Journal of Fish Biology (2008) **72,** 2095–2117
doi:10.1111/j.1095-8649.2008.01875.x, available online at http://www.blackwell-synergy.com

Genomic resources and microarrays for the common carp *Cyprinus carpio* L.

D. R. Williams*, W. Li*†¶, M. A. Hughes*, S. F. Gonzalez‡§, C. Vernon*, M. C. Vidal‖§§, Z. Jeney#, G. Jeney#, P. Dixon**, B. McAndrew††, R. Bartfai‡‡†, L. Orban‡‡|||, V. Trudeau¶¶, J. Rogers##, L. Matthews##, E. J. Fraser*, A. Y. Gracey*** and A. R. Cossins*†††

**Laboratory for Environmental Gene Regulation, University of Liverpool, Crown Street, Liverpool L69 7ZB, U.K., ‡Department of Veterinary Pathobiology, Faculty of Life Sciences, University of Copenhagen. Stigbolen 7, DK 1870 Frederiksberg, Denmark, ‖School of Life Sciences, Huxley Building, Keele University, Staffordshire ST5 5BG, U.K., #Research Institute for Fisheries, Aquaculture and Irrigation, Szarvas 5540, Hungary, **Centre for Environment, Fisheries and Aquaculture, Weymouth Laboratory, Weymouth Dorset DT4 8UB, U.K., ††Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, U.K., ‡‡Reproductive Genomics Group, Temasek Life Sciences Laboratory, National University of Singapore, Singapore 117604, Singapore, |||Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore, ¶¶Centre for Advanced Research in Environmental Genomics, Department of Biology, University of Ottawa, Ottawa, K1N 6N5 Canada, ##Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, U.K. and ***Marine Environmental Biology, University of Southern California, Los Angeles, CA 90089, U.S.A.*

The common carp is an important fish species satisfying ornamental, food and recreational fisheries' needs worldwide, but in common with other cyprinid fishes, it is particularly renowned for its environmental tolerance. Investigating the mechanistic basis of growth, disease and environmental tolerance is greatly enhanced by access to a comprehensive list of gene sequences and post-genomic technologies. The current status of genomic resources is described for this species including 40 k cDNA clone collections, their associated expressed sequence tags (ESTs) and a developing series of 13 k–26 k cDNA microarrays fabricated from amplicons. The arrays have been directed at questions of response to environmental stress (cold and hypoxia), viral and bacterial disease and ectoparasite infection. Consequently, clones from a wide range of tissues were prepared. The authors discuss how these resources were generated and their application. Evidence is presented supporting that the carp microarray may also be useful as a heterologous set of probes in studies of other fish species.

Key words: cDNA; expressed sequence tags; microarray; transcripts.

†††Author to whom correspondence should be addressed. Tel.: +44 (0) 151 795 4510; fax: +44 (0) 151 795 4331; email: cossins@liv.ac.uk

¶Present address: European Bioinformatics Institute, Hinxton, Cambridge, U.K.

§Present address: Immune Disease Institute, Harvard Medical School, 800 Huntington Avenue, Boston, MA 02115, U.S.A.

§§Present address: CICBiogune, Functional Genomics Unit Lab 3, 48160 Derio, Spain.

†Present address: Department of Molecular Biology, Nijmegen Center for Molecular Life Sciences, P. O. Box 9101, 6500 HB Nijmegen, The Netherlands.

INTRODUCTION

The common carp *Cyprinus carpio* L., 1758 has become a widely used model species for environmental research. This is because the carp and other closely related members of the Cyprinidae are very tolerant of environmental stress, particularly including thermal fluctuations and hypoxia (Wendelaar Bonga, 1997; Feder, 2007). Also, in many parts of the world, the common carp is an important food source, which is the object of a substantial aquacultural interest to improve production and protect against disease (Balon, 1995; Horvath & Orban, 1995; Hulata, 1995, 2001). All these interests have prospered from the adoption of conventional hypothesis-driven approaches including the more recent identification of candidate genes, such as the Δ9-desaturase in cold adaptation (Tiku *et al.*, 1996). However, recent developments in post-genomic science now offer broad, open screening approaches to understanding the mechanisms underpinning environmental tolerance, immunity, disease resistance and feed conversion efficiency (Gracey & Cossins, 2003; Cossins & Crawford, 2005). This hypothesis-free or model-free approach depends entirely on the ready availability of large-scale gene resources including cDNA clones as indicators of the relative levels of expression, and also EST and genomic sequence data as the means of establishing gene identity.

The Liverpool carp microarray project was initiated in the year 2000 to explore the transcriptome of multiple tissues responding to chronic cooling and hypoxia regimes both singly and in combination (Gracey *et al.*, 2004; Fraser *et al.*, 2006). To maximize cost efficiencies and universal utility, the construction of a cDNA clone set and the sequence identification of the inserts were undertaken. Amplifying the cloned inserts provided the gene products to print onto a cDNA microarray to enable a broad transcriptomic overview of gene regulatory responses. Subsequently, four other projects contributed batches of cDNA clones generated from specific tissues of fish including those subjected to parasite attack or bacterial–viral challenge. At the same time, the Orban group from Singapore supplied a collection of 7008 clones from differentiating an adult common carp testis. Recently, some of the available cDNA clones have been collated into a less redundant collection for the fabrication of a new version of the carp cDNA microarray, which is now composed of 26 112 probes. Currently, it is estimated that this set of probes consists of *c.* 12 300 unique genes. Here, the origins of each cDNA library and how the latest version of the sequence database for carp and the corresponding cDNA microarray has been constructed are described. The latest production version of the 26 k common carp microarray is available on application to the corresponding author.

MATERIALS AND METHODS

BIOINFORMATIC ANALYSIS OF EST SEQUENCES

For this project, a user-configurable, automated pipeline was created for expressed sequence tags (EST) sequence assembly, identification, annotation and databasing, termed EST-ferret. It was written in the Perl programming language (Baxevanis & Ouellette, 2005) on the Red Hat Linux 7.2 platform, and included the MySQL database management system. Ferret includes all the necessary steps for (i) cleanup and trimming

of ESTs, (ii) submission to external sequence repositories, (iii) sequence clustering, (iv) identification by basic local alignment search tool (BLAST) homology searches and by searches of protein domain databases, (v) annotation with computer-addressable terms and (vi) production of outputs for direct entry into microarray analysis packages. It was constructed from several widely used, open-source algorithms, as listed in Table I, including the CAP3 assembly programme (Huang & Madden, 1999), to yield ESTs containing at least 40 Phred20 base pairs (bps). The typical output from ferret can be inspected at http://www.legr.liv.ac.uk/, choosing carpBASE-Browser for carpBASE 2.1.

