ELSEVIER

Aquat. Living Resour. 15 (2002) 193–196

Aquatic Living Resources

www.elsevier.com/locate/aquliv

Perspective Note

Use of emerging genomic and proteomic technologies in fish physiology

John Parrington, Kevin Coward *

Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK

Received 8 March 2002; accepted 15 April 2002

Abstract

The sequencing of the human genome represented a watershed in the biological sciences. Post-genomic biology will be marked not only by vastly greater knowledge about the human genome and those of other species but will see an increasing application of novel and highly sophisticated technologies. Genomic strategies, together with those that look at the proteome of cells and tissues, are likely to revolutionize scientific research over the coming years. The ease with which novel and homologous genes can be isolated using the new databases and technologies, and the ability to study the expression of thousands of genes simultaneously at a global cellular level will be the major factors in this revolution. Genomic information is already being used to further our understanding of physiology and gene evolution in fish. Furthermore, the highly compact pufferfish (*Takifugu rubripes*) genome is being used extensively as a model to interpret those of tetrapods. Currently, studies of the fish genome are limited to gene evolution and to a much lesser extent, environmental toxicology. However, as interpretation of fish genomes gathers pace, we are likely to see the increasing involvement of other key areas such as reproduction, growth, pathology of disease, and flesh development/quality. Here, we present some of the advanced genomic technologies currently available and discuss how these might influence our knowledge of fish biology.

Keywords: Genomic; Proteomic; Fish; Teleost; Physiology

Post-genomic tools and strategies look set to revolutionize the way in which biological research is undertaken. It is becoming possible to glean enormous amounts of information about biological systems within very small time frames. The common application of emerging genomic and proteomic technologies will play a vital role in the progression of science and our biological understanding of many organisms; fish are no exception. Of course, biotechnology already plays a significant role in the aquaculture industry. Sophisticated approaches such as ploidy manipulation, gynogenesis (all maternal inheritance), androgenesis (all paternal inheritance), gene transfer and marker-assisted selection are regularly utilized in at least some cultured species (Thorgaard, 1995). These biotechnological approaches are, however, only geared at the genetic improvement of desirable broodstock traits (such as sex control, growth rate, environmental tolerance and disease resistance), largely by chromosome set manipulation and improved selection. Although molecular techniques have further allowed us to isolate a suite of key genes such as the vitellogenin receptor, various cathepsins (B, D and L), and the receptors for growth hormone, thyroid hormone and prolactin (e.g. Brooks et al., 1997; Prat et al., 1998; Calduch-Giner et al., 2001; Marchand et al., 2001; Santos et al., 2001; Kwon et al., 2001) from commercially important fish, we still know relatively little about their regulation and global interaction.

Post-genomic technologies now offer the fish biologist the opportunity to gain far greater insight into the cellular and molecular mechanisms involved in the regulation of key physiological processes. This is particularly true of the various cellular signalling mechanisms controlling these vital processes. Genomic and proteomic technologies can be expected to prove important for fish physiology in three main areas: firstly, as a comparative means of analysing gene evolution; secondly, as a way of undertaking a global analysis of cell and tissue signalling pathways involved in physiological mechanisms regulating growth, reproduction, pathology of disease, stress physiology, and flesh quality; and finally, as a way of rapidly identifying novel proteins,

* Corresponding author.
E-mail address: kevin.coward@pharm.ox.ac.uk (K. Coward).

PII: S0990-7440(02)01172-5

and fish homologues of known proteins shown to play important roles in other animal groups.

