 2007b)
Microarrays – oligonucleotide	*D. pulex*, *D. magna*	GFR	Greater specificity and gene coverage	Watanabe *et al.* (2007), *Daphnia* Genomics Consortium (2007)

Table 1. (*Continued*)

	Species	Utility	Notes	References
Toxicological genomic studies	*D. pulex, D. magna*	GD, GFR, VGF	Exposures to cadmium, copper, fenarimol, hydrogen peroxide, ibuprofen, beta-naphthoflavone, pentachlorophenol, propiconazole, zinc	Shaw *et al.* (2007), Poynton *et al.* (2007), Soetaert *et al.* (2006, 2007a, 2007b), Connon *et al.* (2008), Heckmann *et al.* (2006), Soetaert *et al.* (2007b)
Cell lines	*D. pulex*	VGF	Not yet immortalized	Robinson *et al.* (2006)
Transformation lines	*D. pulex*	VGF	In progress	Robinson *et al.* (2006)
Daphnia stocks and mutants	Genus *Daphnia*	VGF	Includes mapping panels	*Daphnia* Genomics Consortium (2007)

Notes: CGA – candidate gene approach for finding known genes of interest; EST – expressed sequence tag; GPM – gene position mapping of predicted loci onto the genome; GD – gene discovery of novel genes; GFR – gene function and regulation predictions; GSS – genome sequence survey; Mb – mega bases; PCR – polymerase chain reaction; RFLP – restriction fragment length polymorphisms; SSR – simple sequence repeat; VGS – validation of gene structure predictions; VGA – validation of local gene arrangements; VGF – validation of gene function predictions.

preferred in lifecycle tests, because it reaches reproductive maturity (~3 days) about three times as quickly as *D. magna* or *D. pulex* (~9 days).

Regulatory tests

To date, standardized procedures have been adopted by numerous environmental protection agencies throughout the world (*i.e.*, American Public Health Association, US Environmental Protection Agency, American Society for Testing and Materials, International Standardization Organization, Environment Canada and Organization for Economic Cooperation and Development, European Commission). More information on standardized test methods is found in Cooney (1995) and Versteeg *et al*. (1997). Information from regulatory tests is used for three primary purposes: (i) criterion development (*e.g.*, establishing regulatory limits); (ii) testing chemical safety and (iii) compliance monitoring. In addition, these data are often deposited in toxicological databases, where they become primary sources for risk assessors. The inclusion of toxicogenomic endpoints has been discussed in context of all these functions (Cook *et al*., 2007).

Advantages in using a model system in ecology and evolution for toxicology

Although the questions addressed by toxicologists, ecologists and evolutionary biologists are similar, there are differences in directives that influence the methods in which they are addressed. The advent of genomic tools, should bridge these (or some of these) differences and expand the uses of *Daphnia* in toxicology. This section discusses the advantages of using *Daphnia* to address toxicological problems, but with approaches drawn from ecology and evolutionary biology.

Cross-disciplinary nature of toxicology

The science of toxicology is inherently multidisciplinary, borrowing and improving on most all of the basic sciences to test its hypotheses. As Micheal Gallo (1996) stated, 'toxicology has drawn its strength and diversity from its proclivity to borrowing.' The need for model species that span disciplinary boundaries is only part of the case for including *Daphnia.* While *Daphnia* were incorporated in toxicological study early in its modern expansion (Warren, 1900), toxicologists traditionally have not exploited the biological attributes of *Daphnia* that other disciplines sought to expand. Rather, they have worked to reduce the contributions of biological variation and focused instead on the role of the chemical surroundings. Perhaps this difference was the result of regulatory

directives or the imprecision noted by Hebert (1987) in correlating genetic change with phenotype prior to the advent of molecular biology. Regardless of the cause for this early split in philosophy, there are substantial reasons to reconsider – at least some – toxicological problems for which *Daphnia* are used in light of modern ecological and evolutionary practices.

Guided by similar questions

Toxicology is focused on understanding response outcomes to pollutant exposure, often with the goal of defining levels that are well tolerated for the organism. Typically, this goal is accomplished by integrating responses at the level of the individual (or below, such as organ tissues or cell lines). Organisms are not passive targets of their external environment, however, and collectively their ranges of response define population-level effects. The magnitude of this impact depends on the organisms' ability to alter their tolerance limits (*i.e.*, acclimate) and over time, these limits re-structure within populations, as the genotypes that favour acclimation are selected (*i.e.*, adaptation). Understanding the dynamic responses that shape tolerance limits are often further complicated, because tolerant phenotypes can be costly to maintain resulting in ecological tradeoffs. Currently, risk assessment techniques only account for the increased fitness associated with acclimation or adaptation, without debiting their associated costs. With this situation in mind, an integration of ecology, population genetics and evolutionary biology practices would provide tools that could improve the derivation of safety limits and ultimately, risk predictions (*i.e.*, susceptibility).

The US National Institute of Environmental Health Sciences has recently suggested such an alignment of philosophies (http://www.niehs.nih.gov/external/plan2006/). This suggestion has highlighted the need for integrated research teams to investigate the complex interplay between genes and the environment in order to identify the environmental factors that influence disease risk. Two areas of research focus towards this endeavour include: (i) an 'expansion of our understanding of environmental influences on genome maintenance/stability …' and (ii) 'concerted efforts to improve our understanding of epigenetic influences on health.' More simply stated, these objectives encompass understanding the environmental influence on the limits and underlying mechanisms of genetic adaptation and physiological acclimation. Then, a growing area of research interest in both environmental toxicology and evolutionary biology is the relationship between genome function and environment. For example, recent analyses of whole genome sequences show that there

are often many more predicted genes than there are genes with known function. One possibility for this observation is that the expression, regulation and function of many genes may be highly context dependent and may only manifest in particular environments. The importance of understanding the genetic basis of interactions between genotype and environment is reflected in a renewed interest in phenotypic plasticity and its relationship to adaptive evolution (Miner *et al*., 2005; Pigliucci, 2005; West-Eberhard, 2003). This shift in focus is also reflected by an increase in research on the relationship between environmental factors and epistatic interactions among genes that may make a substantial contribution to variation in complex traits such as disease susceptibility (Carlborg and Haley, 2004). Increasingly, the connection between genes and phenotypes is being established by combining genomic information with QTL studies. These studies focus on deciphering regulatory networks of polymorphic genes utilizing a combination of QTL analysis and microarray expression profiles (eQTLs) (Bing and Hoeschele, 2005; Carlborg *et al*., 2005; de Koning *et al*., 2005). The combination of these two methods, referred to as genetical genomics (de Koning *et al*., 2005; Jansen and Nap, 2001) is a powerful approach to inferring gene transcriptional relationships (Li *et al*., 2005) and has been utilized to demonstrate that regulation of many genes has a heritable basis (Cheung and Spielman, 2002; Hubner *et al*., 2005; Morley *et al*., 2004). A genetical genomic approach utilized in an organismal system in which environmental conditions can be accurately and systematically manipulated will significantly advance our understanding of the relationship between the phenotype and the underlying genotypic and environmental effects.

As previously highlighted (see section *Biological research using daphniids*), *Daphnia* possess several biological attributes to integrate disciplines and address such research needs and this situation will only be strengthened as genomic resources mature.

Change in response over time

The *Daphnia* system is poised to become a leading research model for understanding environmental influences on gene regulation and subsequent stressor induced acclimation and adaptation. One reason for this emerging utility derives from their life cycle, which during sexual reproduction produces embryos that diapause (*e.g.*, delayed development). Diapausing embryos, which encase in ephippia, are resistant to harsh environmental conditions (*e.g.*, desiccation, freezing) and represent a 'bank of genetic diversity from which an existing population can draw new genotypes (Mort, 1991).' Experimentally, sediments contain banks

of diapaused *Daphnia* that provide access to past populations, as these can be hatched from sediments several decades old (Cáceres, 1998; Hairston *et al*., 1995; Kerfoot *et al*., 1999). Thus, egg banks allow the past products of evolution to be resurrected and evaluated against their current descendants in a controlled setting.

The resting egg bank of *Daphnia* has been used to study the influence of natural stressors (*i.e.*, cyanobacteria, predation) on the distribution of stressor-induced phenotypes in populations and through time (Cousyn *et al*., 2001; Hairston *et al*., 2001). These studies have demonstrated rapid adaptive change (*i.e.*, acquisition and loss of phenotypes) in presence of stress. Current studies have extended these applications to investigate the influence of metal pollution on the structure of *Daphnia* communities (Pollard *et al*., 2003), examining pollution and subsequent recovery. In cases in which the egg banks of interest are no longer viable, the DNA is still accessible to genetic probes for centuries (Limburg and Weider, 2002).

Benefits in applying genomic tools

Genomics add a new level of knowledge to traditional toxicology studies of model test species such as *Daphnia.* Traditional toxicity assays, although informative as to the levels of a chemical that may be toxic to a population, do not provide information on the mechanism by which a chemical has its toxic effect. In addition, many chemicals are found at sublethal levels in the environment, affecting populations by altering the general physiology, reproductive capacity or the ability of an organism to fight disease. Examining changes in gene expression provides a means to identify biochemical pathways that are altered in an organism after even a low-level chemical exposure. Genomics can provide a level of detail that is absent in general toxicity studies, indicating mode of action, differences between low- and high-dose effects, effects caused by exposures to complex mixtures, providing detailed early biomarkers or a 'canary' to indicate exposure or potential effects of an exposure (Klaper and Thomas, 2004). The potential of genomics for studies of exposure and effects in toxicology and environmental risk assessment is now recognized by the US Environmental Protection Agency (Dix *et al*., 2006; Gallagher *et al*., 2006). This potential is illustrated by a pilot project using *Daphnia* in a proof of concept experiment to integrate metabolomic, proteomic and genomic responses to pollutant exposures (http://www.epa.gov/heasd/edrb/comptox.htm).

Identifying mechanisms of action
One of the most promising immediate applications of genomics is in determining the mode of action of a chemical. Gene expression patterns provide clues as to the biochemical pathways that are affected by a particular toxin (Amin *et al.*, 2002). Genomic biomarkers can also distinguish effects of different chemicals. Several studies have now shown that gene expression patterns can be signatures of exposure to a particular chemical (*e.g.*, Bartosiewicz *et al.*, 2001; Merrick and Bruno, 2004). Organisms under stress may show a generalized pattern of gene expression associated with a stress response. Unique gene expression patterns, however, are also present in each of these studies. Chemicals with similar modes of action provide similar expression patterns (*e.g.*, heavy metals, Andrew *et al.*, 2003) even within a chemical class, exposures can be distinguished based upon gene expression patterns (Hamadeh *et al.*, 2002; Poynton *et al.*, 2007; Watanabe *et al.*, 2007).

Improved biomarkers
Traditional field studies for ecological risk assessments try to identify what factors are affecting a population after an insult has already occurred. Molecular biomarkers are taken from laboratory to field studies to diagnose the effects of the many different stressors and organism may be exposed to in its environment. Genomic biomarkers are likely more sensitive and more specific than other biomarkers, which could be highly valuable for field assessments. For example, gene expression patterns are already being used in field studies, to determine when various fish species have been exposed to endocrine disrupting chemicals (Larkin *et al.*, 2002) and studies demonstrate that other chemicals (such as metals) that phenotypically cause the same reproductive issues as endocrine disruptors actually have different mechanisms of action (Klaper *et al.*, 2006; Shaw *et al.*, 2007).

Cross-species interpretations
Another issue with laboratory toxicology studies is the ability to predict the response of one organism using data from another related organism. Will all *Daphnia* species, for example, have the same susceptibilities to a chemical in question? As noted by Shaw *et al.* (2006), there is often a disconnect between species studied in the laboratory and natural populations exposed in the field. Will *Daphnia* toxicity assays correctly predict what will happen to other important aquatic invertebrates? Currently, arbitrary extrapolation factors are employed to provide a conservative estimate of minimum exposure limits to protect the most

sensitive species. From the use of genomic data and biochemical pathway homology across species, direct comparisons can be made of genomic changes related to these biochemical pathways among species. If different pathways are affected when two species are exposed to the same chemical, the data could indicate that a receptor is present or triggered in one species and not in the other or the more resistant species may have additional pathways that are triggered for detoxification. Differences in levels of gene expression may indicate higher chemical sensitivity caused by physiological modification or genetic alteration.

Biosensor/predictive models

Some would argue that genomics for *Daphnia* will never be used for toxicology modelling as standard *Daphnia* toxicity tests are cheap and provide enough relevant information. We argue that because of their ecological importance and distribution, their ease of use, the resources developed through the DGC and most importantly their ability to act as sensitive sensors for other organisms, *Daphnia* present the ideal species group to use for proof-of-principle toxicogenomic modelling efforts. We already see *Daphnia* incorporated into sensor systems for freshwater for human consumption (*e.g.*, de Hoogh *et al*., 2006), in which a change in *Daphnia* mortality levels or behaviour triggers alarms for water intake systems. In the study of de Hoogh *et al*. (2006), *Daphnia* mortality signalled that an unknown chemical had been released into the River Meuse in the Netherlands. Potential chemical suspects were identified and then traditional toxicity assays were used to determine if the suspected chemical was the cause of the *Daphnia* deaths. In this case, in particular, *Daphnia* from the monitoring device could be sampled for RNA expression and compared with a database of gene expression patterns recorded by laboratory screening of thousands of chemicals and mixtures. This direct sampling would provide information to inform what type of chemical was involved and what biochemical pathways are altered to better predict the potential impact on sensitive animal or human populations.

Daphnia *toxicogenomics*

Developing genomics resources in the model *Daphnia* is the most cost-effective and ecologically relevant investment currently proposed for models of toxicology and toxicogenomics. As discussed above, there is significant information known on their behaviour, ecology, population genetics, reproduction and physiology that can now through the efforts of the DGC be linked to genes and gene expression data. The linkage between

phenotype, specifically those related to survival and fitness and genomic biomarkers will ultimately be necessary to make genomic information relevant for toxicology and the environmental sciences. The DGC is providing a means to link genomic data to environmentally relevant phenotypic characteristics that are currently elusive for some model organisms and too expensive to explore in any context but the laboratory for many others.