The assembly of the ESTs used two consecutive rounds of the CAP3, generating groups corresponding to separate gene families (main groups) and individual genes (sub-groups). The BLAST package (Altschul *et al.*, 1997) was used to search for alignment against a range of sequence databases (in order of Swissprot, RefSeq proteins and the NCBI nr; Table II) to deliver an identity from the top hit using an E-value criterion of $<1e^{-15}$. This identity was used to search the annotation databases, including Gene Ontology (Ashburner *et al.*, 2000), KEGG (http://www.genome.jp/kegg/) and Enzyme Commission (EC; http://www.expasy.org/enzyme/) databases and Conserved Domain Database (CDD), thereby generating a functional annotation of each identified clone. CDD (Marchler-Bauer *et al.*, 2005) and InterPro searches were performed to provide additional protein domain information for each contig. PatSearch (Grillo *et al.*, 2003) and RepeatMasker (http://www.repeatmasker.org) were also used to identify UTR repeats. The entire analysis was run as a single batch process from a configuration file that was set up with the appropriate parameter settings. cDNAs that failed to yield a significant BLASTx result ($<1e^{-15}$) were annotated in carpBASE as unclassified. Finally, EST-ferret was used to generate reports in flat file formats to serve as inputs

TABLE I. Open source programmes used in the construction of the EST-ferret pipeline

Programme	Resources from	Description
Phred	http://www.phrap.org and bge@u.washington.edu	Reads DNA sequencer trace data, calls bases, and assigns quality values to the bases
Cross_match	phg@u.washington.edu	Masks sequences
CAP3	xqhuang@cs.iastate.edu	Assembles sequences and generates consensus sequences (Huang & Madan, 1999)
NCBI BLAST	ftp://ftp.ncbi.nlm.nih.gov	Basic Local Alignment Search Tool for similarity search
GOprofiler	http://legr.liv.ac.uk/	Assigns GO annotations for sequences according to BLAST best hits against Swiss-Prot
ECprofiler	http://legr.liv.ac.uk/	Scans enzyme database by using BLAST best hits against Swiss-Prot
PatSearch	ftp://www.pesolelab.it/	Finds UTR patterns from UTRsite
RepeatMasker	nilah@geospiza.com or http://www.geospiza.com/	Masks repeats from RepBase
Perl	http://www.perl.org/	A scripting language
MySQL	http://www.mysql.com	A SQL Database Management System
Apache HTTP Server	http://www.apache.org	A web server software
PHP	http://www.php.net	The Hypertext Pre-processor

TABLE II. Databases used by EST-ferret for the assembly, BLAST alignment and annotation of carp ESTs

Database	Resource from	Description and reference
Common carp mitochondrial genome	http://www.ncbi.nlm.nih.gov/entrez/viewer.fcg?db=nucleotide&val=5835023	GenBank accession number NC_001606
1027 known carp protein sequences	Retrieved *via* NCBI Entrez	Known common carp sequences available in GenBank
UniProtKB–Swiss-Prot protein	ftp://us.expasy.org/databases/uniprot/knowledgebase/uniprot_sprot.fasta.gz	A well-curated protein sequence database
RefSeq Vertebrate: for mammalian and other species	ftp://ftp.ncbi.nih.gov/refseq/release/vertebrate_mammalian and ftp://ftp.ncbi.nih.gov/refseq/release/vertebrate_other/	Through heavily manual curation on know proteins and NCBI Genome Annotation Projects
Nr	ftp://ftp.ncbi.nlm.nih.gov/blast/db/nr.tar.gz	Non-redundant protein collections in NCBI
Zebrafish protein	ftp://ftp.ncbi.nih.gov/refseq/D_rerio/mRNA_Prot/zebrafish.protein.faa.gz	Zebrafish protein sequences generated through NCBI RefSeq and NCBI Genome Annotation projects
1902 fugu known protein sequences	Retrieved *via* NCBI Entrez	
Human protein	ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens/protein/protein.fa.gz	Human protein sequences generated through NCBI RefSeq and NCBI Genome Annotation projects
Mouse protein	ftp://ftp.ncbi.nlm.nih.gov/genomes/M_musculus/protein/protein.fa.gz	Mouse protein sequences generated through NCBI RefSeq and NCBI Genome Annotation projects
Rat protein	ftp://ftp.ncbi.nih.gov/genomes/R_norvegicus/protein/protein.fa.gz	Rat protein sequences generated through NCBI RefSeq and NCBI Genome Annotation projects
FlyBase:	ftp://flybase.net/genomes/Drosophila_melanogaster/current/fasta/dmel-all-translation-r4.2.1.fasta	Fly proteins sequences from FlyBase project
Worm protein	ftp://ftp.wormbase.org/pub/wormbase/acedb/WS147/wormpep147.tar.gz	Worm protein translations of all predicted and confirmed genes
Yeast protein	ftp://genome-ftp.stanford.edu/pub/yeast/data_download/sequence/genomic_sequence/orf_protein/orf_trans_all.fasta.gz	Yeast translations of all systematically named ORFs

TABLE II. Continued

Database	Resource from	Description and reference
Zebrafish gene collection (ZGC)	ftp://ftp1.nci.nih.gov/pub/MGC/fasta/dr_mgc_cds_aa.fasta.gz	Zebrafish full-length cDNA collection
Zebrafish WZ contigs	http://www.genetics.wustl.edu/fish_lab/assemblies/wzcontigs.gz	Washington University zebrafish EST assembly
Fugu EST assembly	http://fugu.biology.qmul.ac.uk/Download/	HGMP Fugu EST assembly
UTRdb	ftp://bighost.ba.itb.cnr.it/pub/Embnet/Database/UTR	Contains UTR sequences
UTRsite	ftp://bighost.ba.itb.cnr.it/pub/Embnet/Database/UTR/UTRSite	Contains UTR patterns
RepBase	http://www.girinst.org	Updated repeat elements
CDD (Pfam, Smart, Kog and Cog)	ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/	A conserved domain database and search service
Gene Ontology data	http://www.geneontology.org/GO.current.annotations.shtml and http://www.godatabase.org/dev/database/	GO annotations include molecular function, cellular component & biological process
Enzyme database	ftp://ca.expasy.org/databases/enzyme/release/enzyme.dat	Includes descriptions for all known enzymes
KEGG enzyme search	http://www.genome.ad.jp/kegg-bin/mk_point_html and http://www.genome.ad.jp/kegg/kegg2.html	A biochemical pathway database

for gene annotation and gene expression profiling tools, and also the carpBASE series of MySQL web-browsable databases (Table V).

PHASE I LIBRARY PRODUCTION

Stocks of common carp were obtained from fish farms and were maintained in 2000 l tanks at 30° C for at least 4 weeks. Some individuals were removed to separate 1000 l tanks where they were subjected to staged cooling and hypoxia regimes as described previously (Gracey *et al.*, 2004).

RNA isolation and first-strand cDNA synthesis

Total RNA was extracted with TRIzol (Invitrogen, Madison, WI, U.S.A.), and then poly(A) + RNA was purified over oligo-dT resin (Clontech, Mountain View, CA, U.S.A.), following the manufacturer's instructions. Adaptors containing the rare asymmetrical restriction sites for *Sfi*I were incorporated into the cDNA using a template-switching mechanism at the 5′ end of RNA transcript (SMART; Clontech, originally described as Capfinder) (Zhu *et al.*, 2001; Wellenreuther *et al.*, 2004), enabling directional cloning of the cDNA into the pTRiplEx2 vector *Sfi*I sites. Different 5′ adaptors were used for each tissue (see underlined sequence tag below) to allow tissue source to be identified by sequence. For SMART PCR amplification of first-strand cDNA, the 5′ SMART PCR primer was used in conjunction with the same 3′ adaptor that was used for cDNA synthesis [Table III(a)].