Much of the present focus in fish genomics concerns the analysis of gene evolution (Meyer and Malaga-Trillo, 1999; Postlethwaite et al., 2000; Naruse et al., 2000; Taylor et al., 2001), especially in terms of the evolutionary and functional divergence of genes after duplication. This said, there still remains much debate on the origin of duplicated fish genes (Robinson-Rechavi et al., 2001). Mitochondrial genomes (mitogenomes) are becoming increasingly popular in the study of ancient divergences amongst many animal groups (Curole and Kocher, 1999) but are proving especially useful in fish (Miya et al., 2001), the most diversified group of all vertebrates, comprising over 23 000 species. A database containing groups of orthologous genes with one copy from man, mouse, chicken, xenopus and several copies from ray-finned fish (Actinopterygii) has recently been developed. Known as "Wanda", this curated database is expected to prove useful in the comparative study of gene evolution (Van de Peer et al., 2002). Further genomic resources for fish include ZFIN (Sprague et al., 2001), COG (Tatusov et al., 2001), PALI (Balaji et al., 2001) and HOVERGEN (Duret et al., 1994). A genomic study of the three-spined stickleback *Gasterosteus aculeatus* (L.) recently provided many clues to the genetic changes that define a species (Peichel et al., 2001). With the completion of the human genome, it is clear that emphasis will soon switch from data gathering to interpretation. Evolutionary comparisons show that the vertebrate genome is highly conserved. Comparative genomics, using "model" organisms, is likely to provide a powerful tool for such interpretation. Perhaps the best known fish genome at present is that of the pufferfish *Takifugu rubripes* (Temminck and Schlegel). Whilst ray-finned fish generally contain more copies of genes than other vertebrates, the pufferfish only possesses a very small genome and is thereby a highly useful model system (Elgar et al., 1996; Graves, 1998). Remarkably, the Fugu genome contains essentially the same genes and regulatory sequences in approximately 400 million bases as compared to the 3 billion bases that make up the human genome. A further compact genome being developed is that of another pufferfish, *Tetraodon nigroviridis*, a species 20–30 million-years distant from *T. rubripes* (Crollius et al., 2000). The zebrafish *Brachydanio rerio* (Hamilton) is also proving useful as a test system for functional genomics and therapeutic development application. Zebrafish embryos offer a system that rapidly provides efficacy, toxicity and specificity data within an in vivo setting (Nasevicius and Ekker, 2001).

Although fish genomes are being extensively used in the analysis of gene evolution, genomic approaches have yet to receive the full attention of more physiologically minded scientists. As the interpretation of fish genomes gathers pace, we are likely to see increasing efforts to exploit this multitude of information in those areas beneficial to the aquaculture industry. These areas might, for example, include the specific signalling events controlling resource partitioning between growth and reproduction, oocyte sequestration of the yolk precursor vitellogenin, fertilization/egg activation and the pathology of disease. Until recently, studies of cellular signalling pathways have necessarily tended to focus on particular signalling pathways or responses in isolation. With recent advances in genomic and proteomic approaches, it is now possible to undertake a much more systematic comparison of the changes in mRNAs and proteins that occur downstream of a signalling pathway at the global level. In contrast to traditional approaches which focus upon one or a small subset of genes, modern gene expression profiling allows us to study global patterns of gene expression, as it allows simultaneous assessment of the expression levels of thousands of mRNAs. A key technique in such global gene expression measurement is the DNA micro-array in which glass or silicon slides are arrayed with cDNAs or oligonucleotides complementary to transcripts from several thousand genes (Friend and Stoughton, 2002). The most sophisticated of these micro-arrays, also known as "DNA chips", are wafers no bigger than a thumbprint. The total mRNA expression from two samples, for instance, cells before and after a signalling response, can be compared on a single micro-array by converting each mRNA into a cDNA probe that incorporates a different fluorochrome (Deyholos and Galbraith, 2001). After probe hybridization, fluorescence measurements are converted into ratios of relative gene expression that reflect the difference in mRNA levels between samples. Alternatively, samples can be radioactively labelled and compared separately on identical micro-arrays. An important aspect of such analysis will be its ability to reveal the full scope of regulatory pathways. In what has been called the "guilt-by-association" approach, if genes in a tissue are found to switch on and off together in response to a stimulus, this can be a good indication that they are linked together in the same signalling pathway. The DNA micro-array is likely to become an incredibly powerful technique with which one can compare a normal fish gene population with one that has been manipulated experimentally. Such manipulations might involve ecotoxicity tests using endocrine-disrupting chemicals for example. Aquatic ecotoxicologists are already beginning to harness the power of genomic technology (Wise et al., 2002).