Daphnia genomics initiative

Community-based approach to Daphnia *genomics*

Proliferation of the consortium approach to big science

Initiating and managing a large-scale genomics initiative and developing the genetic tools and bioinformatic infrastructure to support these efforts are challenging tasks. Perhaps the most important component is fostering a user community with the level of collaboration and shared expertise to take full advantage of developing genomic resources. One method currently employed to accomplish this objective is to build a coordinated consortium with the common goal of advancing a particular model system. The consortium approach works well by ensuring the backing of an organized group with a vested interest in utilizing and managing the vast amount of data from such a project. For instance, most of the completed and ongoing genome sequence projects have employed this approach to secure funding and to achieve success, as demonstrated by the diverse array of taxa and broad utility in terms of both basic and applied issues that have been the focus of genome projects. For large eukaryotic genomes, international cooperative efforts are a critical component of future success.

Rapidly expanding genome sequence data

An overview of current and planned genome projects recorded in NCBI's GenBank, which is by no means exhaustive, documents 338 eukaryotic and 1,471 prokaryotic genome sequence projects either completed or in progress (From NCBI Entrez Genome Project; http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html; accessed 31 October 2007). These data reveal the taxonomic focus of existing genome projects. For example, the vast majority of eukaryotic genome projects are focused on five taxonomic groups; mammalian vertebrates, insects, fungi, protists and plants (Fig. 2). Some interesting historical trends are imbedded in this array of genome

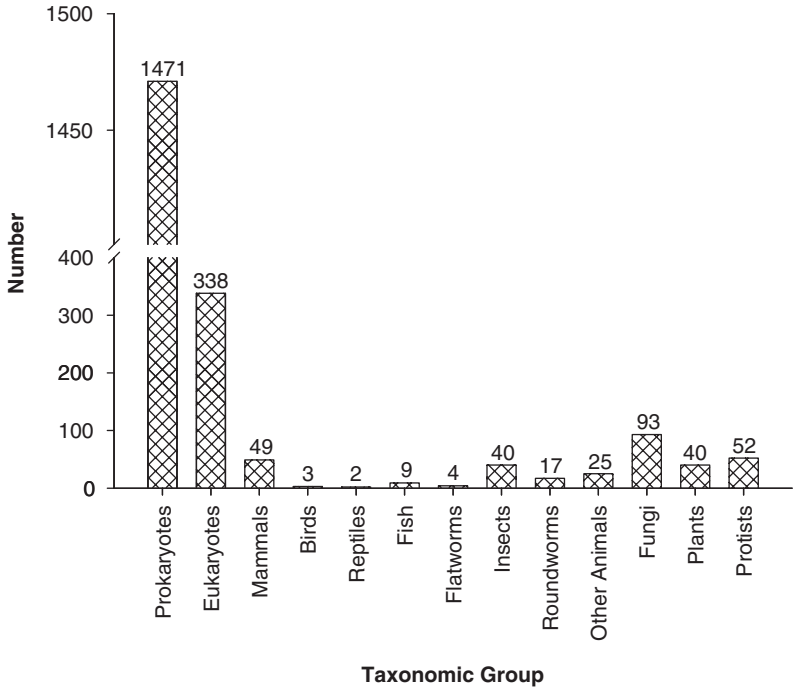

Fig. 2. Taxonomic representation of genome projects. Data compiled from NCBI EntrezGenome Project; http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html; accessed October 31, 2007.

projects. The initial projects were driven primarily by two factors: (i) genome size (smaller genomes are simpler and cheaper to sequence than larger ones) and (ii) utility as a model system for basic genetic or human health research. *Chlamydomonas*, *Caenorhabditis*, *Drosophila* and humans were among the first whole-genome sequences completed for these reasons. Subsequently the focus turned to a broader set of criteria including consideration of economic and phylogenetic utility in informing the selection of genomes. An example of the former is rice (*Oryza*) and of the latter is the tunicate (*Ciona*). Recently, the repertoire of genome projects has included organisms that are important ecological model systems such as the ever so useful plant *Arabidopsis* and models for toxicology and environmental monitoring including the zebrafish (*Danio*) and the freshwater microcrustacean *Daphnia.* These examples highlight the changing perspectives and prioritization contributing to the current diversity of completed and ongoing genome projects and the increasing availability of genomic data for toxicological application.

Why a community consortium?

There are a number of advantages to developing a community consortium. Foremost is the increased availability of data and techniques to researchers in the science community at large. Consortia function in a number of ways to facilitate the development and utility of a model system. One primary function is to provide an open-access platform as a repository for genomic reagents and data. Experimental model systems, like *Daphnia*, are ultimately used for study in a large number of biological contexts. These diverse contexts range from developmental and evolutionary biology to toxicology and ecosystem monitoring and assessment. The result will be the generation of multiple complex data tracks. Integrating these complex biological databases requires a common web-based forum that will promote synergistic research and enhance the utility of the system as a whole. For example, the phenomenal success of *Drosophila* as a genetic model is attributed to the strength of its research community and the community's devotion at creating and maintaining shared resources, ranging from mutant stocks (http://flystocks.bio.indiana.edu/) and cell lines to vectors, clones, microarrays (http://dgrc.cgb.indiana.edu/) and a continuously improved genome sequence annotation and curated research literature (http://flybase.bio.indiana.edu/). A second function of a coordinated consortium is to provide a level of oversight for the community by openly generating quality control standards, including controlled methodologies, nomenclature, test samples and reagents such as DNA libraries and microarray probes. There are numerous examples of the recent adoption of standards for genomic data collection, such as microarray experiments (Ball and Brazma, 2006) and toxicological studies (Mattingly *et al*., 2004). For toxicological studies in *Daphnia*, there is already a number of experimental and agency protocols that require uniformity to ensure that data collected in separate facilities and studies are comparable. Given the sensitivity of genomic assays (gene expression profiling) to test conditions and to the physiological/developmental state of isolates under investigation, these standards are required if genomic data are to be compared across experiments. A third function of a consortium and perhaps the single most important factor requiring a community of investigators is in leveraging the technical difficulties and high cost associated with creating genomics tools for a new species.

Outcomes of a consortium approach

The outcomes of a well-developed genome consortium manifest in a number of ways. A community-based approach facilitates inter-institutional and international collaboration. This approach facilitates

synergistic interdisciplinary research by promoting a common set of tools and sets a forum for the free exchange of ideas and results. Because consortia draw on the combined efforts of many individuals, they enable a scale of science that is well outside the realm of the individual investigator. Resources such as a complete genome sequence and expression arrays are costly to develop and the facilities to conduct such projects are limited. The likelihood of generating financial and technical support to conduct these large-scale projects is greatly increased by coordinated efforts and the presence of a well-organized and documented consortium. The result is a much more rapid development of the diverse set of tools required for genomic level science.

Development of the Daphnia *Genome Consortium (DGC)*
Recognizing the need for a well-developed genomic model system for ecological, evolutionary and toxicogenomic studies the *Daphnia* Genome Consortium (DGC) was initiated in the autumn of 2001 and held its first meeting the following year. This consortium aimed to develop a model system that would address these issues:

(1) A need for tractable models to study the genetic basis and evolution of organismal responses to environmental stressors, as individuals and among populations.
(2) A need for appropriate model systems to dissect the interaction between genotype and environment, *i.e.*, phenotypic plasticity.
(3) A need for model aquatic systems with a wide distribution amenable to use as biological indicators for ecosystem monitoring and risk assessment.
(4) A need for model systems to study the genetic plus environmental basis of gene regulation.

The traditional model organisms (*i.e., Escherichia*, *Saccharomyces*, *Arabidopsis*, *Caenorhabditis*, *Drosophila*, *Danio* and *Mus*) were selected for genomics by earlier research groups because of their utility in development, cell biology and genetics. Unfortunately, all of these systems lack significant biological context outside of the laboratory. Thus, despite the deep understanding of the molecular and developmental properties of these species, we know almost nothing about the natural environmental factors that lead to their evolution or govern their responses to environmental stressors. The inaccessibility of one or more life stages of these species in nature does not inspire confidence that this

situation will change in the near future. One of the ultimate goals of biology is to understand how organisms and populations respond to and evolve in variable environments. With this goal in mind, the DGC has been developing a new model system with clear application to toxicogenomics. The microcrustacean *Daphnia*, because of the biological attributes enumerated earlier is an ideal candidate for further genomic development. The ecological diversity for this organism provides a unique opportunity to ask if independent lineages of *Daphnia* evolve to meet environmental challenges in the same way (Pfrender *et al*., 2000).

Until recently, the main limitation of the *Daphnia* system was the lack of well-developed genetic tools, but rapid progress has been made on this front. The first international *Daphnia* Genomics Consortium (DGC) meeting was held at Indiana University in October 2002 (*Daphnia* Genomics Consortium, 2007). Consisting of diverse scientists from 17 countries, the DGC's goal is 'to develop the *Daphnia* system to the same depth of molecular, cell and developmental biological understanding as other model systems, but with the added advantage of being able to interpret observations in the context of natural ecological challenges.' *Daphnia* is now one of the best genomically characterized organisms with a deeply understood ecology. Although *D. magna* is more commonly used for toxicological research and significant tools are in place and will continue to grow for this species, *D. pulex* was first chosen for genomics because of its natural history is pertinent to a greater number of investigators. Its geographic range is vast compared with other narrowly endemic taxa in North America (Hebert and Finston, 1993, 1996, 1997). It is closely allied to a 'complex' of hybridizing species that have adapted to live in a great diversity of habitats within only the last few million years (Colbourne *et al*., 1998). In certain lineages, the sexual phase of reproduction is altogether lost, therefore enabling comparative studies on the consequences of shuffling the genome by recombination (Paland and Lynch, 2006).

In the following sections, we outline some of the significant genomic developments in the *Daphnia* system as a direct result of adopting a consortium philosophy. The centerpiece of this effort is the generation and annotation of a draft genome sequence assembly for *D. pulex* through the combined work of the DGC and the US Department of Energy's Joint Genome Institute. In parallel, we have developed extensive cDNA libraries and sequences, microarray-based gene expression systems, a large number of polymorphic microsatellite markers to facilitate population genetic studies and genetic mapping and an open-access web portal to maintain and distribute these data (Table 1).

In the sections below, we detail our progress in each of these areas and outline progress in developing QTL mapping panels for community use. The consortium approach allowed for the coordination of multiple simultaneous efforts across varied institutions, providing extremely rapid progress in tool building and application of these emerging tools for the fields of toxicological and ecological genomics. The *Daphnia* experience is an exemplar for a coordinated community collaboration that benefits the biological scientific community at large.

The genomic toolbox for Daphnia *– linking genomic resources via the genome sequence*

The discovery and functional analysis of ecologically relevant genes is critical to the goals of toxicological genomics. At one end of a spectrum of analytical approaches, unique patterns of gene expression can be used as indicators of specific environmental stressors. At the other end, these patterns of expression implicate identifiable genes and specific genetic regulatory pathways in the response of organisms to stressors. In either case, understanding the functionality of expressed genes and their place within a network of interacting genes is highly informative. Nevertheless, gaining an understanding of gene function in a novel organism is a challenging task that requires multiple tactics and a suite of phenotypic and genomic tools (Fig. 3).

A central component of our efforts is the recent generation of the complete genome sequence of *D. pulex*. With this tool in hand, we have a complete catalogue of the coding and noncoding components of a *Daphnia* genome. To understand the functional relevance of these components, however, requires tools to link the phenotypic response of organisms with the genome, tools to examine patterns of transcriptional and translational variation and tools to systematically isolate the function of particular genes. In essence, the research community must have a set of methodologies that implicate particular genes in the genetic basis of organismal response and then another set to verify the functional role of these genes.

There is a number of approaches to associate gene function with the catalogue represented by the genome sequence. Establishing a direct connection between phenotypic variation and physical locations with in the genome that influence this variation can be accomplished using a QTL approach. This methodology utilizes a large set of recombinant individuals and a genetic map based on recombination frequencies to make the link. There is also a need to establish the transcriptional and

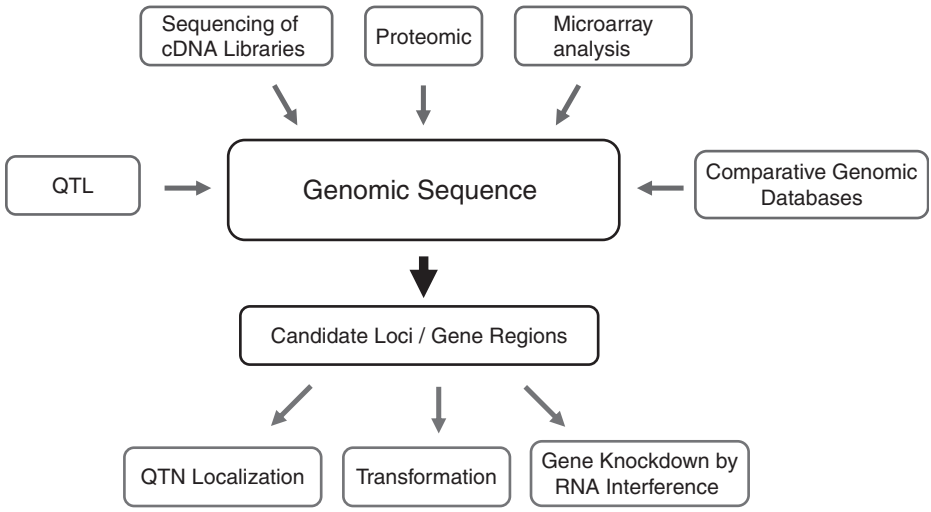

Fig. 3. Strategies for applying the *Daphnia* genomic tool box for functional analysis of toxicologically relevant genes and gene networks. A series of tools (blue) are available to probe the genome (black), via candidate gene, global expression profiling or genetic mapping approaches. These are validated using standard molecular biological approaches (red). (Modified from material presented by Michael Lynch to the Joint Genome Institute on behalf of the *Daphnia* Genomics Consortium, April 28, 2003; See Colour Plate Section in this book).

translational responses of organism to environmental stressors. A first step towards a complete understanding of translation response is to develop a rich collection of cDNA libraries representing as large a fraction of the coding portion of the genome as possible. *Daphnia* is an ideal organism for this task as a common genetic background can be exposed to any number of environmental perturbations and the resulting transcriptional response captured for analysis. Once a library of expressed genes is developed the use of microarray chips to assay gene expression patterns and 2D gels or mass spectrophotometry to assay protein production provides tremendous insight. Finally, a comparative approach can take advantage of the growing understanding of gene function in other phylogenetically related species. This latter approach will be among the most informative in the short run.