Normalization

All libraries were normalized using the same pool of poly(A) + RNA that was used for first-strand synthesis. Normalization was performed essentially as described (Carninci

TABLE III. Primers used for cloning of carp cDNAs

(a) SMART PCR amplification in phase I cDNA cloning

3′ adaptor (first-strand cDNA synthesis and SMART PCR)
5′-ATTCTAGAGGCCGAGGCGGCCGACATGTTTTTTTTTTTTTTTVN-3′

5′ tissue-specific adaptors (first-strand cDNA synthesis):
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCGGG-3′
Liver (no tag)
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCCGCGGG-3′ Kidney
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCAGCGGG-3′
Intestine
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCTGCGGG-3′ Gill
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCGCCGGG-3′ Muscle
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCTCCGGG-3′ Brain
5′-AAGCAGTGGTATCAACGCAGAGTGGCCATTATGGCCACCGGG-3′ Heart

5′ SMART PCR primer
5′-AAGCAGTGGTATCAACGCAGAGT-3′

(b) SMART PCR amplification in phase II cDNA cloning
CDS-3M adapter (first strand cDNA synthesis) 5′-AAGCAGTGGTATCAACGCA GAGTGGCCGAGGCGGCC$(T)_{20}$VN-3′; N = A,C,G or T; V = A,G or C
SMART IV oligo (first-strand cDNA synthesis) – 5′-AAGCAGTGGTATCAACGCA GAGTGGCCATTATGGCCGGG-3′
SMART PCR Primer – 5′-AAGCAGTGGTATCAACGCAGAGT-3′

et al., 2000), except that the first-strand cDNA was prepared with template switching (Gracey *et al.*, 2004). Second-strand synthesis of the normalized cDNA was performed using long-distance SMART PCR with five to eight cycles. Alternatively, the normalized first-strand cDNA was subjected to a round of subtraction.

Subtraction

To synthesize subtracting driver, 384 well plates of picked cDNA clones were plated on LB-agarose, colonies were scraped and plasmid DNA obtained using a Nucleospin Plasmid DNA extraction kit (Macherey-Nagel, Düren, Germany). Using the plasmid DNA as a template, the cDNA inserts were amplified by long-distance PCR, using primers that flank their cloning site. The PCR products were digested with *Hae*III and then labelled with biotin (PanVera LabelIT; Invitrogen). Normalized first-strand cDNA from single or pooled tissues was then subtracted (CoT $>$ 100) against the biotin-labelled PCR products using the same hybridization protocol as used for normalization (Carninci *et al.*, 2000), and the resulting subtracted cDNA was converted to double-stranded cDNA using long-distance PCR, with limited cycles (five to eight cycles).

SSH capture

A small number of libraries were prepared from the capture of full-length equivalent of suppressive subtractive hybridization (SSH) libraries that were enriched for cold-responsive and hypoxia-responsive genes. Details of the full-length capture and the SSH technology are described in detail elsewhere (Diatchenko *et al.*, 1996, 1999; Gracey *et al.*, 2001). To prepare an actual cDNA library of bacterial clones containing amplified carp cDNA inserts, each source of cDNA was digested with *Sfi*I restriction enzyme and size-fractionated by gel excision or using CHROMA SPIN 400 (Clontech) size-exclusion spin columns which favour inserts $>$600 bps but excludes small inserts ($<$100 bps) from the resulting libraries.

Cloning

All SMART PCR amplified cDNAs were digested with *Sfi*I restriction site-specific enzyme and size-fractionated by agarose gel electrophoresis. The cDNA was ligated into pTriplEx2 plasmid (Clontech) and transformed into DH10B electro-competent *Escherischia coli* cells (Invitrogen). The transformants were plated onto LB-Amp plates containing Xgal and IPTG for blue–white screening. Individual clones were transferred in 384 well-plate positions containing LB-broth in 10% (v/v) glycerol with ampicillin.

Removal of redundant clones

Based on microarray expression data and sequencing of all phase I libraries, redundant clones were removed, allowing additional spotting capacity. Clones were removed based on the criteria of lacking a BLAST identity and either no signal on the microarray or no interesting changes in expression after experimental treatment. An MG2 microarray robot (Genomics Solutions, Huntingdon, U.K.) was used to re-array the selected clones into new 384 well plates in a process termed as 'cherry picking'. The phase I microarray was cherry picked back to 6144 clones from 13 824 clones which equated to a reduction to 16 $\times$ 384 well plates. Table IV gives an overview of the contributing tissues and treatment types used for the construction of these libraries.

PHASE II LIBRARY PRODUCTION

Source of clones and libraries

The new cherry-picked collection from the first phase libraries was supplemented by the contribution of individual carp tissue libraries. These were produced by researchers

TABLE IV. The source of tissues used for construction of cDNA libraries and clone sets for the common carp processed in Liverpool

	Treatment type	Tissues sampled (libraries)	Number of clones produced (GenBank accession numbers)	Plate numbers (series A)	Plate numbers (series B)	Origin of tissue samples
Phase Ia	Cold, acute and chronic	G, B, H, M, L, I, HK (13)	13,824 (CA963982–CA970467; CF660356–CF663121)	1–40	1–16	Liverpool, U.K.
Phase Ib	Hypoxia, acute and chronic	M, H, K, B, I, G (1)				Liverpool, U.K.
Phase IIa	Cold	B (3)	3456 (EX821642–EX824491)	51–55, 58–61	17–27	Liverpool, U.K.
Phase IIb	Apoptosis during infection	HK (1)	3840 (EX824492–EX827081)	62–71		Keele, U.K.
Phase IIc	Ectoparasite infection and immune responses	Skin (2)	4608 (EC391458–EC394905)	72–83		Copenhagen, Denmark
Phase IIIa	Bacterial challenge	H, G, Sp (3)	7680 (EX879026–EX886245)		28–47	Szarvas, Hungary
Phase IIIb	Viral challenge	HK, G (2)				Weymouth, U.K.

Tissues: G, gill; B, brain; H, heart; M, epaxial skeletal muscle; I, intestinal mucosal scrapings; L, liver; HK, head kidney; Sp, spleen. Series B plates for phase I and II were cherry-picked from the corresponding series A plates. These plates were augmented by clones from phase III, with 384 well microtitre plates numbered consecutively.

to address specific scientific questions and thus were directed at the tissue relevant to that question (*e.g.* skin, head kidney and brain). Libraries were prepared again as directional full-length clones but were not subjected to normalization and subtraction. RNA was converted into first-strand cDNA according to the instructions of Creator SMART cDNA Library Construction kit (Clontech) except that the cDNA synthesis primers were modified to allow the use of a single primer for SMART PCR. Unlike phase I libraries, the 5′ end of both primers used in cDNA synthesis were identical allowing a single primer to be used during cDNA amplification [Table III(b)]. The single primer was sequence specific to each end of the first-strand cDNA synthesis primers, whereas previously the oligo-dT cDNA synthesis primer was also employed as a PCR primer which is likely to be less sequence specific.

For these libraries, size-fractionation using CHROMA SPIN 400 spin columns, *Sfi*I digested SMART cDNA was cloned into commercially prepared *Sfi*I predigested pDNRLib vector (Clontech). This vector does not offer blue–white screening but instead contains a sucrase gene in the multiple cloning site to prevent the growth of vector recombination clones. These ligations were transformed into DH10B electro-competent *E. coli* and the resulting clones collected into 384 well plates with LB-broth in 10% (v/v) glycerol and chloramphenicol.