Use of micro-arrays in this manner is unsuitable for comparing global differences in gene expression between single cells. However, there are now amplification techniques that can generate cDNA from as few as a hundred or so cells that can then be used to make a probe for micro-array analysis. Another emerging technique is polymerase chain reaction serial analysis of gene expression (PCR-SAGE). In SAGE, mRNA is converted to cDNA and short expressed sequence tags (10 base pairs in length) are isolated from a defined position (Velculescu et al., 2000). By combining SAGE with PCR amplification, it has been possible to create a SAGE catalogue from as few as nine

human oocytes (Neilson et al., 2000) and it should ultimately be applicable to single cells. Studies indicate that SAGE can generate gene expression profiles with a quantitative similarity to those generated by micro-array analysis (Ishii et al., 2000). A key task in the future development of PCR-SAGE will be to make it a truly quantitative approach, using such innovations as real time PCR (Freeman et al., 1999), so that SAGE catalogues prepared from a single cell truly reflect the mRNA population of that cell.

While enormously powerful, approaches to global analysis of gene expression focused solely at the mRNA level nevertheless cannot provide a complete picture of the changes taking place in the cell during a signalling response and there is some concern over the correlation between protein level and mRNA. Genomic approaches also provide little or no information about co- and post-translational modifications such as phosphorylation, glycosylation, prenylation, ADP-ribosylation, or processing by limited proteolysis. This is particularly important as many signalling responses initially involve such post-translational changes, phosphorylation in particular. Proteomic approaches must be included in any study that aims to uncover the global pattern of protein expression, including post-translational modifications, within the cell (Dongre et al., 2001). The rapidly expanding field of proteomics relies heavily upon the use of two-dimensional electrophoresis of protein samples. This technique allows simultaneous separation of thousands of proteins at a time, excellent reproducibility, and ability to exhibit post-translational modifications. Reproducibility has been greatly increased by the use of IPG (immobilized pH gradient) technology and precast PAGE (polyacrylamide gel electrophoresis) gels (Fey and Larsen, 2001). Comparison of the proteome of two samples has traditionally been complicated by variations between gels. However, fluorescence 2D differential gel electrophoresis, where the protein samples are labelled with different fluorophores, has now made it possible to compare two samples on the same gel (Unlu et al., 1997). By linking such a technique to imaging software, an automated spot picker and protein identification techniques such as MALDI-TOF based peptide mapping (matrix-assisted laser desorption/ionization—time of flight mass spectometry), it is now possible both to compare the protein profiles of two tissues or before and after a tissue response, and to identify the proteins that are different in each case (Dongre et al., 2001). Ultimately, it seems likely that the same arraying technologies that are revolutionizing genomics will also transform proteomics, but at present “protein chips” that will allow the study of protein–protein interactions within the cell, are still in the early stages of development, with many technical problems still to be solved (Figeys and Pinto, 2001). Proteomic techniques are, however, already being utilized in fish biology. For example, Pineiro et al. (2001) report the use of proteomics in the characterization and partial sequencing of species-specific sarcoplasmic polypeptides in a marine fish of high commercial value.

The other main area where genomic technologies may be expected to transform fish physiology research is in identifying novel signalling proteins and homologues of known proteins in a variety of fish species. A variety of fish species are the subject of genome initiatives. As part of these initiatives, EST (expressed sequence tag) databases, cDNA databases, and DNA arrays are being created based on many different tissues from a range of species. These include model species such as zebrafish, medaka *Oryzias latipes* (Temminck and Schlegel) and pufferfish, but also commercially important species such as Atlantic salmon *Salmo salar* (L.) and tilapia *Oreochromis niloticus* (L.). Generation of these resources can be expected to transform our ability to identify important new proteins involved in fish tissue and cellular mechanisms as it will allow the rapid isolation of novel proteins along with homologues shown to be important in other species. It will be thus possible to generate a range of antibodies, fluorescent probes and other molecular tools that can used to study the physiological function of these proteins. One potential drawback that might hamper the future exploitation of some fish genomes, is that owing to genome duplication on the fish lineage (though there is still some debate on this issue; Robinson-Rechavi et al., 2001), many gene families appear to be typically half the size in land vertebrates than they are in fish (Meyer and Malaga-Trillo, 1999). This might be a factor that impedes the analysis and accurate interpretation of fish genomes.