Currently the leading model systems for the study of gene function and genome structure in *Daphnia* are a growing number of insect systems. The number of complete genome sequences available for species in this group is rapidly expanding, with over 40 projects completed or in

progress (From NCBI Entrez Genome Project; http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html; accessed 31 October 2007). Functional information derived from these organisms combined with a bioinformatic analysis of the *Daphnia* genome will be a valuable starting point to develop a functional annotation of *Daphnia.* The transfer of information will also flow in the opposite direction. At present there is a conspicuous lack of a relevant outgroup for these numerous insect systems with comparable genomic level infrastructure in the form of a complete genome sequence and available tools for gene expression and QTL mapping studies. Crustaceans, in particular *Daphnia*, are a logical candidate as an outgroup for comparative genomic studies. The close relationship of the crustacea and insects is clearly supported by both molecular and morphological studies (Averof and Akam, 1995a, 1995b; Boore *et al*., 1998; Friedrich and Tautz, 1995; Nardi *et al*., 2003; Regier and Shultz, 1997; Regier *et al*., 2005).

The approaches listed above all serve to increase our understanding of how *Daphnia* respond to environmental stressors and in essence allows for the development of candidate loci and gene regions linked to an organism's response. To increase this understanding to another level, a clear demonstration of the functional relationships among these genes and gene regions, will require the development of techniques to systematically assay genetic variants in natural populations (QTNs) and to knock out genes or transform them into novel genetic backgrounds. In the following sections, we outline the current development of these tools in the *Daphnia* system and suggest priorities for future expansion.

The utility box for toxicological genomic investigations using *Daphnia* is summarized in Table 1.

cDNA sequencing projects

The high-throughput sequencing and analysis of transcribed genes, archived in cDNA libraries, is a powerful method of discovering genes for toxicological studies using *Daphnia.* The straightforward approach of creating cDNA libraries from extracted mRNA within selected tissues, then picking a large number of clones at random for sequencing, was originally proposed to characterize new genes and to facilitate the identification of coding regions in genomic sequences (Adams *et al*., 1991). cDNA sequencing continues to serve as a necessary component of whole-genome sequencing projects, by improving the annotation of the *D. pulex* genome sequence, for instance. The growing number of *Daphnia* cDNA libraries – which are created under a variety of environmental

stressors and life-stages – are also providing insights into the physiological, developmental and cellular responses of animals to toxicants in two ways. First, by clustering the sequences to represent unique transcripts among libraries, classes of genes with shared putative functions are identified. Gene transcripts with similar functions are either enriched or missing from certain libraries compared with others, then producing clues about the general mode of action and biological effects of specific toxicants. Second, both the cDNA clones and their sequences are reagents for the fabrication of microarrays. This tool for monitoring the simultaneous expression patterns of thousands of genes under controlled experimental conditions is valuable, only if the DNA from genes of interest is present on the array to observe the experimental results from cDNA hybridization. Genes of greatest interests to researchers are those whose expression is specific to chemical challenges. These genes, therefore, are most likely to be absent from an array created from a limited diversity of cDNA libraries. This section describes the *Daphnia* cDNA library production and EST sequencing projects that are currently underway to support genomic-level investigations and we summarize some preliminary findings.

Daphnia pulex *cDNA libraries and EST sequencing*

The *D. pulex* cDNA libraries are key genomic tools for the overall DGC efforts at mounting the freshwater crustacean *Daphnia* as a model system for ecological and toxicological genomics. At present, 37 libraries are constructed and partially sequenced from three different isolates of *D. pulex*, *araneta* strain (Table 2). The diversity of conditions and developmental stages represented among these libraries stems from the input of a research community involved in a variety of research programs. The list includes conditions to discover stress response genes, of both natural (*e.g.*, UV, starvation, hypoxia, predation, salt, infection) and anthropogenic causes (metals, acidification, nanoparticles, depleted calcium). Other libraries contrast genes expressed in males compared with females, juveniles compared with adults and at high compared with low doses of environmental toxicants. In many cases, the libraries reveal genes associated with phenotypic plastic responses to changes in the environment, such as the parthenogenetic production of males and haemoglobin synthesis under the control of a juvenoid hormone and hypoxia (Rider *et al.*, 2005) or the modulation of the carapace into defence structures against predators (Tollrian, 1993, 1995). The libraries may also bring to light shared symptoms or physiological outcomes from diverse environmental hardships. In total, over 135,000 cDNA clones are

Table 2. The *Daphnia pulex* cDNA libraries supporting genomic research activities.

Daphnia pulex library ID	Date created	Condition, developmental stage	# Arrayed clones	# Sequenced clones	# Nuclear ESTs	# Clusters[a]	# Clones[a]	% Clone diversity	% Organelle ESTs
Nonnormalized libraries									
Log52	26-Jun-03	Unchallenged, mixed	10,000	1,648	1,414	804	1,435	56	9.0
Log50–1	18-Mar-04	Hypoxia, adult	2,304	2,304	3,355	1,039	1,823	57	10.6
Log50–2	18-Mar-04	Hypoxia, juvenile	3,840	3,840	5,567	1,524	3,033	50	12.7
Log50–3	1-Apr-04	Low dose UV exposure, mixed	2,304	2,304	2,620	1,013	1,433	71	23.6
Log50–4[b]	1-Apr-04	High-dose UV exposure, mixed	2,304	384	450	188	243	77	22.9
Log50–5	5-Mar-04	Unchallenged, juvenile	1,536	1,152	1,580	553	827	67	16.7
Log50–6	19-Jul-04	Low-dose cadmium, mixed	2,688	2,688	4,048	1,209	2,170	56	12.8
Log50–7	19-Jul-04	Low-dose arsenic, mixed	4,224	4,224	6,370	1,867	3,399	55	14.6
Log50–8	19-Jul-04	Low-dose zinc, mixed	4,224	4,224	6,817	1,535	3,709	41	6.5
Log50–9	19-Jul-04	High-dose mixed metals, mixed	4,608	4,608	7,185	1,863	3,770	49	14.4
Log50–10[b]	22-Jun-04	Unchallenged, mixed	3,840	384	390	159	232	69	14.6
Log50–11[b]	18-Dec-03	Unchallenged, mixed	4,992	384	405	167	238	70	30.0
Log50–12	8-Sep-04	Invertebrate (*Chaoborous*) predation, adult	4,608	4,608	6,542	2,034	3,511	58	17.7
Log50–13	8-Sep-04	Food starvation, juvenile	2,304	2,304	2,826	924	1,378	67	18.3
Log50–14	8-Sep-04	Food starvation, adult	2,304	2,304	2,684	860	1,291	67	23.1
Log50–15[b]	8-Sep-04	Microcystis fed, juvenile	2,304	384	307	150	175	86	44.8

Log50–16[b]	8-Sep-04	Microcystis fed, adult	2,304	384	368	164	208	79	39.2
Log50–17	8-Sep-04	Fish predation, juvenile	3,840	3,840	4,750	1,548	2,638	59	22.5
Log50–18[b]	8-Sep-04	Fish predation, adult	2,300	384	425	177	249	71	20.3
Log50–19[b]	8-Sep-04	Methyl farnesoate hormone, juvenile	2,300	384	413	170	227	75	31.7
Log50–20	8-Sep-04	Methyl farnesoate hormone, adult	3,840	3,840	4,833	1,323	2,604	51	11.8
Total			76,808	51,718	72,179				
Normalized libraries									
Log50–21	22-Jun-04	Unchallenged, mixed	3,840	5,376	8,962	3,413	4,762	72	0.7
Chosen One-1[b]	7-Feb-06	Females, juvenile	384	384	211	98	121	96	8.0
Chosen One-2	7-Feb-06	Females, adult	3,456	3,456	5,313	2,252	2,821	84	1.5
Chosen One-3	7-Feb-06	Males, adult	4,224	4,224	5,425	2,168	2,883	83	21.7
Chosen One-4	10-Apr-06	Low-dose nickel, mixed	4,224	4,224	6,484	2,865	3,599	86	0.5
Chosen One-5	10-Apr-06	Low-dose copper, mixed	4,224	4,224	6,852	2,963	3,685	85	1.3
Chosen One-6	10-Apr-06	Acid stress pH 6.0, mixed	3,840	3,840	6,626	2,870	3,514	86	0.5
Chosen One-7	10-Apr-06	High salinity, mixed	3,840	3,840	6,121	2,645	3,275	85	2.6
Chosen One-8	4-Apr-06	Fullerene nanoparticle, mixed	4,224	4,224	5,643	2,428	3,044	84	14.1
Chosen One-9	11-May-06	Bacterial infection, mixed	3,456	3,456	5,639	2,553	2,935	90	1.1
Chosen One-10	2-May-06	High-dose mixed metals, mixed	3,840	3,840	4,398	2,030	2,452	91	3.5

Table 2. (*Continued*)

Daphnia pulex library ID	Date created	Condition, developmental stage	# Arrayed clones	# Sequenced clones	# Nuclear ESTs	# Clusters[a]	# Clones[a]	% Clone diversity	% Organelle ESTs
Chosen One-11	2-May-06	Low-dose mixed metals, mixed	3,456	3,456	5,407	2,447	2,967	89	0.8
Chosen One-12	11-May-06	Low-dose monomethylarsinic III, mixed	4,224	4,224	6,274	2,768	3,387	87	4.5
Chosen One-13	10-May-06	Titanium dioxide nanoparticle, mixed	4,224	4,224	5,742	2,490	3,037	86	3.4
Chosen One-14	10-May-06	Microcystis fed, mixed	3,072	3,072	4,734	2,052	2,522	86	8.1
Chosen One-15	10-May-06	Calcium starvation, mixed	3,840	3,840	5,309	2,278	2,887	84	11.1
Total			58,368	59,904	89,140				

Notes: cDNA clones were sequenced from both ends, except for library Log52 that was sequenced only from the 5′ end. Clone diversity is calculated by dividing the # of clusters (including clusters of 1 EST) by the # of clones. This estimate is inflated, especially for nonnormalized libraries, by ignoring clones containing organelle transcripts (6–45% of ESTs are mitochondrial, depending on library). By contrast, the normalized libraries typically contain between <1% and 10% organelle ESTs.

[a]These numbers are of clusters and clones of nuclear genes only.

[b]Libraries failing stringent quality control checks and were, therefore, excluded from high-throughput EST sequencing.

archived within 384 well plates, providing over 161,000 ESTs of nuclear genes. Mining of these sequence data, with reference to the biological conditions of the animals when the gene transcripts were sampled, uncovers regulatory genetic pathways specific to how *Daphnia* cope with environmental challenges (study in progress).

Isolate Log52 is the source of the first cDNA library, created to produce experimental *D. pulex* microarrays. This work is also a pilot study that aimed to improve protocols for fabricating subsequent libraries, which are enriched for full-length cDNA and optimized for gene discovery. A detailed characterization of this initial library assured high-quality cDNA resources for *Daphnia* (Colbourne *et al.*, 2007). Of 1,648 sequenced clones, only 9% contained mitochondrial genes. The average molecular weight of cDNA inserts within the large size fraction was 847 bp, while 64–68% of the cDNAs were full-length or close to full-length. With few exceptions, this level of quality was met and often exceeded in the 20 nonnormalized Log50 libraries – whose average insert sizes range between 575 and 819 bp – and in the 15 normalized 'Chosen One' libraries, with average insert sizes between 819 and 1504 bp (unpublished data).

Alternative splice variants of abundantly transcribed genes are more likely to be detected in standard libraries created without normalization. Moreover, data on the relative number of specific transcripts sampled from libraries that represent an array of experimental conditions may be indicative of differentially expressed loci, which possibly deserve further study (Audic and Claverie, 1997). The benefits in sampling from nonnormalized cDNA libraries, however, are offset by the cost of sequencing redundant clones. Normalization procedures reduce the number of redundant copies of gene transcripts, therefore increasing the gene discovery rate during high-throughput EST sequencing. Sequencing from normalized *D. pulex* libraries resulted in a 15% decrease in the average number of EST contaminants from mitochondrial genes and gained an average of 20% genes discovered (Table 2).

The gene inventory of Daphnia *compared with model insects*

An important use for the large cDNA sequence data is in identifying *Daphnia* genes whose functions may be inferred from their sequence similarity to well-studied loci in genetic model species. This method is the candidate gene approach to uncovering genes of interest for toxicology. Although crustaceans and insects have divergent evolutionary histories for some 600 million years, both *Daphnia* and the model insect *Drosophila* (fruitfly) are members of a monophyletic group called

Pancrustacea (Boore *et al*., 1998). Consequently, these two species, plus related taxa, are expected to share ancestral genes that are central to their biology and development. A fraction of the gene inventory of *Daphnia* is also expected to be uniquely crustacean or specific to daphniids, given the numerous lineage-specific adaptations associated with their distinct ecologies and life histories.