Cherry picking of phase II clones

The new clones from skin, brain and head kidney added at phase II were cherry-picked from 33 × 384 well plates and reduced to 11 plates. Thus, the BLAST identities for these sequences on carpBASE (http://www.legr.liv.ac.uk/) were examined and sequences with the same top hit were reduced in number, usually not more than five clones. As expected, the redundant clones consisted of primarily ribosomal rRNA subunit transcripts and surprisingly a large number of orphan nuclear receptor NR5A2 transcripts. Those clones without a high BLAST score termed as 'unclassifiable EST' were subjected to BLASTN and BLASTX at a lower stringency of $<e^{-5}$. Any sequence that did not have a BLAST ID after this was rejected from the new cherry-picked list. Expression data were not used as a factor for exclusion. The rejected sequences primarily contained very short inserts or the vector's stuffer fragment used in the original commercial vector construction. The relationship between the plate location of a clone in the original plate and in the cherry-picked collection is available at http://legr.liv.ac.uk/carpbase/carpbase_5_0/index.htm in the form of an Excel file.

PHASE III LIBRARY PRODUCTION

Singapore cDNA clone collection

These libraries were generated in the Orban (Singapore) laboratory as described by Christoffels *et al.* (2006). Briefly, full-length cDNA was synthesized from testis of 59–60, 80 and 100 days-old individuals using Creator SMART Library Construction kit (Clontech) according to the manufacturer's instruction. After *Sfi*I restriction enzyme digestion, the adaptors and short cDNAs were removed by CHROMA SPIN 400 spin columns (Clontech). The size-fractionated cDNA pool was then cloned into a pBluescript-based vector with a modified MCS site containing *Sfi*I sites (Christoffels *et al.*, 2006) and transformed into *E. coli* XL10-Gold cells. Clones were picked into thirty, twenty and ten 96 well plates from the libraries generated from testes collected at 60, 80 and 100 days post-fertilization (dpf), respectively, and their inserts were sequenced using M13 forward primer.

Two SSH libraries using RNA isolated from the testis of 70 day and 100 day individuals were also produced. Two sets of subtractive hybridizations were performed: 70 dpf male gonad (driver) from 100 dpf testis (tester) and 100 dpf testis (driver) from 70 dpf male gonad (tester). The PCR-Select cDNA Subtraction kit (Clontech) was used to enrich developmental stage-specific fragments from the SMART cDNA template according to the recommendations of the manufacturer. The selectively amplified cDNA

fragments (on average 400–800 bp in length) were ligated into pGEM-T (Promega, Madison, WI, U.S.A.) cloning vector and transformed into *E. coli* XL10-Gold cells. Two thousand five hundred clones were picked in 96 well plates and inserts sequenced using M13 forward primer.

The libraries were transferred to Liverpool as frozen bacterial stocks in 96 well plates. The EST sequences were compared with the existing dataset using EST-ferret, and clones missing from the Liverpool collection were added to carpBASE v5. This collection of 7008 clones was then incorporated into the microarray pipeline by reformatting into 384 well plates in preparation for PCR amplification and microarray processing.

Eurocarp cDNA clone collection

Most recently cDNA clones have been generated from head kidney, spleen and gill from carp exposed to bacterial and viral pathogenesis. These libraries were constructed using the Creator SMART cDNA Library Construction kit and again with the cDNA synthesis primers modified to allow subsequent single primer PCR amplification (see phase II). A total of 7680 common carp cDNA clones from this source has been added to v5 of the microarray from this collection. Clones were sequenced from the 5′ end of the cDNA using M13 forward sequencing primer.

AURATUS.CA cDNA clone collection

Initial partial cDNA sequences from goldfish *Carassius auratus* (L., 1758) were obtained by two strategies: traditional homology cloning and subtractive suppressive hybridization (Martyniuk *et al.*, 2006). Genes involved in the synthesis or degradation of key neuropeptides, neurotransmitters, hormone, peptides and hormone receptors in goldfish brain, pituitary and gonad were targeted. To produce the brain-enriched cDNA library, cDNA derived from muscle was subtracted from a mixed sample of cDNAs

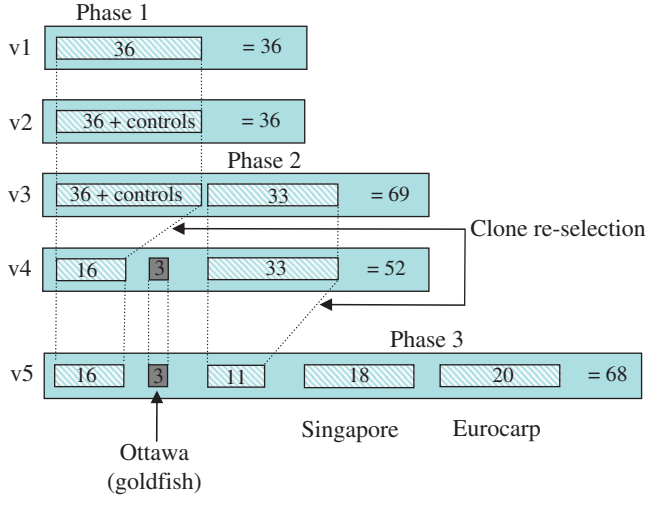

FIG. 1. Plates contributing to the various builds of carpArray versions (v) 1–5 and their clone library sources. The numbers refer to the number of 384 well plates in each collection, as indicated in Table I. Phase I and II plates were cherry picked as indicated to a smaller number of plates, allowing addition of clones from other collections. v4 and v5 include clones from Ottawa, Singapore and from the EU Eurocarp project.

obtained from the telencephalon and hypothalamus using the PCR-Select Subtraction kit (Clontech). About 1152 goldfish cDNAs were arrayed with 18 816 carp cDNAs to obtain the first generation of the mixed-species arrays (Fig. 1).

STORAGE AND CURATION OF cDNA CLONES

Clones were stored in LB-broth containing 10% (v/v) glycerol and appropriate antibiotic in 384 well microtitre plates at −80° C. Replica plates were produced and were stored in separate freezers.

SEQUENCE CHARACTERIZATION OF cDNA CLONE SETS

Each clone was subjected to single-pass Sanger sequencing from the 5′ end. The resulting sequences were assembled into a minimally redundant set of contigs, and their identities were assessed by BLAST against sequence databases.

Although the sequence analysis and annotation had already been performed by the Orban group on their testis libraries, these sequences have also been processed through the EST-ferret pipeline and these results have been included within carpBASE v5 (http://www.legr.liv.ac.uk/carpBASE).

MICROARRAY FABRICATION

PCR products for each cDNA clone were generated to support microarray printing. Insert sizes ranged between 0·3 and 2 kb with the vast majority being *c.* 1 kb. The PCR master mix was designed with view to contribute to the microarray print solution. Thus, commercial PCR buffers containing detergents were avoided and a Tricine PCR buffer was routinely used (300 mM Tricine pH 8·5, 500 mM KCl, 20 mM $MgCl_2$) as a standard alternative. The master mix (30 ml) sufficient to charge 4 × 96 well microtitre plates (70 μl per well or two 384 well plates at 30 μl per well) contained 3 ml Tricine 10× PCR buffer, 60 μl vector forward primer (100 μM), 60 μl vector reverse primer (100μM), 60 μl dATP (100 mM), 60 μl dCTP (100 mM), 60 μl dGTP (100 mM), 60 μl dTTP (100 mM), 100 μl Taq polymerase (5U/μl; Bioline) and 26 667 μl PCR grade water. Control features were applied using the Lucidea Universal Scorecard kit (GE Healthcare, Chalfont St Giles, U.K.).