These are exciting times for the fish physiologist. As we enter the post-genomic era and become acquainted with the almost bewildering array of genomic and proteomic technologies that are becoming available, the opportunities for highly technological and elegant research will become plentiful. On the one hand, the ease with which novel and homologous genes can now be isolated will mean that molecular tools can be rapidly created for functional studies. Cross-species comparisons will become far more feasible and wide reaching. On the other hand, the emphasis of post-genomic approaches on global analysis represents a paradigm shift in the biological sciences. One criticism of modern molecular biology has been that although its reductionist approach has been remarkably successful in identifying large numbers of novel genes involved in cell physiology and embryonic development, it has been less successful at showing how these interact functionally within the context of the living cell, tissue or organism. The ability to simultaneously study the expression of thousands of genes at both the mRNA and protein level offers the possibility of going beyond the limits of such reductionism, albeit from a highly molecular basis. The combination of these two features of post-genomic biology will in no uncertain terms revolutionize our understanding of fish physiology, evolution, adaptation and development. Ultimately, these advances ought to prove very fruitful in the many species of fish that are cultured commercially as well as in those species used purely for laboratory research.

Furthermore, the use of fish as a "model system" in the interpretation of the genomes from other taxa, should yield significant insight into the mammalian genome.

References

Balaji, S., Sujatha, S., Kumar, S.S., Srinivasan, N., 2001. PALI – a database of Phylogeny and alignment of homologous protein structures. Nucl. Acid. Res. 29, 61–65.

Brooks, S., Tyler, C.R., Carnevali, O., Coward, K., Sumpter, J.P., 1997. Molecular characterization of ovarian cathepsin D in the rainbow trout, *Oncorhynchus mykiss*. Gene 201, 45–54.

Calduch-Giner, J., Duval, H., Chesnel, F., Boeuf, G., Perez-Sanchez, J., Boujard, D., 2001. Fish growth hormone receptor: molecular characterization of two membrane-anchored forms. Endocrinology 142, 3269–3273.

Crollius, H.R., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quertier, F., Saurin, W., Bernot, A., Weissenbach, J., 2000. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. Genome Res. 10, 939–949.

Curole, J.P., Kocher, T.D., 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. 14, 394–398.

Deyholos, M.K., Galbraith, D.W., 2001. High-density microarrays for gene expression analysis. Cytometry 43, 229–238.

Dongre, A.R., Opiteck, G., Cosand, W.L., Hefta, S.A., 2001. Proteomics in the post-genome age. Biopolymers 60, 206–211.

Duret, L., Mouchiroud, D., Gouy, M., 1994. HOVERGEN: a database of homologous vertebrate genes. Nucl. Acids Res. 25, 2360–2365.

Elgar, G., Sandford, R., Aparicio, S., Macrae, A., Venkatesh, B., Brenner, S., 1996. Small is beautiful: comparative genomics with the puffer fish (*Takifugu rubripes*). Trends Genet. 12, 145–150.

Fey, S.J., Larsen, P.M., 2001. 2D or not 2D. Two-dimensional gel electrophoresis. Curr. Opin. Chem. Biol. 5, 26–33.

Figeys, D., Pinto, D., 2001. Proteomics on a chip: promising developments. Electrophoresis 22, 208–216.

Freeman, W.M., Walker, S.J., Vrana, K.E., 1999. Quantitative RT-PCR: pitfalls and potential. Biotechniques 26, 112–125.

Friend, S.H., Staughton, R.B., 2002. The magic of microarrays. Scient. Am. February, 34–41.

Graves, J.A., 1998. Background and overview of comparative genomics. Inst. Lab. Animal Res. J. 39, 48–65.

Ishii, M., Hashimoto, S., Tsutsumi, S., Wada, Y., Matsushima, K., Kodama, T., Aburatani, H., 2000. Direct comparison of GeneChip and SAGE on the quantitative accuracy in transcript profiling analysis. Genomics 68, 136–143.