A recent analysis of the Log52 cDNA sequences provides some insight about the level of sequence conservation between *Daphnia* and its fellow arthropods. Of 787 assembled gene sequences, ~68% matched to a similar sequence in at least one insect proteome (Colbourne *et al*., 2007). By also comparing to the nematode genome, 21% of the genes are either derived within *Pancrustacea* or lost within nematodes. These results suggest that the elaborate functional genetic database for the fruitfly (Drysdale *et al*., 2005) may be a valuable resource for making predictions about the biological function for a majority of *Daphnia* genes. Understandably, the link between gene sequence similarity and function can be tenuous, partly because of lineage-specific expansions and extinctions of ancestral loci or the invention of new genes. For example, within the Log52 cDNA sequence dataset, 13 sequences encode genes with putative homologues in insects that specifically bind charged atoms like metals. These include a lineage-specific expansion of the *D. pulex* ferritin genes. In contrast to flies that have three ferritins, *Daphnia* has six or more loci that probably code protein subunits. Expression data from microarray experiments suggest that at least three gene duplicates have modified functions, based on their different transcriptional responses to metals and their sex-biased expression (Colbourne *et al*., 2007). Further experiments are required to determine which locus, if any, has retained the ancestral gene function.

Community resource

The *D. pulex* and *D. magna* EST sequences are mapped to specific clones within the archived cDNA libraries (Table 1), which can, therefore, be retrieved by researchers for their experiments. As the pace of gene discovery quickens with high-throughput and computational methods, so will the demand for reagents to validate predicted functions, through detailed gene-by-gene investigations. At present, however, the cDNA sequences are playing a vital role in the annotation of the newly assembled *D. pulex* genome sequence. They facilitate the delineations of intron/exon boundaries, mark the positions of transcribed and untranslated regions and help to identify regulatory regions of the genome. Their predicted gene translations enable more accurate similarity searched

against protein databases and are more useful for pattern matching and comparison of *Daphnia* functional proteomics data. Moreover, the full set of assembled cDNA sequences were used to design 10,000 oligonucleotides that are unique to single loci in the genome, for printing microarrays to detect the transcriptional signatures of exposures to toxicants.

Gene expression profiling

The use of microarray gene expression profiling for toxicology research has been a key development linking molecular approaches with traditional toxicology studies. Large-scale sequencing efforts have resulted in the creation of several different microarray platforms for toxicologenomics research in *Daphnia* and studies using these arrays are beginning to be published (Connon *et al*., 2008; Poynton *et al*., 2007; Shaw *et al*., 2007; Soetaert *et al*., 2006, 2007a, 2007b; Watanabe *et al*., 2007). Data from these experiments are being used for a variety of purposes, with special interest in the environmental toxicology community for biomonitoring, which has been called 'canary on a chip' (Klaper and Thomas, 2004). Other uses for mRNA abundance data include gene function discovery (Hughes *et al*., 2000; Shaw *et al*., 2007), genetic regulatory network analysis (Tavazoie *et al*., 1999), defining mechanisms of toxicity and clearance (Waring *et al*., 2001a, 2001b) and evaluating the effects of genetic and environmental variation on transcript levels (discussed in section *Toxicogenomics database – a role for wFleaBase*). Harnessing the power of transcriptional profiling will require solutions to a number of vexing problems, including general challenges for gene expression analysis, as well as, those unique to the ecological aspects involved in this work. A brief description of options for expression profiling in *Daphnia* will be followed by a summary of their current uses and a discussion of the advantages and challenges of conducting microarray studies with daphniids.

Microarray platforms for Daphnia

The number of options available for comparative expression profiling on a genome-wide scale continues to increase, from cDNA amplicon arrays or custom oligonucleotide arrays to high-throughput pyrosequencing technologies (*e.g.*, Illumina®, Solexa®; see Cook *et al*., 2007 for review). Newer pyro-technologies have a significant advantage in that transcripts are sequenced directly, so there is no ascertainment bias caused by absence of (potentially unknown) sequences from an array.

While there are ongoing efforts to apply pyrosequencing technologies to *Daphnia*, these approaches are still too costly and in the research and development stage (Darren Bauer, personal communication). By comparison, the ease and comparatively minimal costs of microarrays makes it highly probably that this technology will be a workhorse for many years to come. With this in mind, several groups have developed microarrays for probing *Daphnia*'s expressed genome (*i.e., D. magna*, Connon *et al*., 2008; Poynton *et al*., 2007; Soetaert *et al*., 2006, 2007a, 2007b; Watanabe *et al*., 2007; *D. pulex*, Shaw *et al*., 2007). All of these arrays relied on cDNA libraries for source material, but differed with respect to design.

One microarray format that has been utilized for *Daphnia* research followed the approach described by Gracey *et al*. (2001), which involved arraying unknown PCR-amplified cDNA clones randomly picked from a collection of high-quality cDNA libraries. These blind arrays were used to identify differentially regulated targets that were then sequenced for characterization/annotation. One of the major challenges associated with this approach lies on the analysis end, as there is an unknown and uneven amount of replication on the array. The observation of repetitive annotations, however, suggests that they are not the product of chance events, which can be addressed statistically using permutation tests to estimate the likelihood that random processes would place highly represented annotations on a list of significant annotations. This microarray platform has been used for *D. magna* and *D. pulex* and proved successful as a gene discovery tool differentiating male- and female-specific responses (Eads *et al.*, 2007, 2008) and following exposure to pollutants (*i.e.*, metals, ordinance related compounds; Poynton *et al*., 2007; Shaw *et al*., 2007).

The other array format that has been utilized with *Daphnia* involved spotting known cDNA amplicons. These were derived from: (i) a large EST project that sequenced through a cDNA library from unexposed mixed aged organisms (Watanabe *et al*., 2005); (ii) suppressive subtractive hybridization (SSH) between adults and juveniles (Soetaert *et al*., 2006) and (iii) SSH on populations exposed to selected stressors (*i.e.*, cadmium, lufenuron, pH, hardness, kerosene and ibuprofen; Connon *et al*., 2008). SSH, which identifies genes expressed in one population, but not in the other, was used as a means of establishing condition-specific targets (Diachenko *et al*., 1996; Lisitsyn *et al*., 1993). Theoretically, using stressor-specific SSH to create microarrays maximises the specificity of the response and reduces the need to have all genes represented. Retrospect analysis suggested, however, that the SSH

derived spots actually underperformed random nonspecific cDNA in identifying genes regulated by the conditions from which the SSH probes were generated. This does bring into question the utility of using SSH and, if it is used, suggests that a series of SSH treatments should be used rather than just one. This microarray platform has been used for *D. magna* to identify differentially regulated genes following exposure to fenarimol, propiconazole, ibuprofen and cadmium (Connon *et al.*, 2008; Heckmann *et al.*, 2006; Soetaert *et al.*, 2006; Soetaert *et al.*, 2007a, 2007b).

The declining cost of oligonucleotide synthesis has increased their attractiveness as a microarray platform, largely because of the time and effort required to generate cDNA for spotting. Such an approach has been applied by Watanabe *et al.* (2007) for *D. magna.* This approach, however, requires careful attention be paid to probe design in order to minimize cross-hybridization of related sequences, so splice variants and closely related multigene families can be easily distinguished. Such intricate design is now possible for *D. pulex* because of the complete genome sequence and the diverse EST project (Table 1). A robust set of oligonucleotide (70mers) probes is now available for this species. One challenge by using oligonucleotide arrays is their lower signal in proportion to their length. Such arrays are thus susceptible to signal loss caused by sequence mismatches especially in distantly related study populations. It remains to be examined how broad the oligonucleotide set that is available for *D. pulex* can be applied within the *D. pulex* complex of species.

Benefits of using Daphnia *in microarray studies*

Transcriptional profiling has already begun to provide important insights into the biology of *Daphnia.* First, because the animals can be bred clonally, separation of genetic and random environmental components (*e.g.*, subtle changes in food concentration from beaker to beaker) is possible. In addition, transgenerational effects can be tested or removed because of the quick generation times and large numbers of progeny produced in culture. Investigators can, therefore, examine how the norm of reaction (*i.e.*, the range of phenotypes a given genotype is able to produce, given variable environmental conditions) of gene expression changes from one clonal genotype to another. This approach will have unique power in *Daphnia* for the dissection of the genetic basis of gene expression, sometimes called genetical genomics (Jansen and Nap, 2001). A current problem in the field is how to account for the effect of common descent, sometimes called 'phylogenetic inertia' (Blomberg and Garland, 2002), on gene expression. Recent work in this area using a traditional

phylogenetic comparative method (Whitehead and Crawford, 2006) demonstrates an approach to control for the effects of phylogenetic distance (assuming neutral drift to be the dominant component of population-level differences) and shows some patterns of gene expression to be under natural selection. More work in this area is clearly warranted, because studies at the population-level are an integral part of an ecological approach to toxicology. Another important attribute of *Daphnia* is the ability to bring field samples directly into the lab and rear them in a common garden environment. In some cases, extinct populations can even be 'resurrected' by hatching dormant ephippia from sediment core samples, providing an unparalleled opportunity to examine microevolutionary patterns. Although longitudinal studies (*e.g.*, over time) of natural populations have not yet been reported, with the proliferation of the genomics utility box they will probably be an important component of toxicology research in *Daphnia.* Together, these features provide virtually limitless opportunities to study the interplay of genetics and environmental conditions in toxicological assays.

Many of the challenges facing transcriptional profiling in environmental toxicology are not unique to the field, including a need for appropriate quality control and analytical rigour. There are, however, some issues arising from the use of natural populations and ecological stressors that demand extra attention. Foremost among these are the avoidance or characterization that confounds biological variation, such as cryptic infection by parasites (see Ford and Fernandes, 2005) or regional variation in water chemistry affecting baseline culture conditions. Thorough genetic characterization of study populations is also critical, as is the attention to age, developmental stage or reproductive status of the animals that are being assayed. Finally, integration of expression data with toxicological or other types of data will be an important determinant of the utility of microarray data. In this area, the DGC has made considerable progress in creating databases and internet tools for community use. Developments in this area are highlighted later in this chapter (see section *Toxicogenomics database – a role for wFleaBase*).

Genetic maps and QTL analysis

The availability of genomic information greatly facilitates the mapping and characterization of genes responsible for complex phenotypes. Typically, complex traits are under the control of multiple interacting loci, the segregation thereof leading to a continuous distribution of phenotypes. Such traits are also termed quantitative traits. The genomic

location of quantitative trait loci or QTLs, can be estimated using various methodologies, all of which are based on the cosegregation of genetic markers with known location on a genetic map and the QTLs in question. One of the most powerful methods is line-cross mapping: by crossing two lines with different phenotypes of interest and divergent genotypes at markers loci, one can determine what part of the genome cosegregate with the QTLs (Lynch and Walsh, 1998). This methodology has been previously used in toxicity studies in rodents (Mcclearn *et al*., 1993) and plants (Dong *et al*., 2006). With high-resolution mapping, this method can identify putative candidate genes, which can then be functionally characterized (Mackay, 2001).

QTL panels in Daphnia

With their long history as model organisms in ecotoxicological studies and the availability of increasing amount of genomic information, *Daphnia* provide an ideal system to genetically characterize QTLs for resistance and responses to environmental toxicity. We now have at our disposal the genomic sequence of *D. pulex*, as well as an EST (expressed sequence tags) database of *D. pulex and D. magna.* Moreover, an extensive number of polymorphic markers are available for *Daphnia* species from the direct cloning and development of simple sequence repeat motifs (SSRs) (Table 1). For example, microsatellite markers have been developed for *D. pulicaria*, many of which cross-amplify in other *Daphnia* species (Colbourne *et al*., 2004). In addition, bioinformatic analyses of the available genome and EST sequences from *D. pulex* will yield many additional SSR loci and single nucleotide polymorphisms (SNPs) useful for fine resolution mapping. The tools are then available to relatively quickly develop genetic markers in many *Daphnia* species. An important advantage of *Daphnia* as a model system for QTL analysis is the ability to clonally propagate individual *Daphnia* genotypes by parthenogenesis. Not only does this property of *Daphnia* increase the power of scoring population-level phenotypes such as toxicity tolerance by allowing replication of genotypes across environments, but also recombinant lines resulting from crosses can be kept in the laboratory for extended periods and repeatedly used for mapping of a wide variety of traits. Cyclical parthenogenesis, however, also has a drawback, as this phenomenon makes line crossing more difficult. Environmental triggering of sexuality varies greatly among lines and the production of sexual eggs is linked to diapause, which is often difficult to break in the laboratory. Performing sexual crosses in *Daphnia* most often leads to substantial losses in numbers. As a result, segregation ratios may be

subject to a strong bias. Careful choice of parental clones will help to alleviate this potential source of bias.

A genetic linkage map has now been produced for *D. pulex* based on 129 recombinant F_2 lines (Cristescu *et al.*, 2006). The map comprises 185 microsatellite markers distributed over 12 linkage groups, with an average inter-marker distance of 7 cM. Notably, a substantial number of the markers (more than 20%) show evidence of a homozygote deficiency in the F_2 generation, probably because these markers are linked to recessive deleterious alleles. Indeed, genetic load seems to be high in *Daphnia* populations (De Meester and De Jager, 1993), with levels of inbreeding depression being among the highest ever reported (Haag *et al.*, 2002). This general pattern of high inbreeding depression indicates that many strains of *Daphnia* may harbour a large load of deleterious recessive alleles, which can lead to patterns of segregation distortion in line-cross mapping panels. Line-cross designs that minimize the impact of deleterious recessive alleles should be utilized in *Daphnia.* There are two possible strategies to reduce the impact of deleterious recessive alleles on subsequent QTL analyses. One method is to construct highly inbred lines to purge the genetic load and select the highest fitness inbreds for use as the parental generation. These lines are then crossed to produce an F_1 generation and then selfed to produce recombinant F_2s (Fig. 4A). This approach requires several rounds of sexual reproduction to inbreed the parental lines and adds a considerable amount of time and effort to the initial phase of mapping panel construction. An alternative strategy to avoid the consequences of fixing deleterious alleles is to conduct two initial crosses using four different outbred parental clones (four-grandparent design, Bradshaw *et al.*, 1998). The resulting F_1s are then crossed reciprocally to produce F_2 generation individuals (Fig. 4B). This design has the advantage of avoiding fixing deleterious recessive alleles in the recombinant F_2 lines. In addition, greater phenotypic diversity can be incorporated in the F_2 mapping panel. The major limitation to the four-grandparent design is the larger number of markers required to distinguish chromosomal segments from each parental line. Given the possibility of long-term maintenance of recombinant F_2's through asexual reproduction, the additional effort required to create highly inbred parental lines may be well warranted and below we describe our efforts to construct QTL panels using the four grand parent design.