While cDNA clones were stored in 384 well format, the PCR reactions were performed in 96 well plates. The transfer of clones between formats was achieved using a 96 well microplate replicator (Boekel, Feasterville, PA, U.S.A.). The 96 well PCR reaction was robotically re-arrayed into the 384 well format to serve as PCR stock plates. More recently, lower volume reactions were performed in 384 well PCR plates to avoid the need to reformat the PCR products after PCR. An aliquot from each well was then directly transferred to the 384 well plate used by the microarray robot, which was termed the 'print plate'. The print plate contains the spotting buffer that was mixed with the PCR product to give the appropriate final concentration of buffer and PCR product.

Microarray versions 1–4 were printed using a contact printer using 48 pins (MG2; Genomic Solutions, Huntingdon, U.K.). Version 1 was printed on slides treated with poly L-lysine, while version 2–5 were printed on GAPS II coated slides (Corning, Lowell, MA, U.S.A.). The v 5 microarray was printed using a non-contact piezo-deposition printer (ArrayJet, Dalkieth, U.K.) fitted with a low volume head unit and 48 plate changer. For printing, the dried amplicons were resuspended in 50% glycerol, 0·05% Triton-X100 with overnight rehydration to give a final DNA concentration of 100–200 ng μl^{-1}. The change from contact to non-contact printing considerably improved spot uniformity.

RESULTS

LIBRARIES AND CLONE SETS

Directionally cloned, full-length cDNA libraries were generated in three separate production phases for a series of tissues and treatment conditions (Fig. 1 and Table IV). To ensure that genes expressed only under specific conditions were included, tissues taken from common carp were used that between them had experienced a wide range of treatments: cold, hypoxia, bacterial and viral infection, parasite attack, all in acute and chronic phases. At the time of library construction for phase I clones, the cost of sequencing was significantly higher than it is today. To reduce these costs in production and characterization of phase I libraries, the redundancy was minimized and gene representation maximized by subjecting the libraries to normalization (reducing the frequency of abundant genes while increasing the representation of rare genes) followed by iterative rounds of subtraction (Carninci *et al.*, 2000; Gracey *et al.*, 2004). This was done using labelled probes from already collected clones with redundant inserts to capture clones by hybridization from new cDNA pools. This resulted in a collection of reduced redundancy relative to many other collections. To this, subsequent clone collections were added including those submitted by the Singapore and Ottawa groups which together comprised a collection of 26 202 clones from 33 different libraries generated from 13 different tissues and six different treatments. This collection, which was used in construction of carpARRAY v5 comprised 2766 contigs each represented by multiple probes and 9521 singletons, representing a possible 12 287 genes.

carpBASE

Successive versions of carpBASE [*i.e.* versions (v) 1–3; Table V] included increasing numbers of clones. carpBASE v5 was specifically constructed to assist with the analysis and interpretation of the newest carp microarray (version 5) which included the clones listed above. This microarray included large

Table V. The developing carpBASE series of databases of carp cDNA ESTs

Database version	Clone plates	Number of ESTs	Date produced	Analysed using
carpBASE 1	*A1-30,32-35,39-40	9202	January 2002	EST-ferret 1.1
carpBASE 2	A1-30,32-35,39-40	9202	June 2004	EST-ferret 2.0/1
carpBASE 3	A1-30,32-35,39-40, A51-55, 58-83, C1-3	18 104	September 2005	EST-ferret 2.1
carpBASE 5	*B1-16,B17-27, *D1-19, *E1-20	22 661	April 2007	EST-ferret 2.2

*A and B refer to the original and reduced cDNA collections described in Table I. C refers to 1152 goldfish (Ottawa), D refers to 7008 testis carp cDNA clones (Singapore) and E refers to 7680 carp cDNA clones (Eurocarp project).

clone collections that had been reduced by cherry picking single clones from redundant entries (*in silico* normalization) to reduce redundancy and increase gene representation. carpBASE v5 is a collation only of the clones represented on the microarray rather than all known carp clones. As a consequence, a full interrogation of gene coverage and identity is best achieved by exploring both carpBASE versions 3 and 5. The link http://legr.liv.ac.uk/carpbase/carpbase_5_0/index.htm provides the key to relate clones from series A (original plates) to series B plates (cherry-picked plates).

EST-ferret was designed to provide a two-stage clustering of sequences to identify 'main-groups' (gene families) with a 60% identity cut-off criterion, and sub-groups with a 90% criterion (genes or isoforms). The entry for each contig includes a list of all clones that were used in its construction, with an indication of contig length. The FASTA versions of each clone and contig are provided. The top BLAST hits against a series of different databases were displayed to allow the consistency of returned identities to be gauged, and links to the full list of hits are also provided. Searches were also conducted of the (i) translated sequences against the conserved domain database (CDD), (ii) the untranslated nucleotide sequence against UTR and repeat info databases, all which furnishes additional functional annotation to assist the interpretation of clones lacking a conventional BLAST identity. Where an identity was generated, then each contig/singleton was categorized within the gene ontology (GO) nomenclature, using terms in each of the 'biological process', 'cellular compartment' and molecular function' domains. Table VI indicates the distribution of sequences within the biological process domain and Table VII indicates the distribution of identified genes across the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways categories, and shows a coverage of metabolic pathways that in many cases matches that of the more extensive zebrafish *Danio rerio* (Hamilton, 1822) EST collection. Both the tables indicate a broad representation of genes within the major functional categories, which in many cases matches the representation of genes in the much larger mouse cDNA collection. Genes within the Enzyme Classification (http://www.expasy.org/enzyme/) and KEGG (Kanehisa *et al.*, 2006) annotation schemes were also categorized, as appropriate. The resulting identities were used in the construction of the carpBASE series of databases using the EST-ferret informatics pipeline. Each version corresponds to a version of the carp microarray as indicated in Table VIII.

To assess the degree of coverage of genes within carpBASE v5, Table VIII compares the number of identified genes in carpBASE in each Enzyme Classification category with that for zebrafish. The latter species has *c.* 20-fold greater number of ESTs than common carp, and for which there is more confidence that the gene collection is closer to completion. Yet for many of the categories, the number of genes was similar, including, for example, glycolysis–glycogenolysis, the triarboxylic acid (TCA) cycle and pentose-phosphate shunt. For some other pathways (alanine–aspartate metabolism, *etc.*), the common carp collection had up to twice the number in zebrafish, which may relate to the proposed recent whole genome duplication event (David *et al.*, 2003). But for some pathways, gene representation was much lower in the common carp, which probably represents the limitations of the common carp EST collection.