We are at present developing resources for QTL analysis in *D. magna* and *D. pulex.* In order to minimize segregation of deleterious recessive alleles in the F_2 generation, inbred clones will be used as parentals in the crosses, then hopefully purging the experimental system of deleterious

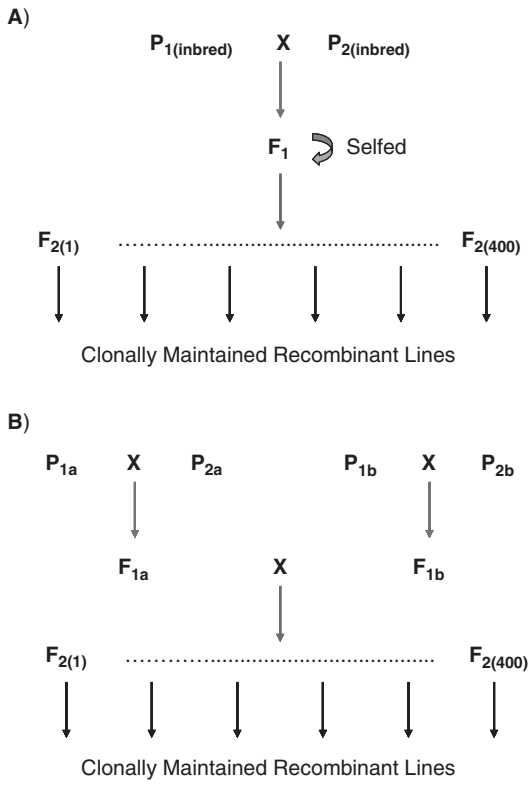

Fig. 4. Alternative crossing designs for QTL panel construction. (A) Inbred parental lines and selfed F_1. (B) Four-grandparent design with outcrossed F_1s. (See Colour Plate Section in this book).

alleles. In addition, to avoid biased segregation of markers linked to genes involved in diapause, it is necessary to use hatching conditions that will not favour one of the parental clones. These precautions should allow us to produce a high-quality recombinant mapping populations that will be available to the whole community of *Daphnia* researchers. For *D. magna*, we are developing markers based on simple sequence repeat motifs from two sources: (i) genomic DNA libraries enriched for repetitive sequences yielded more than 60 polymorphic microsatellite markers and (ii) to supplement these markers, we identified repetitive sequences in a *D. magna* EST database (Watanabe *et al.*, 2005). Utilizing an EST database allows us to develop markers closely linked to coding regions, increasing the probability to detect QTLs (Coulibaly *et al.*, 2005; Vasemagi *et al.*, 2005). A total of 330 EST sequences containing

tandemly repeated DNA have been identified and are currently being tested for polymorphism in *D. magna*. In *D. pulex*, there are in excess of 500 polymorphic SSR loci available (Colbourne *et al*., 2004) and markers linked to virtually every open reading frame can be developed from the complete genome sequence. Availability of these QTL resources in the future will make it possible to efficiently map QTLs involved in a variety of toxicological and ecologically relevant traits, including toxicological responses. There is a strong interest in evolutionary toxicological studies on *Daphnia* examining patterns of genetic adaptation to local ambient pollution levels (Lopes *et al*., 2004). Both the identification of QTLs and the analysis of gene expression profiles using cDNA microarrays (Soetaert *et al*., 2006) are promising avenues for linking genetic adaptation to specific genes. QTL panels from *D. magna* and *D. pulex* will provide a comparative basis for examining the generality of mapping results and will help bridge the existing gap between a model for toxicology (*D. magna*) and a resource-rich model for ecological and evolutionary genomic studies (*D. pulex*). The shared use of common mapping populations for the analysis of different traits will also allow the investigation of correlations among complex phenotypes.

Currently, the genomic tools available for *D. pulex* are more advanced than for *D. magna*. Given the importance of *D. magna* as a prime model organism in toxicological studies and the investment of many research groups in developing genomic resources for this species, there would be enormous benefit for the community to have the genomic tools (including a complete genome sequence) for this organism advanced as quickly as possible. In fact, the *D. magna* genome project has been launched by the DGC (http://dgc.cgb.indiana.edu/display/magna/Home).

Toxicogenomics database – a role for wFleaBase

Toxicological genomics research has a new resource in the *Daphnia* genome and the new genome database wFleaBase (http://wFleaBase.org/; Colbourne *et al*., 2005) provides useful access to it. New genome sequencing projects and communities are facing large informatics tasks for incorporating, curating and annotating and disseminating sequence and annotation data. Biologists should now expect rapid access to new genomes, including basic annotations from well-studied model organisms and predictions to locate potential new genes, to make sense of them. Expertise from existing genome projects can be leveraged into building such tools. The Generic Model Organism Database (GMOD; Stein *et al*., 2002) project has this goal, to fully develop and extend a genome

database tool set to the level of quality needed to create and maintain new genome databases. The wFleaBase database, which is constructed on the GMOD platform, provides scientists with rapid access to this *Daphnia* genome, facilitating new discoveries and understanding for sciences such as toxicological genomics.

Genome database components

wFleaBase is built with common GMOD database components and open source software shared with other genome databases. Use of common components facilitates rapid construction and interoperability. The GMOD ARGOS replicable genome database template (www.gmod.org/argos/) provides a tested set of integrated components. The genome access tools of GMOD – GBrowse (Stein *et al*., 2002), BioMart (Durinck *et al*., 2005) and BLAST (Altschul *et al*., 1997) – are available for searching the *D. pulex* genome. The GMOD Chado relational database schema (www.gmod.org/chado/) is used for managing an extensible range of genome information. Middleware in Perl and Java were added to bring together BLAST, BioMart, sequence reports, searches and other bioinformatics programs for public access. Another aid to integrating and mining these data is GMOD Lucegene (www.gmod.org/lucegene/), that forms a core component for rapid data retrieval by attributes, GBrowse data retrieval and databank partitioning for Grid analyses. wFleaBase operates on several Unix computers including Apple Macintosh OSX and Intel Linux and is portable enough to run on laptop computers for field studies. Genome maps include homologies to nine eukaryote proteomes, marker genes, microsatellite and EST locations and gene predictions. The assemblies and predicted genes can be searched by BLAST and linked to genome maps. BioMart provides searches of the full genome annotation sets, allowing selections of genome regions with and without specific features.

Genome annotations at wFleaBase produced by several groups are provided for map viewing and data mining, including contributions listed in Table 1. Gene predictions with SNAP (Korf, 2004) have been generated to locate new as well as known genes. SNAP guided by protein homology evidence is one of the better *ab initio* predictors when: (i) new genes are sought and (ii) there are no close relatives with an experimentally verified genome annotation. SNAP works well on the range of eukaryote genomes (plant to animal, small to big) with minimal homology data. A drawback is that SNAP overpredicts genes, but such aggressiveness can prove useful in identifying gene-like features.

The TeraGrid project (www.teragrid.org) is part of a shared cyber infrastructure for sciences, funded primarily by NSF. TeraGrid provides

collaborative, cost-effective scientific computing infrastructure much in the same way the GMOD initiative is building common tools for genome databases. The TeraGrid system is particularly suitable for genome assembly, annotation, gene finding and phylogenetic analyses. TeraGrid computers have been employed to annotate and validate the assembly of *D. pulex*. Results include homologies to nine eukaryote proteomes, gene predictions, marker genes, microsatellite and EST locations. Proteome comparisons included 217,000 proteins drawn from source genome databases, Ensembl and NCBI, for human, mouse, zebrafish, fruitfly, mosquito, bee, worm, mustard weed and yeast.

Database uses

wFleaBase provides a resource to biologists interested in comparing *Daphnia* to known genomes, finding novel and known genes, genome structure and evolution and gene function associations. These known genes provide useful access to this new genome for many researchers interested in locating a particular gene or gene family. The known gene matches also offer searches and cataloguing gene contents by known functions. Figure 5 summarizes known model organism genes found in *Daphnia* and two insects.

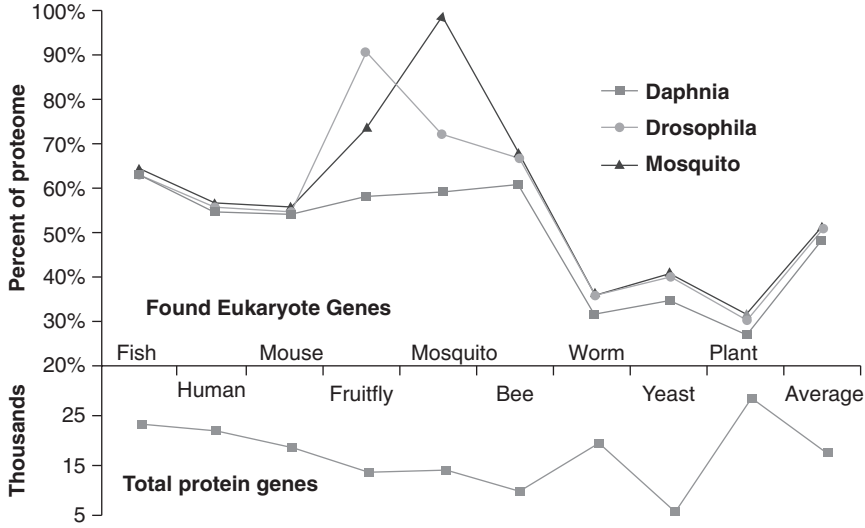

Fig. 5. Percent of full gene sets of nine eukaryote genomes found in new genomes *Daphnia pulex, Drosophila virilis,* with out-group *Anopheles gambia.* Lower line shows count of protein gene sets. *D. virilis* has 90% similarity to model fruitfly *D. melanogaster and A. gambiae* has 100% similarity to itself. (See Colour Plate Section in this book).

BLAST searches in wFleaBase

Searching for known gene sequences or gene families with BLAST remains one of the best ways to probe genomes. For example, one might be interested in cytochrome P450 genes and want to search for their presence in the *Daphnia* genome. Using the mouse P450 gene, MGI:88607 or GenBank:NP_067257, which is involved in haeme and iron ion binding, monooxygenase and oxidoreductase activity, a BLAST search against predicted proteins with BlastP returned several high-scoring matches. The best was to 'scaffold_3-snapho.108.' The BLAST report contained usual statistics and alignment values and also provided visualization by linking to a genome map view of the match. As a check of this predilation, a search of the chromosome 'scaffold' DNA with tBlastN returned several matches and the best was also at scaffold_3. The genome map view of this BLAST search linked to the same predicted gene, 'scaffold_3-snapho.108,' which is shown in Fig. 6. Two *Daphnia*

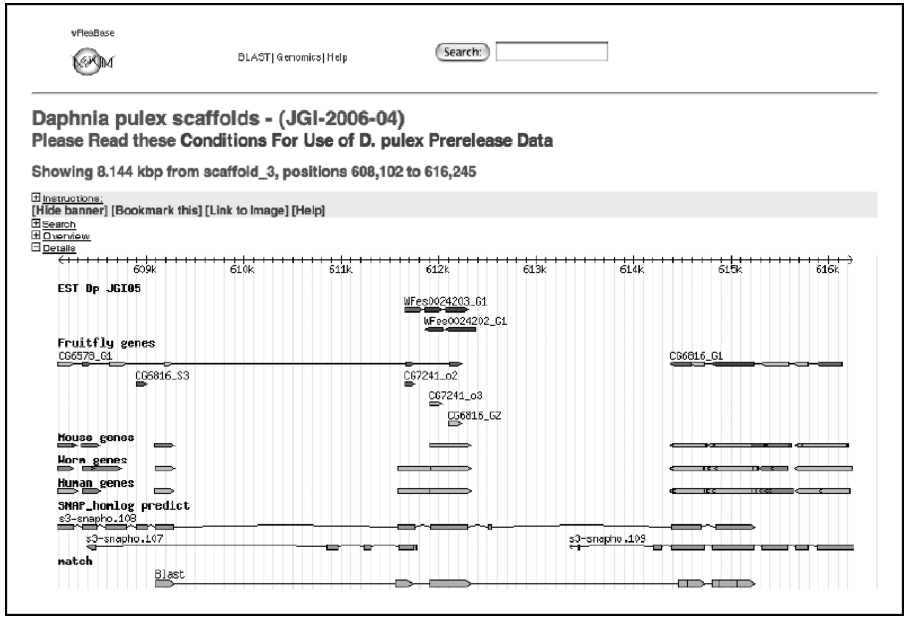

Fig. 6. Cytochrome P450 gene located on *Daphnia pulex* genome. This GBrowse map view at wFleaBase is returned from a BLAST search for mouse gene MGI:88607 (GenBank:NP_067257) and matches well the predicted *Daphnia* gene s3-snapho.108. Two *Daphnia* EST matches part of this gene and homologous genes from fruitfly, mouse, worm and human match.

ESTs are also found at this location and the source material can be readily accessed for follow-up studies of this gene.

Genome maps
Maps of the genome form the core, with BLAST searches, of discovery tools for bioscientists. Maps including available annotations from several groups are provided using GBrowse. The BLAST result reports include hyperlinks from each alignment match to the respective genome map, as well as to sequence and GFF annotation results. As seen in Fig. 6, this tool allows one to view evidence of common gene predictions and features in homologous regions.

Gene functions and biological processes
To provide an assessment of possible gene gain and loss among *Daphnia*, gene matches to Gene Ontology (GO) categories by species were tabulated and provided at section genome-summaries/gene-GO-function-association, which are discussed in functional detail in Colbourne *et al*. (2007). These may indicate species differences in functional categories when statistically significant deviations are indicated. While low counts, suggestive of missing genes, may be caused by divergence of genes, extra gene matches more strongly suggest categories in which species differ. Among the interesting effects, *Daphnia* may have higher gene counts than insects or *Caenorhabditis elegans* for catalytic activity (GO:0003824), hydrolase activity (GO:0016787), peptidase activity (GO:0008233) and transferase activity (GO:0016740). There is also a suggestion of lower gene numbers for receptor activity, protein binding and enzyme regulator activity.

The gene matches were high-scoring segment pair groupings and include various events: gene duplications, alternate splice exons within genes, new genes that appear composed of exons from other genes, as well as computational artifacts. Detailed evidence pages provide links to GBrowse genome map views showing all secondary high-scoring segment pair groupings. Proteome sources in this analysis were those organism with extensive GO annotations: Dmel fruitfly, mouse, *C. elegans* worm and yeast. GO-Slim groupings are used for Biological Process, Molecular Function, Cell Location (125 categories).

Genome data mining
An emerging trend among bioscientists and bioinformaticians is to use data mining of large subsets of genome data, often focused on summary information for a range of common attributes. These data are used in spreadsheets and simple databases or analyses. The Ensembl project has

produced BioMart (Durinck *et al*., 2005), used at wFleaBase for searches of the full *Daphnia* genome by various attributes of homologous genes, such as known toxicogenomic genes. One can retrieve tables of matching gene locations, Ids and annotations or retrieve the genome sequences for these in bulk Fasta format, which is suitable for BLAST and sequence tools. With BioMart, one can select or exclude genome regions with the available annotations and download tables or sequences of the selection set. For instance, one could select the regions of the *Daphnia* genome in which there are gene homologs similar to fish but fruitfly homologs are missing or regions with gene predictions in which there are no known homologs.

Future additions
Additions planned for this *Daphnia* genome database include experimental data from toxicology studies, such as gene expression in environmentally stressed populations. Metabolic pathways and cell cycle views of identified genes will be added to this database. These will provide an aid to quickly classify genes that belong to common and toxicologically important biochemical processes. Future growth of this database as a toxicogenomic resource will benefit much from use and contributions of data from the scientific community.

Concluding remarks

Daphnia are without question an established model species for toxicological studies. This sentinel species of inland aquatic habitats is among the most well-studied animals for toxicological testing and for ecological research. Their position in the emerging field of toxicological genomics will progress concomitantly with the development of genomic resources. In general, these resources are expected to provide reliable diagnostic tools that determine whether a defined ecosystem (or individual) is exposed to toxic agents (Andrew *et al*., 2003), whether substances are harmful (Hayes and Bradfield, 2005) and whether a population is susceptible to certain chemicals (Pedra *et al*., 2004). Genomic tools offer the promise of providing data that are more sensitive, more quickly obtained, less expensive and better able to detect all bioreactive compounds, including parent compounds and their metabolites, than traditional monitoring tools. When linked to well-defined biological endpoints – including cellular, physiological, life-history and population-level responses – they also promise to reveal insights into the modes of toxicity that are shared among varying classes

of compounds and across species. For *Daphnia*, genomic technologies benefit from anchoring to abundant existing knowledge about the response of species to toxicological and ecological challenges. These challenges can at times have important and unexpected manifestations by their combined actions. For example, research suggests that certain toxicants negatively affect the ability of populations to respond appropriately to predation threat (Hunter and Pyle, 2004) and to the normal cycles of population growth and reproduction (Oda *et al.*, 2005). *Daphnia* also provide the experimental platform to probe past populations and integrate genomic response through evolutionary time scales.

Genomic resources are rapidly becoming available for *Daphnia* research by the coordinated efforts of the DGC. Over the past five years, the genomic utility box (Table 1) for *Daphnia* has experienced rapid growth. The box started with a set of molecular markers for genetic mapping and identifying genes of interest by QTL analyses; a pilot microarray platform to identify synchronized gene responses to experimental conditions; and cDNA libraries and sequences used to uncover the existence of conserved genes of interest that have already been characterized in other model systems. The utility box for *Daphnia* now includes a fully sequenced genome and draft annotation. The genome sequence provides a resource that will greatly facilitate research to identify major genes and their interactions (Fig. 3) that can account for the successes or failures of individuals at coping with chemicals and of populations at adapting to these challenges. Given the complexities of genome-wide studies, however and the scope of toxicological research, which often necessitates validation of results by detailed functional studies in the lab and field, integration would benefit from an approach similar to studies in development, cell and molecular biology. As a starting point, this involves building a research community around a common set of model species. Research using *Daphnia* is at present focused on two species. Expanding this research to include comparative studies among many species relies on creating the maximum set of resources for both *D. magna* and *D. pulex*. From this foundation, studies using other daphniids or even more distantly related species are made easier by comparing results to a reference model system for toxicological genomics. This strategy should lead to a marked improvement in our ability to extrapolate toxicological responses and biological outcomes across animal populations, through ecological food webs and from animal models to humans – a goal that is made possible by the general evolutionary conservation of genes and their interactions within biochemical pathways.

Acknowledgements

We thank Luc De Meester and Dieter Ebert for helpful comments on the manuscript. cDNA libraries were created by Darren Bauer and Zack Smith in the laboratories of Kelley Thomas and John Colbourne. Sequencing was produced by Erika Lindquist and Jeff Boore at the Joint Genome Institute. Kelley Thomas and Michael Lynch are key contributors of the *Daphnia* Genomics Consortium. Support for the *Daphnia* genomic utility box was provided in part by the U.S. National Science Foundation and by the DOE Joint Genome Institute under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, Los Alamos National Laboratory under Contract No. W-7405-ENG-36, and in collaboration with the *Daphnia* Genomics Consortium, http://daphnia.cgb.indiana.edu. wFleaBase was developed at the Genome Informatics Lab of Indiana University with support to Don Gilbert from the National Science Foundation and the National Institute of Health. Other sources of support include NSF DEB-022183 (JKC, JRS); NIEHS-SBRP ES ES07373-07 (JRS); NSF DEB-021212487 (MEP); a USU Center for Integrated BioSystems research grant (MEP); NERC NER/D/S/2002/00413 (RMS, AC); The School of Public and Environmental Affairs, Indiana University, and The Center for Genomics and Bioinformatics at Indiana University, which is supported in part by the METACyt Initiative of Indiana University, funded in part through a major grant from the Lilly Endowment, Inc. Our work benefits from and contributes to the *Daphnia* Genomics Consortium.

References

Adamowicz, S. J., Hebert, P. D. N. and Marinone, M. C. (2004). Species diversity and endemism in the *Daphnia* of Argentina: a genetic investigation. *Zool. J. Linn. Soc. Lond.* 140,171–205.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F. *et al.* (1991). Complementary-DNA sequencing – expressed sequence tags and Human Genome Project. *Science* 252,1651–1656.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25,3389–3402.

Amin, R., Hamadeh, H. K., Bushel, P. R., Bennett, L., Afshari, C. A. and Paules, R. S. (2002). Genomic interrogation of mechanism(s) underlying cellular responses to toxicants. *Toxicology* 27,555–563.

Anderson, B. (1944). The toxicity thresholds of various substances found in industrial wastes as determined by the use of *Daphnia magna*. *Sewage Work J.* 16,1156–1165.

Anderson, B. G. (1945). The toxicity of DDT to *Daphnia*. *Science* 102,539.

Andrew, A., Warren, A. J., Barchowsky, A., Temple, K. A., Klei, L., Soucy, N. V., O'Hara, K. A. and Hamilton, J. W. (2003). Genomic and proteomic profiling of responses to toxic metals in human lung cells. *Environ. Health Perspect.* 111, 825–838.

Audic, S. and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7,986–995.

Averof, M. and Akam, M. (1995a). Hox genes and the diversification of insect and crustacean body plans. *Nature* 376,420–423.

Averof, M. and Akam, M. (1995b). Insect-crustacean relationships: Insights from comparative developmental and molecular studies. *Philos. Trans. R. Soc.* B347, 293–303.

Ball, C. A. and Brazma, A. (2006). MGED standards: Work in progress. *Omics* 10, 138–144.

Banta, A. (1939). *Studies on the physiology, genetics and evolution of some cladocera*. Carnegie Institution, Washington, DC.

Bartosiewicz, M., Jenkins, D., Penn, S., Emery, J. and Buckpitt, A. (2001). Unique gene expression patterns in liver and kidney associated with exposure to chemical toxicants. *J. Pharmacol. Exp. Ther.* 297,895–905.

Bing, N. and Hoeschele, I. (2005). Genetical genomics analysis of a yeast segregant population for transcription network inference. *Genetics* 170,533–542.

Blomberg, S. P. and Garland, T. (2002). Tempo and mode in evolution: Phylogenetic inertia, adaptation and comparative methods. *J. Evolution. Biol.* 15,899–910.

Boore, J. L., Lavrov, D. V. and Brown, W. M. (1998). Gene translocation links insects and crustaceans. *Nature* 392,667–668.

Bradshaw, H. D., Otto, K. G., Frewen, B. E., McKay, J. K. and Schemske, D. W. (1998). Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics* 149,367–382.

Brede, N., Thielsch, A., Sandrock, C., Spaak, P., Keller, B., Streit, B. and Schwenk, K. (2006). Microsatellite markers for European *Daphnia*. *Mol. Ecol. Notes* 6,536–539.

Breukelman, J. (1932). Effect of age and sex on resistance of daphnids to mercuric chloride. *Science* 76,302.

Cáceres, C. E. (1998). Interspecific variation in the abundance, production and emergence of *Daphnia* diapausing eggs. *Ecology* 79,1699–1710.

Campbell, A. K., Wann, K. T. and Matthews, S. B. (2004). Lactose causes heart arrhythmia in the water flea *Daphnia pulex*. *Comp. Biochem. Physiol. B* 139,225–234.

Carlborg, O. and Haley, C. S. (2004). Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* 5,618–625.

Carlborg, O., De Koning, D. J., Manly, K. F., Chesler, E., Williams, R. W. and Haley, C. S. (2005). Methodological aspects of the genetic dissection of gene expression. *Bioinformatics* 21,2383–2393.

Carpenter, S. R., Kitchell, J. F., Hodgson, J. R., Cochran, P. A., Elser, J. J., Elser, M. M., Lodge, D. M., Kretchmer, D., He, X. and Vonende, C. N. (1987). Regulation of lake primary productivity by food web structure. *Ecology* 68,1863–1876.

Cheung, V. and Spielman, R. S. (2002). The genetics of variation in gene expression. *Nat. Genet.* 32,522–555.

Chiavelli, D. A., Marsh, J. W. and Taylor, R. K. (2001). The mannose-sensitive hemagglutinin of *Vibrio cholerae* promotes adherence to zooplankton. *Appl. Environ. Microbiol.* 67,3220–3225.

Colbourne, J. K. and Hebert, P. D. N. (1996). The systematics of North American *Daphnia* (*Crustacea: Anomopoda*): A molecular phylogenetic approach. *Philos. Trans. R. Soc. Lond. B* 351,349–360.

Colbourne, J. K., Crease, T. J., Weider, L. J., Hebert, P. D. N., Dufresne, F. and Hobaek, A. (1998). Phylogenetics and evolution of a circumarctic species complex (*Cladocera: Daphnia pulex*). *Biol. J. Linn. Soc. Lond.* 65,347–365.

Colbourne, J. K., Eads, B. D., Shaw, J. R., Bohuski, E., Bauer, D. and Andrews, J. (2007). Sampling *Daphnia*'s expressed genes: Preservation, expansion and invention of crustacean genes with reference to insect genomes. *BMC Genomics* 8,217.

Colbourne, J. K., Hebert, P. D. N. and Taylor, D. J. (1997). Evolutionary origins of phenotypic diversity in *Daphnia*. In *Molecular Evolution and Adaptive Radiation* (eds T. J. Givnish and K. J. Sytsma), pp. 163–188, Cambridge University Press, London.

Colbourne, J. K., Robison, B., Bogart, K. and Lynch, M. (2004). Five hundred and twenty-eight microsatellite markers for ecological genomic investigations using *Daphnia*. *Mol. Ecol. Notes* 4,485–490.

Colbourne, J. K., Singan, V. R. and Gilbert, D. G. (2005). wFleaBase: The *Daphnia* genome database. *BMC Bioinformatics* 6,45.

Colbourne, J. K., Wilson, C. C. and Hebert, P. D. N. (2006). The systematics of Australian Daphnia and Daphniopsis (Crustacea: Cladocera): A shared phylogenetic history transformed by habitat-specific rates of evolution. *Biol. J. Linn. Soc.* 89,469–488.

Connon, R., Hooper, H. L., Lim, F. L., Moore, D. J., Watanabe, H., Soetart, A., Cook, K., Sibly, R. M., Orphanides, G., Maund, S. J., Hutchinson, T. H., Moggs, J., De Coen, W., Iguchi, T. and Callaghan, A. (2008). Linking molecular and population stress responses in *Daphnia magna* exposed to cadmium. *Environ. Sci. Technol.* 42,2181–2188.

Cook, J., Denslow, N. D., Iguchi, T., Linney, E. A., Miracle, A., Shaw, J. R., Viant, M. R. and Zacharewski, T. R. (2007). 'Omics' approaches in the context of environmental toxicology. In *Genomic Approaches for Cross-Species Extrapolation in Toxicology* (eds R. DiGiulio and W. H. Benson), Taylor and Francis, Washington, DC.

Cooney, J. (1995). Effects-toxicity testing. In *Fundamentals of Aquatic Toxicology* (ed. G. Rand), 2nd Edition, pp. 71–102, Taylor and Francis, Washington, DC.

Coulibaly, I., Gharbi, K., Danzmann, R. G., Yao, J. and Rexroad, C. E. (2005). Characterization and comparison of microsatellites derived from repeat-enriched libraries and expressed sequence tags. *Anim. Genet.* 36,309–315.