TABLE VI. The distribution of carp genes from carpBASE v5 within the gene ontology nomenclature

Biological process	4211
Cell communication	616
Cell–cell signalling	62
Cell adhesion	167
Cell cycle	227
Cell growth	39
Cell motility	66
Cell organization and biogenesis	634
Cell proliferation	110
Response to endogenous stimulus	77
Response to external stimulus	128
Signal transduction	577
Metabolism	2833
Amine metabolism	160
Amino acid and derivative metabolism	141
Biosynthesis	697
Carbohydrate metabolism	227
Catabolism	282
Electron transport	180
Energy pathways	341
Lipid metabolism	249
Nitrogen metabolism	167
Nucleobase, nucleoside, nucleotide and nucleic acid metabolism	937
Organic acid metabolism	269
Phosphorus metabolism	271
Protein metabolism	1362
Regulation of metabolism	610
Regulation of cellular metabolism	569
Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism	513
Transcription	513
Regulation of transcription	497
Regulation of transcription, DNA-dependent	469
Transcription, DNA-dependent	478
Stress response	328
Transport	999
Cation transport	127
Electron transport	180
Intracellular protein transport	138
Intracellular transport	265
Ion transport	193
Protein transport	242
Secretory pathway	80
Vesicle-mediated transport	137
Development	418
Cell differentiation	160
Morphogenesis	154
Organ development	128

TABLE VI. Continued

Regulation of development	32
System development	97
Developmental process	418
Physiological process	3845
Cellular physiological process	3599
Localization	1073
Organismal physiological process	434
Regulation of physiological process	851
Regulation of biological process	943
Regulation of catalytic activity	70
Regulation of cellular process	882
Regulation of cellular physiological process	825
Regulation of signal transduction	93
Regulation of physiological process	851
Negative regulation of physiological process	228
Positive regulation of physiological process	143
Regulation of metabolism	610
Reproduction	61
Response to stimulus	531
Response to abiotic stimulus	124
Response to biotic stimulus	249
Response to chemical stimulus	114
Response to oxidative stress	22
Response to external stimulus	128
Immune response	222
Response to stress	328
Response to unfolded protein	32
Response to wounding	108
Response to endogenous stimulus	77
Response to DNA damage stimulus	68
Response to pest, pathogen or parasite	145
Sensory perception	51

CARP MICROARRAY

A series of customized common carp microarrays have been fabricated (Table VIII) in which the authors have (i) added increased numbers of clones, (ii) decreased the redundancy of probes and thus increased the representation of genes, initially using normalization and subtraction techniques, but subsequently by cherry-picking, (iii) selected representative clones from multi-clone clusters identified through assembly techniques in EST-ferret, (iv) moved from in-house slide chemistry in early versions to commercial sources of slides in later versions and (v) included external standards. The recent transition from contact-printed microarrays (versions 1–4) to non-contact, piezo-printed (ink-jet) microarrays has generated more consistent spot shapes and dimensions and more consistent spot locations.

TABLE VII. Comparison of gene representation in pathways of metabolism between the zebrafish and carp gene collections. Data for zebrafish were obtained from <http://www.genome.jp/kegg-bin/mk_point_html> and that for carp were from carpBASE v5. The right hand column expresses the representation for carp as a percentage of that defined for the more extensively sequenced and annotated zebrafish EST collection

KEGG pathway categories	carpBASE5: number of enzymes	Zebrafish: number of enzymes	Ratio of carpBASE5/ zebrafish (%)
Carbohydrate metabolism			
Glycolysis–gluconeogenesis	22	21	105
Citrate cycle (TCA cycle)	12	10	120
Pentose phosphate pathway	10	10	100
Pentose and glucuronate interconversions	2	4	50
Fructose and mannose metabolism	11	7	157
Galactose metabolism	7	4	175
Ascorbate and aldarate metabolism	2	3	67
Starch and sucrose metabolism	12	10	120
Aminosugars metabolism	7	10	70
Inositol phosphate metabolism	8	10	80
Pyruvate metabolism	14	12	117
Glyoxylate and dicarboxylate metabolism	5	9	56
Nucleotide sugars metabolism	2	3	67
Energy metabolism			
Oxidative phosphorylation	6	6	100
Methane metabolism	3	3	100
Carbon fixation	10	11	91
Nitrogen metabolism	6	5	120
Lipid metabolism			
Fatty acid metabolism	12	14	86
Biosynthesis of steroids	8	10	80
Bile acid biosynthesis	6	6	100
Androgen and oestrogen metabolism	6	11	55
Urea cycle	6	10	60
and metabolism of amino groups			
Glycerolipid metabolism	8	9	89
Nucleotide metabolism			
Purine metabolism	25	25	100
Pyrimidine metabolism	16	18	89
Amino acid metabolism			
Glutamate metabolism	11	11	100
Alanine and aspartate metabolism	16	8	200
Glycine, serine	17	11	155
and threonine metabolism			
Valine, leucine	17	15	113
and isoleucine degradation			
Arginine and proline metabolism	10	21	105
Histidine metabolism	7	10	120
Tyrosine metabolism	7	10	100

TABLE VII. Continued

KEGG pathway categories	carpBASE5: number of enzymes	Zebrafish: number of enzymes	Ratio of carpBASE5/ zebrafish (%)
Benzoate degradation *via* hydroxylation	1	4	50
Tryptophan metabolism	11	7	157

DISCUSSION

The carpARRAY project has evolved over the past 7 years as new cDNA clones became available and as microarray fabrication technology matured. The research project that initiated the generation of clone sets, EST collections and microarrays was directed at environmental stress and multiple tissues, but subsequent applications to disease and parasites were more directed at specific tissues, particularly including those involved in conferring disease resistance or responses to parasite infection (spleen, head kidney or affected tissues such as skin). Throughout this work, most libraries were generated from tissues that had been exposed to a range of treatments, and over the full time-course of exposures, in order to ensure that transcripts that were rare in one condition or time, but which was more abundant in another condition or time, would still be included in the collection. Early work generated significant redundancy in the clone collections.

While in principle only one representative amplicon probe is required for each expressed gene, in practice it is more difficult to ensure that all isoforms are included rather than being discarded through an incomplete understanding of gene diversity. The common carp is widely thought to have experienced a whole genome duplication event some 12–15 million years ago (David *et al.*, 2003), and thus is expected to have approximately double the number of copies of many genes compared to unduplicated relatives, such as the grass carp (Orban & Wu, 2008). So, in early versions of the carp microarray, all available clones were included and this allowed the presence of isoforms to be explored using the expression data to discriminate between genes rather than just the sequence data. Indeed, refined clustering analysis has demonstrated that cDNA probes are sufficient to separate the properties of closely related isoforms, such as aldolase A and aldolase B, and parvalbumin paralogues (unpubl. obs.). However, as the number of cDNA clones accumulated and exceeded the capacity of the printing platform, it became necessary to reduce the number of clones represented, the selection being directed at clones lacking a BLAST-defined identity together with uninteresting expression properties determined across >1100 high-quality microarrays. Some of these clones may represent cloning artefacts or concatenated products that do not correspond to transcriptional products.