Cousyn, C., De Meester, L., Colbourne, J. K., Brendonck, L., Verschuren, D. and Volckaert, F. (2001). Rapid, local adaptation of zooplankton behavior to changes in predation pressure in the absence of neutral genetic changes. *Proc. Natl. Acad. Sci. USA* 98,6256–6260.

Cristescu, M. E., Colbourne, J. K., Radivojac, J. and Lynch, M. (2006). A microsatellite-based genetic linkage map of the water flea, *Daphnia pulex*: On the prospect of crustacean genomics. *Genomics* 88,415–430.

Davis, J. (1977). Standardization and protocols of bioassays – their role and significance for monitoring, research and regulatory usage. In *Proceedings of the 3rd Aquatic Toxicity Workshop,* (eds E. P. W. R. Parker, P. G. Wells and G. F. Westlake), Environmental Protection Service, Technical Report No. EPS-5-AR-77-1, Halifax, NS.

Daphnia Genomics Consortium (2007). http://daphnia.cgb.indiana.edu (Accessed on October 26).

de Bernardi, R. and Peters, R. H. (1987). Why *Daphnia*. In '*Daphnia*' *Memorie dell'Istituto Italiano di Idrobiologia Dr Marco De Marchi* (eds R. H. Peters and R. de Bernardi), Vol. 45, Consiglio Nazionale Delle Ricerche Instituto Italiano Di Idrobiologia, Verbania Pallanza.

De Hoogh, C., Wgenvoort, A. J., Jonker, F., Van Leerdam, J. A. and Hogenboom, A. C. (2006). HPLC-DAD and Q-TOF MS techniques identify cause of *Daphnia* biomonitor alarms in the river Meuse. *Environ. Sci. Technol.* 40,2678–2685.

de Koning, D. J., Carlborg, O. and Haley, C. S. (2005). The genetic dissection of immune response using gene-expression studies and genome mapping. *Vet. Immunol. Immunopathol.* 105,343–352.

De Meester, L. (1996). Local genetic differentiation and adaptation in freshwater zooplankton populations: Patterns and processes. *Ecoscience* 3,385–399.

De Meester, L. and De Jager, H. (1993). Hatching of *Daphnia* sexual eggs. I. Intraspecific differences in the hatching responses of *D. magna* eggs. *Freshwater Biol.* 30,219–226.

Denslow, N., Colbourne, J. K., Dix, D., Freedman, J. H., Helbing, C. C., Kennedy, S. and Williams, P. L. (2007). Selection of surrogate animal species for comparative toxicogenomics. In *Genomic Approaches for Cross-Species Extrapolation in Toxicology* (eds R. DiGiulio and W. H. Benson), Taylor and Francis, Washington, DC.

Diachenko, L. B., Ledesma, J., Chenchik, A. A. and Siebert, P. D. (1996). Combining the technique of RNA fingerprinting and differential display to obtain differentially expressed mRNA. *Biochem. Biophys. Res. Commun.* 219,824–828.

Dix, D., Gallagher, K., Benson, W. H., Groskinsky, B. L., McClintock, J. T., Dearfield, K. L. and Farland, W. H. (2006). A framework for the use of genomics data at the EPA. *Nat. Biotechnol.* 24,1108–1111.

Dodson, S. I. and Hanazato, T. (1995). Commentary on effects of anthropogenic and natural organic-chemicals on development, swimming behavior and reproduction of *Daphnia*, a key member of aquatic ecosystems. *Environ. Health Perspect.* 103,7–11.

Dong, Y. J., Ogawa, T., Lin, D. Z., Koh, H. J., Kamiunten, H., Matsuo, M. and Cheng, S. H. (2006). Molecular mapping of quantitative trait loci for zinc toxicity tolerance in rice seedling (*Oryza sativa* L.). *Field Crops Res.* 95,420–425.

Drysdale, R. A., Crosby, M. A. and the Fly Consortium. (2005). FlyBase: Genes and gene models. *Nucleic Acids Res.* 33,D390–D395.

Duodoroff, P. and Katz, M. (1950). Critical review of literature on the toxicity of industrial wastes and their components to fish. I. Alkalies, acids and inorganic gases. *Sewage Ind. Wastes* 22,1432–1458.

Dudycha, J. L. (2001). The senescence of *Daphnia* from risky and well tolerated habitats. *Ecol. Lett.* 4,102–105.

Dudycha, J. L. (2003). A multienvironment comparison of senescence between sister species of *Daphnia*. *Oecologia* 135,555–563.

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A. and Huber, W. (2005). BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* 21,3439–3440.

Eads, B., Colbourne, J. K., Bohuski, E. and Andrews, J. (2007). Profiling sex-biased gene expression during parthenogenetic reproduction in *Daphnia pulex*. *BMC Genomics* 8,464.

Eads, B. D., Andrews, J. and Colbourne, J. K. (2008). Ecological genomics in *Daphnia*: Stress responses and environmental sex determination. *Heredity* 100,184–190.

Eaton, D. and Klaassen, C. D. (1996). Principles of toxicology. In *Casarett and Doull's Toxicology: The Basic Science of Poisons* (ed. C. Klaassen), 5th Edition, McGraw-Hill, New York.

Ebert, D., Carius, H. J., Little, T. and Decaestecker, E. (2004). The evolution of virulence when parasites cause host castration and gigantism. *Am. Nat.* 164,S19–S32.

Edmondson, W. (1987). *Daphnia* in experimental ecology: Notes on historical perspectives. In '*Daphnia*' *Memorie dell'Istituto Italiano di Idrobiologia Dr Marco De Marchi* (eds R. H. Peters and R. de Bernardi), Vol. 45, Consiglio Nazionale Delle Ricerche Instituto Italiano Di Idrobiologia, Verbania Pallanza.

Elendt, B. and Bias, W. R. (1990). Trace nutrient deficiency in *Daphnia magna* cultured in standard medium for toxicity testing: Effects of the optimization of culture conditions on life history parameters of *Daphnia magna*. *Water Res.* 24,1157–1167.

Ford, A. T. and Fernandes, T. F. (2005). Better the devil you know? A precautionary approach to using amphipods and daphnids in endocrine disruptor studies. *Environ. Toxicol. Chem.* 24,1019–1021.

Fox, J. A. (2004). New microsatellite primers for *Daphnia galeata mendotae*. *Mol. Ecol. Notes* 4,544–546.

Friedrich, M. and Tautz, D. (1995). Ribosomal DNA phylogeny of the extant arthropod classes and the evolution of myriapods. *Nature* 376,165–167.

Gallagher, K., Benson, W. H., Brody, M., Fairbrother, A., Hasan, J., Klaper, R., Lattier, D., Lundquist, S., McCarroll, N., Miller, G., Preston, J., Sayre, P., Smith, B., Street, A., Troast, R., Vu, B., Reiter, L., Farland, W. and Dearfield, K. (2006). Genomics: Applications, challenges and opportunities for the U.S. Environmental Protection Agency. *Hum. Ecol. Risk Assess.* 12,572–590.

Gallo, M. (1996). History and scope of toxicology. In *Casarett and Doull's Toxicology: The Basic Science of Poisons* (ed. C. Klaassen), 5th Edition, McGraw-Hill, New York.

Glover, C. N. and Wood, C. M. (2005). Physiological characterisation of a pH- and calcium-dependent sodium uptake mechanism in the freshwater crustacean, *Daphnia magna*. *J. Exp. Biol.* 208,951–959.

Gracey, A. Y., Troll, J. V. and Somero, G. N. (2001). Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *Proc. Natl. Acad. Sci. USA* 98, 1993–1998.

Haag, C. R., Hottinger, J. W., Riek, M. and Ebert, D. (2002). Strong inbreeding depression in a *Daphnia* metapopulation. *Evolution* 56,518–526.

Haber, F. (1924). Zur geschichte des gaskrieges (On the history of gas warfare). In *Funf Vortrage aus den Jahren 1920–1923 (Five Lectures from the Years 1920–1923)*, pp. 76–92, Springer, Berlin.

Hairston, N., Van Brunt, R. A., Kearns, C. M. and Engstrom, D. R. (1995). Age and survivorship of diapausing eggs in a sediment egg bank. *Ecology* 76,1706–1711.

Hairston, N. G., Holtmeier, C. L., Lampert, W., Weider, L. J., Post, D. M., Fischer, J. M., Cáceres, C. E., Fox, J. A. and Gaedke, U. (2001). Natural selection for grazer resistance to toxic cyanobacteria: Evolution of phenotypic plasticity? *Evolution* 55,2203–2214.

Hairston, N. G., Lampert, W., Cáceres, C. E., Holtmeier, C. L., Weider, L. J., Gaedke, U., Fischer, J. M., Fox, J. A. and Post, D. M. (1999). Lake ecosystems – rapid evolution revealed by dormant eggs. *Nature* 401,446.

Hall, S. R., Tessier, A. J., Duffy, M. A., Huebner, M. and Cáceres, C. E. (2006). Warmer does not have to mean sicker: Temperature and predators can jointly drive timing of epidemics. *Ecology* 87,1684–1695.

Hamadeh, H., Bushel, P. R., Jayadey, S., Martin, K., DiSorbo, O., Sieber, S., Bennett, L., Tennant, R., Stoll, T., Barrett, J. C., Blanchard, K., Paules, R. S. and Afshari, C. A. (2002). Gene expression analysis reveals chemical-specific profiles. *Toxicol. Sci.* 67,232–240.

Hayes, K. and Bradfield, R. (2005). Advances in toxicogenomics. *Chem. Res. Toxicol.* 18,403–414.

Hebert, P. (1987). Genetics of *Daphnia*. In 'Daphnia' *Memorie dell'Istituto Italiano di Idrobiologia Dr Marco De Marchi* (eds R. H. Peters and R. de Bernardi), Vol. 45, Consiglio Nazionale Delle Ricerche Instituto Italiono Di Idrobiologia, Verbania Pallanza.

Hebert, P. and Ward, R. D. (1972). Inheritance during parthenogenesis in *Daphnia magna*. *Genetics* 71,639–642.

Hebert, P. D. N. (1974a). Ecological differences between genotypes in natural populations of *Daphnia magna*. *Heredity* 33,327–337.

Hebert, P. D. N. (1974b). Enzyme variability in natural populations of *Daphnia magna* II. Genotypic frequencies in permanent populations. *Genetics* 77,323–334.

Hebert, P. D. N. (1974c). Enzyme variability in natural populations of *Daphnia magna* III. Genotypic frequencies in intermittent populations. *Genetics* 77,335–341.

Hebert, P. D. N. and Finston, T. L. (1993). A taxonomic reevaluation of North American *Daphnia* (*Crustacea: Cladocera*). I. The *Daphnia similis* complex. *Can. J. Zool.* 71,908–925.

Hebert, P. D. N. and Finston, T. L. (1996). A taxonomic reevaluation of North American *Daphnia* (*Crustacea: Cladocera*). II: New species in the *Daphnia pulex* group from the south-central United States and Mexico. *Can. J. Zool.* 74,632–653.

Hebert, P. D. N. and Finston, T. L. (1997). A taxonomic reevaluation of North American *Daphnia* (*Crustacea: Cladocera*): III. The *D. catawba* complex. *Can. J. Zool.* 75,1254–1261.

Heckmann, L. H., Connon, R., Hutchinson, T. H., Maund, S. J., Sibly, R. M. and Callaghan, A. (2006). Expression of target and reference genes in *Daphnia magna* exposed to ibuprofen. *BMC Genomics* 7,175.

Hubner, N., Wallace, C. A., Zimdahl, H., Petretto, W., Schultz, H., Maciver, F., Muller, M., Hummel, O., Monti, J., Zidek, V., Musiolova, A., Kren, V., Causton, H., Game, M., Born, G., Schmidt, S., Muller, A., Cook, S. A., Kurtz, T. W., Wittaker, J., Pravenec, M. and Aitman, T. J. (2005). Integrating transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat. Genet.* 37,243–253.

Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H. Y., He, Y. D. D. *et al.* (2000). Functional discovery via a compendium of expression profiles. *Cell* 102,109–126.

Hunter, K. and Pyle, G. (2004). Morphological responses of *Daphnia pulex* to *Chaoborus americanus* kairomone in the presence and absence of metals. *Environ. Toxicol. Chem.* 23,1311–1316.

Hutchinson, G. (1932). Experimental studies in ecology. I. The magnesium tolerance of *Daphniidae* and its ecological significance. *Int. Rev. Ges. Hydrobiol. Hydrogr.* 28,90–108.

Jansen, R. C. and Nap, J. P. (2001). Genetical genomics: The added value from segregation. *Trends Genet.* 17,388–391.

Jensen, K. H., Little, T., Skorping, A. and Ebert, D. (2006). Empirical support for optimal virulence in a castrating parasite. *PLoS Biol.* 4,1265–1269.

Kerfoot, W. C., Robbins, J. A. and Weider, L. J. (1999). A new approach to historical reconstruction: Combining descriptive and experimental paleolimnology. *Limnol. Oceanogr.* 44,1232–1247.

Kilham, S. S., Kreeger, D. A., Lynn, S. G., Goulden, C. E. and Herrera, L. (1998). COMBO: A defined freshwater culture medium for algae and zooplankton. *Hydrobiologia* 377,147–159.

Klaper, R. and Thomas, M. A. (2004). At the crossroads of genomics and ecology: The promise of a canary on a chip. *BioScience* 54,403–412.

Klaper, R., Rees, C., Carvan, M., Weber, D., Drevnick, P. and Sandheinrich, M. (2006). Gene expression links to endocrine function and reproduction decline after mercury exposure in fathead minnows. *Environ. Health Perspect.* 114,1337–1343.

Klugh, A. and Miller, H. C. (1926). The hydrogen ion concentration range of *Daphnia magna*. *Trans. R. Soc. Can.* 20,225–227.

Koivisto, S. (1995). Is *Daphnia magna* an ecologically representative zooplankton species in toxicity tests. *Environ. Pollut.* 90,263–267.