The most recent version, carpARRAY v5, thus includes representative clones of all assembled carp contigs from the contributing laboratories. It also includes representation of expression variants or probes with interesting expression profiles but which did not possess a BLAST-defined identity. Since, for

TABLE VIII. List of carp cDNA microarray versions produced from 2001 to 2007

Version numbers (date of first production, relevant database)	Plate numbers	Slide source chemistry	Control plates	Number of 384 well plates (probes)
carpARRAY v1 (August 2003, carpBASE 1)	*A1-30, 32–35, 39–40	Poly L-lysine, in-house		36 (13 824)
carpARRAY v2 (September 2005, carpBASE 2)	A1-30, 32–35, 39–40, control plate	Corning GAPS II	Home-made	36 (13 824)
carpARRAY v3 (December 2005, carpBASE 3)	A1-30,32–35,39–40, A51-55, 58-83 C1-3	Corning GAPS II	Lucidea	69 (26 496)
carpARRAY v4 (April 2006)	*B1-16, A51-55, 58-83, *C1-3	Corning GAPS II	Lucidea	50 (19 200)
carpARRAY v5 (October 2007, carpBASE 5)	*B1-16, B17-27, C1-3, *D1-19, *E1-20	Corning GAPS II	Lucidea	68 (26 208)

*A and B refer to the original and reduced cDNA collections described in Table I. C refers to 1152 goldfish (Ottawa), D refers to 7008 testis carp cDNA clones (Singapore) and E refers to 7680 carp cDNA clones (EU Eurocarp project). Lucidea refers to the set of control RNAs commercialized by GE Healthcare.

later versions of carpARRAY, the number of available clones exceeded the capacity of the array, the authors chose to exclude clones that were both unidentified and on the basis of initial array analysis displayed no interesting expression properties. However, under other circumstances, these clones may well be relevant to studies of other tissues or treatments and these can be introduced into later array versions, if desired, simply by substituting them for other clones. Alternatively, increased numbers of features can be printed first by reducing feature size and second by adopting hexagonal packing of spots. Both options are available with the ArrayJet robotic printer.

Sequence data generated from the common carp has informed the wider analysis of fish genes, notably including those of the zebrafish. Thus, Orban and colleagues (Christoffels *et al.*, 2006) have generated and collated 6050 ESTs from carp testis with 9303 non-gonadal ESTs taken from carpBASE and other sequences from GenBank. Of these 15 353 clones, 974 failed to match any other known sequence in zebrafish or any fish species found in dbEST (http://www.ncbi.nlm.nih.gov/dbEST/). They mapped 1843 of these 15 353 carp sequences to the zebrafish genome, of which 1752 matched coding sequences, 91 mapped to intra-genic or intronic regions and 412 represent new splicing variants. Finally, they showed that there was a significant degree of sequence conservation in the 5′ UTR sequences between carp and zebrafish. This study indicates the similarity at the nucleotide between the two species, and this facilitates the linked use of common carp for biochemistry and physiological characterization with the genetic tractability of zebrafish.

The microarrays and databases described here have proved to be the critical elements in several other major physiological studies. Gracey *et al.* (2004) have explored responses of seven major tissues to a chronic cooling regimen. They found that >3000 genes showed significant responses, of which *c.* 1700 possessed a BLAST identity. The scale of this response was surprising, and exceeded expectations based on previous single-gene or small-scale studies, which indicates the need to understand the distributed, system-wide nature of cellular responses to treatment. These *c.* 1700 genes were clustered into 24 groups, one of which consisted of a common-response gene group mainly linked to cell homeostasis functions and which were >90% up-regulated by cold in all tissues. Figure 2 illustrates one part of this group with a 'heat-map' of coherent up-regulated responses of genes involved in ATP synthesis. Other clusters showed highly tissue-specific patterns of cold-induced expression, the GO analysis of which identified the coherent regulation of 'carbohydrate metabolism' and 'energy pathways' categories. They also demonstrated that the gene expression profiles can discriminate between individuals in a population; thus, control fish from a single group that were killed on three successive occasions over 3 weeks were grouped by the day of sampling with 100% accuracy (Cossins & Crawford, 2005).

A second major experiment was conducted in Liverpool to assess the effects of chronic, deep environmental hypoxia treatment at two different temperatures. While the analysis of the full dataset has not been completed, myoglobin genes have been noticed that were clearly hypoxia-upregulated at both 17 and 30° C in liver and other non-muscle tissues (Fraser *et al.*, 2006). This very surprising observation was extended to a range of other non-muscle tissues,

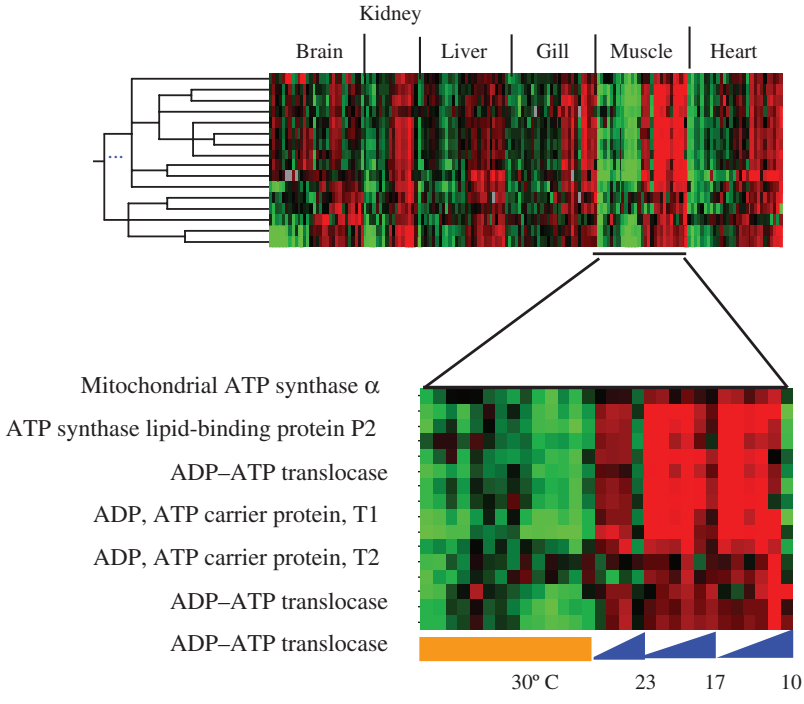

FIG. 2. The cold-inducible expression of genes involved in ATP synthesis in multiple tissues from the common carp. The upper panel displays the responses across six tissues, and the details shown below relates to skeletal muscle only. The genes represented in both panels are listed to the left, there being several probes for each of these genes. Data were taken from Gracey *et al.* (2004).

including brain. Careful alignment of the multiple myoglobin sequences in carpBASEv3 led to the discovery of a second isoform that was expressed only in the brain. To date, the common carp is the only vertebrate known to possess two myoglobin isoforms. Fraser *et al.* (2006) also demonstrated using 2D gel fractionation and mass spectrometric sequence determination that liver expresses a myoglobin protein, amounts of which were increased three-fold following 5 days of hypoxia verifying the microarray data. This work demonstrates the value of discovery-driven screening approaches for identifying both novel genes and novel responses of known genes. It also shows how inclusion of unidentified cDNA clones on the microarray or of clones of undistinguished forms of a particular gene can lead to new and serendipitous discoveries.

Clone sets and the associated sequences can form the basis for much expanded investigation of disease responses in carp. Gonzalez *et al.* (2007) detected immune-related genes in carp challenged by ectoparasitic infection with *Ichthyophthirius multifiliis* from two cDNA libraries. They identified 178 orthologues of immune-related genes of which 92 had not been previously described in *C. carpio*. RT-PCR was used to assess up-regulated and down-regulated responses to infection in skin tissue. Teleost skin has been largely ignored in immune responses, yet this work has revealed that it plays an active role in immune responses.