Koivisto, S., Ketola, M. and Walls, M. (1992). Comparison of 5 cladoceran species in short-term and long-term copper exposure. *Hydrobiologia* 248,125–136.

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* 5,59.

Korovchinsky, N. M. (1997). On the history of studies on cladoceran taxonomy and morphology, with emphasis on early work and causes of insufficient knowledge of the diversity of the group. *Hydrobiologia* 360,1–11.

Larkin, P., Folmar, L. C., Hemmer, M. J., Poston, A. J., Lee, H. S. and Denslow, N. D. (2002). Array technology as a tool to monitor exposure of fish to xenoestrogens. *Mar. Environ. Res.* 54,395–399.

Li, H. Q., Lu, L., Manly, K. F., Chesler, E. J., Bao, L., Wang, J. T., Zhou, M., Williams, R. W. and Cui, Y. (2005). Inferring gene transcriptional modulatory relations: A genetical genomics approach. *Hum. Mol. Genet.* 14,1119–1125.

Limburg, P. A. and Weider, L. J. (2002). 'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proc. R. Soc. Lond. B Biol. Sci.* 269,281–287.

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259,946–951.

Little, T. J. and Ebert, D. (2000). The cause of parasitic infection in natural populations of *Daphnia* (*Crustacea*: *Cladocera*): The role of host genetics. *Proc. R. Soc. Lond. B Biol. Sci.* 267,2037–2042.

Little, T. J., O'Connor, B., Colegrave, N., Watt, K. and Read, A. F. (2003). Maternal transfer of strain-specific immunity in an invertebrate. *Curr. Biol.* 13,489–492.

Lopes, I., Baird, D. J. and Ribeiro, R. (2004). Genetic determination of tolerance to lethal and sublethal copper concentrations in field populations of *Daphnia longispina*. *Arch. Environ. Contam. Toxicol.* 46,43–51.

Lubbock, J. (1857). An account of the two methods of reproduction in *Daphnia* and of the structure of the ephippium. *Philos. Trans. R. Soc. Lond.* 8,352–354.

Lynch, M. (1983). Ecological genetics of *Daphnia pulex*. *Evolution* 37,358–374.

Lynch, M. and Gabriel, W. (1983). Phenotypic evolution and parthenogenesis. *Am. Nat.* 122,745–764.

Lynch, M. and Spitze, K. (1994). Evolutionary genetics of *Daphnia*. In *Ecological Genetics* (ed. L. A. Real), pp. 109–128, Princeton University Press, Princeton, NJ.

Lynch, M. and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., Sunderland.

Mackay, T. F. C. (2001). Quantitative trait loci in *Drosophila*. *Nat. Rev. Genet.* 2,11–20.

Mattingly, C. J., Colby, G. T., Rosenstein, M. C., Forrest, J. N. and Boyer, J. L. (2004). Promoting comparative molecular studies in environmental health research: An overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J.* 4,5–8.

McClearn, G. E., Jones, B., Blizard, D. A. and Plomin, R. (1993). The utilization of quantitative trait loci in toxicogenetics. *J. Exp. Anim. Sci.* 35,251–258.

Merrick, B. and Bruno, M. E. (2004). Genomic and proteomic profiling for biomarkers and signature profiles of toxicity. *Curr. Opin. Mol. Ther.* 6,600–607.

Miller, F., Schlosser, P. M. and Janszen, D. B. (2000). Haber's rule: A special case in a family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint. *Toxicology* 149,21–34.

Miner, B. G., Sultan, S. E., Morgan, S. G., Padilla, D. K. and Relyea, R. A. (2005). Ecological consequences of phenotypic plasticity. *Trends Ecol. Evol.* 20,685–692.

Morley, M., Molony, C. M., Weber, T. M., Delvin, J. L., Ewens, K. G., Spielman, R. S. and Cheung, V. G. (2004). Genetic analysis of genome-wide expression variation in human gene expression. *Nature* 430,743–747.

Mort, M. A. (1991). Bridging the gap between ecology and genetics: The case of freshwater zooplankton. *Trends Ecol. Evol.* 6,41–45.

Mueller, O. (1785). Entomostraca seu insecta testacea, quae in aquis Daniae et Norvegiae reperit, descriptsit et inconibus illustravit. Lipsiae et Harniae.

Nardi, F., Spinsanti, G., Boore, J. L., Carapelli, A., Dallai, R. and Frati, F. (2003). Hexapod origins: Monophyletic or paraphyletic? *Science* 299,1887–1889.

Nielsen, R. (2006). Evolution – Why sex? *Science* 311,960–961.

Oda, S., Tatarazako, N., Watanabe, H., Morita, M. and Iguchi, T. (2005). Production of male neonates in *Daphnia magna* (*Cladocera, Crustacea*) exposed to juvenile hormones and their analogs. *Chemosphere* 61,1168–1174.

Olmstead, A. W. and LeBlanc, G. A. (2003). Insecticidal juvenile hormone analogs stimulate the production of male offspring in the crustacean *Daphnia magna*. *Environ. Health Perspect.* 111,919–924.

Paland, S. and Lynch, M. (2006). Transitions to asexuality result in excess amino acid substitutions. *Science* 311,990–992.

Pedra, J., McIntyre, L., Scharf, M. and Pittendrigh, B. (2004). Genome-wide transcription profile of field- and laboratory-selected dichlorodiphenyltrichloroethane (DDT)-resistant *Drosophila*. *Proc. Natl. Acad. Sci. USA* 101,7034–7039.

Peters, R. H. (1987). *Daphnia* culture. In '*Daphnia*' *Memorie dell'Istituto Italioano di Idrobiologia Dr Marco De Marchi* (eds R. H. Peters and R. de Bernardi), Vol. 45, Consiglio Nazionale Delle Ricerche Instituto Italiono Di Idrobiologia, Verbania Pallanza.

Pfrender, M. E., Spitze, K. and Lehman, N. (2000). Multilocus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Mol. Ecol.* 9,1717–1735.

Pigliucci, M. (2005). Evolution of phenotypic plasticity: Where are we going now? *Trends Ecol. Evol.* 20,481–486.

Pollard, H. G., Colbourne, J. K. and Keller, W. (2003). Reconstruction of centuries-old *Daphnia* communities in a lake recovering from acidification and metal contamination. *Ambio* 32,214–218.

Poynton, H. C., Varshavsky, J. R., Chang, B., Cavigiolio, G., Chan, S., Holman, P. S., Loguinov, A. V., Bauer, D. J., Komachi, K., Theil, E. C., Perkins, E. J., O Hughes, O. and Vulpe, C. D. (2007). *Daphnia magna* exotoxicogenomics provides mechanistic insights into metal toxicity. *Environ. Sci. Technol.* 41,1044–1050.

Regier, J. C. and Shultz, J. W. (1997). Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. *Mol. Biol. Evol.* 14,902–913.

Regier, J. C., Shultz, J. W. and Kambic, R. E. (2005). Pancrustacean phylogeny: Hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proc. R. Soc. Lond. B Biol. Sci.* 272,395–401.

Richard, J. (1895, 1896). Revision des Cladoceres. *Ann. Sci. Nat. Zool. Paleon.* 18,279–389.

Rider, C. V., Gorr, T. A., Olmstead, A. W., Wasilak, B. A. and Leblanc, G. A. (2005). Stress signaling: Coregulation of hemoglobin and male sex determination through a terpenoid signaling pathway in a crustacean. *J. Exp. Biol.* 208,15–23.

Robinson, C. D., Lourido, S., Whelan, S. P., Dudycha, J. L., Lynch, M. and Isern, S. (2006). Viral transgenesis of embryonic cell cultures from the freshwater microcrustacean *Daphnia*. *J. Exp. Zool.* 305A,62–67.

Schierwater, B., Ender, A., Schwenk, K., Spaak, P. and Streit, B. (1994). The evolutionary ecology of *Daphnia*. In *Molecular Ecology and Evolution: Approaches and Applications* (eds B. S. B. Schierwater, G. P. Wagner and R. DeSalle), pp. 495–508, Birkhauser Verlag, Basel, Switzerland.

Schwenk, K., Posada, D. and Hebert, P. D. N. (2000). Molecular systematics of European *Hyalodaphnia*: The role of contemporary hybridization in ancient species. *Proc. R. Soc. Lond. B Biol. Sci* 267,1833–1842.

Shaw, J. R., Colbourne, J. K., Davey, J. C., Glaholt, S. P., Hampton, T. H., Chen, C. Y., Folt, C. L. and Hamilton, J. W. (2007). Gene response profiles for *Daphnia pulex* exposed to cadmium reveal a novel crustacean metallothionein. *BMC Genomics* 8,477.

Shaw, J. R., Dempsey, T. D., Chen, C. Y., Hamilton, J. W. and Folt, C. L. (2006). Comparative toxicity of cadmium, zinc and mixtures of cadmium and zinc to daphnids. *Environ. Toxicol. Chem.* 25,182–189.

Shiga, Y., Sagawa, K., Takai, R., Sakaguchi, H., Yamagata, H. and Hayashi, S. (2006). Transcriptional readthrough of Hox genes Ubx and Antp and their divergent after transcriptional control during crustacean evolution. *Evol. Dev.* 8,407–414.

Soetaert, A., Moens, L. N., Van der Ven, K., van Leemput, K., Naudts, B., Blust, R. and De Coen, W. M. (2006). Molecular impact of propiconazole on *Daphnia magna* using a reproduction-related cDNA array. *Comp. Biochem. Physiol. C* 142,66–76.

Soetaert, A., Van der Ven, K., Moens, L. N., Vandenbrouck, T., van Remortel, P. and De Coen, W. M. (2007a). *Daphnia magna* and ecotoxicogenomics: Gene expression profiles of the antiecdysteroidal fungicide fenarimol using energy-, molting- and life stage-related cDNA libraries. *Chemosphere* 67,60–71.

Soetaert, A., Vandenbrouck, T., van der Ven, K., Maras, M., van Remortel, P., Blust, R. and De Coen, W. J. (2007b). Molecular responses during cadmium-induced stress in *Daphnia magna*: Integration of differential gene expression with higher-level effects. *Aquat. Toxicol.* 83,212–222.

Stein, L. D., Mungall, C., Shu, S. Q., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J. E., Harris, T. W., Arva, A. *et al.* (2002). The generic genome browser: A building block for a model organism system database. *Genome Res.* 12,1599–1610.

Swammerdam, J. (1669). Historia insectorum generalis, of te algemeene verhandeling van de bloedeloose dierkens. t'Utrecht.

Swammerdam, J. (1758). *The Book of Nature; or The History of Insects:* Reduced to Distinct Classes, Confirmed by Particular Instances, Displayed in the Anatomical Analysis of Many Species and Illustrated with Copperplates. John HIll, London.

Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. and Church, G. M. (1999). Systematic determination of genetic network architecture. *Nat. Genet.* 22,281–285.

Tessier, A. J., Leibold, M. A. and Tsao, J. (2000). A fundamental trade-off in resource exploitation by *Daphnia* and consequences to plankton communities. *Ecology* 81, 826–841.

Tollrian, R. (1993). Neckteeth formation in *Daphnia pulex* as an example of continuous phenotypic plasticity – morphological effects of *Chaoborus* kairomone concentration and their quantification. *J. Plankton Res.* 15,1309–1318.

Tollrian, R. (1995). *Chaoborus* crystallinus predation on *Daphnia pulex*: Can induced morphological changes balance effects of body size on vulnerability? *Oecologia* 101,151–155.

U.S.E.P.A. (2002). *Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms*, 5th Edition, US EPA, Office of Water, Washington, DC. 275pp.

Vasemagi, A., Nilsson, J. and Primmer, C. R. (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). *Mol. Biol. Evol.* 22, 1067–1076.

Versteeg, D. J., Stalmans, M., Dyer, S. D. and Janssen, C. (1997). *Ceriodaphnia* and *Daphnia*: A comparison of their sensitivity to xenobiotics and utility as a test species. *Chemosphere* 34,869–892.

Viehoever, A. (1931). Transparent life. *Am. J. Pharm.* 103,252–278.

Viehoever, A. (1936). *Daphnia* – the biological reagent. *J. Am. Pharm. Assoc.* 25,112–117.

Viehoever, A. (1937). The development of *Daphnia magna* for the evaluation of active substances. *Am. J. Pharm.* 109,360–366.

Waring, J. F., Ciurlionis, R., Jolly, R. A., Heindel, M. and Ulrich, R. G. (2001a). Microarray analysis of hepatotoxins in vitro reveals a correlation between gene expression profiles and mechanisms of toxicity. *Toxicol. Lett.* 120,359–368.

Waring, J. F., Jolly, R. A., Ciurlionis, R., Lum, P. Y., Praestgaard, J. T., Morfitt, D. C., Buratto, B., Roberts, C., Schadt, E. and Ulrich, R. G. (2001b). Clustering of

hepatotoxins based on mechanism of toxicity using gene expression profiles. *Toxicol. Appl. Pharm.* 175,28–42.

Warren, E. (1900). On the reaction of *Daphnia magna* to certain changes in its environment. *Q. J. Microsc. Sci.* 43,199–224.

Watanabe, H., Takahashi, E., Nakamura, Y., Oda, S., Tatarazako, N. and Iguchi, T. (2007). Development of a *Daphnia magna* DNA microarray for evaluating the toxicity of environmental chemicals. *Environ. Toxicol. Chem.* 26,669–676.

Watanabe, H., Tatarazako, N., Oda, S., Nishide, H., Uchiyama, I., Morita, M. and Iguchi, T. (2005). Analysis of expressed sequence tags of the water flea *Daphnia magna*. *Genome* 48,606–609.

West-Eberhard, M. (2003). *Developmental Plasticity and Evolution*. Oxford University Press, Oxford.

Whitehead, A. and Crawford, D. L. (2006). Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. USA* 103,5425–5430.

Yampolsky, L. Y. and Galimov, Y. R. (2005). Evolutionary genetics of aging in *Daphnia*. *Zh. Obshch. Biol.* 66,416–424.