The common carp and goldfish are closely related; they diverged <10 million years ago. Trudeau and colleagues (Martyniuk *et al.*, 2006) employed the carp ARRAY v4 containing a series of goldfish brain-enriched cDNA clones generated in their laboratory. They established gene expression profiles in the brain of male goldfish exposed to environmental oestrogens. They identified a large number of gene responses, and for selected genes the responses were verified using quantitative real-time PCR including brain aromatase, secretogranin-III and interferon-related developmental regulator 1. They were able to address the dose-response relationships of these genes, and the microarray approach led to the adoption of a wider set of candidate genes in further analyses. Expansion of the goldfish cDNA collection is ongoing and will include sequences obtained from brain, pituitary, ovary and testes.

Helen Evans (Evans, 2005) has shown that the carpARRAY v1 performed as well with cDNA targets from a range of fish species including the grass carp *Ctenopharyngodon idella* (Valenciennes, 1844), the roach *Rutilus rutilus* (L., 1758) and the clown loach *Botia macracanthus* (Bleeker, 1852). These results, together with those produced by Orban, Christoffels and colleagues (Christoffels *et al.*, 2006), highlight the potential of the common carp microarray for cross-species applications. Although the family of Cyprinidae is the largest among freshwater teleosts, and it contains a number of important species for aquaculture, currently large EST collections and large-scale genomic tools are only available for zebrafish, fathead minnow *Pimephales promelas* Rafinesque, 1820 and common carp (Cossins & Crawford, 2005; Orban & Wu, 2008). The use of these bioinformatic resources and experimental tools for the analysis of 'sequence-poor' cyprinid species will open up new opportunities for the comparative genomic analysis of cyprinids, and perhaps even those species from other closely related taxonomic categories of teleost fish.

We thank G. Govan (Liverpool) for technical assistance in the aquarium, and R. Sreenivasan and H. Y. Kwan (both Singapore) for replicating and shipping the common carp testis clones. We thank Natural Environment Research Council (U.K.) and The European Commission for financial support (EUROCARP, Project 022665).

References

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

Ashburner, M., Ball, J., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A. P., Dolinski, K., Dwight, S., Eppig, J., Harris, M., Hill, D., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J., Richardson, J., Ringwald, M., Rubin, G. & Sherloc, G. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics* **25,** 25–29.

Balon, E. (1995). Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture* **129,** 3–48.

Baxevanis, A. & Ouellette, B. (2005). *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. Hoboken, NJ: Wiley.

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. & Hayashizaki, Y. (2000). Normalization and subtraction of Cap-trapper-selected cDNA to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Research* **10,** 1617–1630.

Christoffels, A., Bartfai, R., Srinivasan, H., Komen, H. & Orban, L. (2006). Comparative genomics in cyprinids: common carp ESTs help the annotation of the zebrafish genome. *BMC Genomics* **7** (Suppl. 5), S2.
Cossins, A. R. & Crawford, D. L. (2005). Opinion – fish as models for environmental genomics. *Nature Reviews in Genetics* **6,** 324–333.
David, L., Blum, S., Feldman, M., Lavi, U. & Hillel, J. (2003). Recent duplication of the common carp (*Cyprinus carpio* L.) genome as revealed by analyses of microsatellite loci. *Molecular Biology and Evolution* **20,** 1425–1434.
Diatchenko, B., Lau, Y.-F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Gurskaya, N., Sverdlov, E. D. & Siebert, P. D. (1996). Suppression subtractive hybridisation: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proceedings of the National Academy of Science of the United States of America* **93,** 6025–6030.
Diatchenko, L., Lukyanov, S., Lau, Y. F. C. & Siebert, P. D. (1999). Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Methods in Enzymology* **303,** 349–380.
Evans, H. (2005). Evolution of the acyl-CoA desaturases in cyprinid fish. PhD Thesis. Liverpool: University of Liverpool.
Feder, M. (2007). Key issues in achieving an integrative perspective on stress. *Journal of Biosciences* **32,** 433–440.
Fraser, J., de Mello, L., Ward, D., Rees, H., Williams, D., Fang, Y., Brass, A., Gracey, A. & Cossins, A. (2006). Hypoxia-inducible myoglobin expression in nonmuscle tissues. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 2977–2981.
Gonzalez, S., Chatziandreou, N., Nielsen, M., Weizhong, L., Rogers, J., Taylord, R., Santos, Y. & Cossins, A. (2007). Cutaneous immune responses in the common carp detected using transcript analysis. *Molecular Immunology* **44,** 1664–1679.
Gracey, A. Y. & Cossins, A. R. (2003). Application of microarray technology in environmental and comparative physiology. *Annual Reviews of Physiology* **65,** 231–259.
Gracey, A. Y., Troll, J. V. & Somero, G. N. (2001). Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *Proceedings of the National Academy of Science of the United States of America* **98,** 1993–1998.
Gracey, A., Fraser, E., Li, W., Fang, Y., Taylor, R., Rogers, J., Brass, A. & Cossins, A. (2004). Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 16970–16975.
Grillo, G., Licciulli, F., Liuni, S., Sbisà, E. & Pesole, G. (2003). PatSearch: a program for the detection of patterns and structural motifs in nucleotide sequences. *Nucleic Acids Research* **31,** 3608–3612.
Horvath, L. & Orban, L. (1995). Genome and gene manipulation in the common carp. *Aquaculture* **129,** 157–181.
Huang, X. & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research* **9,** 868–877.
Hulata, G. (1995). A review of genetic improvement of the common carp (*Cyprinus carpio* L.) and other cyprinids by crossbreeding, hybridization and selection. *Aquaculture* **129,** 143–155.
Hulata, G. (2001). Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica* **111,** 155–173.
Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.-F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. & Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* **34,** D354–D357.
Marchler-Bauer, A., Anderson, J., Cherukuri, P., DeWeese-Scott, C., Geer, L., Gwadz, M., He, S., Hurwitz, D., Jackson, J., Ke, Z., Lanczycki, C., Liebert, C., Liu, C., Lu, F., Marchler, G., Mullokandov, M., Shoemaker, B., Simonyan, V., Song, J., Thiessen, P., Yamashita, R., Yin, J., Zhang, D. & Bryant, S. (2005). CDD: a conserved domain database for protein classification. *Nucleic Acids Research* **33,** D192–D196.

Martyniuk, C., Xiong, H., Crump, K., Chiu, S., Sardana, R., Nadler, A., Gerrie, E., Xia, X. & Trudeau, V. (2006). Gene expression profiling in the neuroendocrine brain of male goldfish (*Carassius auratus*) exposed to 17-ethinylestradiol. *Physiological Genomics* **27,** 328–336.

Orban, L. & Wu, Q. J. (2008). The genetics and comparative genomics of cyprinids. In *Genome Mapping in Fishes and Aquatic Animals* (Kole, C. R. & Kocher, T., eds), pp. 45–83. Berlin: Springer Verlag.

Tiku, P. E., Gracey, A. Y., Macartney, A. I., Beynon, R. J. & Cossins, A. R. (1996). Cold-induced expression of Δ9-desaturase in carp by transcriptional and post-translational mechanisms. *Science* **271,** 815–818.

Wellenreuther, R., Schupp, I., Poustka, A. & Wiemann, S. (2004). SMART™ cDNA amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC Genomics* **5,** 1–8.

Wendelaar Bonga, S. (1997). The stress response in fish. *Physiological Reviews* **77,** 591–625.

Zhu, Y., Macheder, E., Chenchik, A., Li, R. & Siebert, P. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30,** 892–897.