Kwon, J.Y., Prat, F., Randall, C., Tyler, C.R., 2001. Molecular characterization of putative yolk processing enzymes and their expression during oogenesis and embryogenesis in rainbow trout (*Oncorhynchus mykiss*). Biol. Reprod. 65, 1701–1709.

Marchand, O., Safi, R., Escriva, H., Van Rompaey, E., Prunet, P., Laudet, V., 2001. Molecular cloning and characterization of thyroid hormone receptor in teleost fish. J. Mol. Endocrinol. 26, 51–65.

Meyer, A., Malaga-Trillo, E., 1999. Vertebrate genomics: more fishy tales about Hox genes. Curr. Biol. 25, R210–213.

Miya, M., Kagaguchi, A., Nishida, M., 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18, 1993–2009.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., Shimada, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Terado, T., Kimura, H., Nonaka, M., Shima, A., 2000. A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. Genetics 154, 1773–1784.

Nasevicius, A., Ekker, S.C., 2001. The zebrafish as a model system for functional genomics and therapeutic development applications. Curr. Opin. Mol. Therapeut. 3, 224–228.

Neilson, L., Andalibi, A., Kang, D., Coutifaris, C., Strauss, J.F., Stanton, J.A., Green, D.P., 2000. Molecular phenotype of the human oocyte by PCR-SAGE. Genomics 63, 13–24.

Peichel, C.L., Nereng, K.S., Ohgi, K.A., Cole, B.L., Colosimo, P.F., Buerkle, C.A., Schluter, D., Kingsley, D.M., 2001. The genetic architecture of divergence between threespine stickleback species. Nature 414, 901–905.

Pineiro, C., Vazquez, J., Marina, A.I., Barros-Velaquez, J., Gallardo, J.M., 2001. Characterization and partial sequencing of species-specific sarcoplasmic polypeptides from commercial hake species by mass spectrometry following two-dimensional electrophoresis. Electrophoresis 22, 1545–1552.

Postlethwaite, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., Talbot, W.S., 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. Genome Res. 10, 1890–1902.

Prat, F., Coward, K., Tyler, C.R., Sumpter, J.P., 1998. Molecular characterization and expression of two lipoprotein receptors in the rainbow trout, *Oncorhynchus mykiss*. Biol. Reprod. 58, 1146–1153.

Robinson-Rechavi, M., Marchand, O., Escriva, H., Laudet, V., 2001. An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr. Biol. 11, R458–R459.

Santos, C.R., Ingleton, P.M., Cavaco, J.E., Kelly, P.A., Edem, M., Power, D.M., 2001. Cloning, characterization and tissue distribution of prolactin receptor in the sea bream (*Sparus aurata*). Gen. Comp. Endocrinol. 121, 32–47.

Sprague, J., Doerry, E., Douglas, S., Westerfield, M., 2001. The Zebrafish Information Network (ZFIN): a resource for genetic, genomic and developmental research. Nucl. Acids Res. 29, 87–90.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., Koonin, E.V., 2001. The COG database: new developments in the phylogenetic classification of proteins from complete genomes. Nucl. Acids Res. 29, 22–28.

Taylor, J.S., Van de Peer, Y., Braasch, I., Meyer, A., 2001. Comparative genomics provides evidence for an ancient genome duplication event in fish. Phil. Trans. R. Soc. London B Biol. Sci. 29, 1661–1679.

Thorgaard, G.H., 1995. Biotechnological approaches to broodstock management. In: Bromage, N.R., Roberts, R.J. (Eds.), Broodstock Management and Egg and Larval Quality. Blackwell Science, Oxford, UK, pp. 76–93.

Unlu, M., Morgan, M.E., Minden, J.S., 1997. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis 18, 2071–2077.

Van de Peer, Y., Taylor, J.S., Joseph, J., Meyer, A., 2002. Wanda: a database of duplicated fish genes. Nucl. Acids Res. 30, 109–112.

Velculescu, V.E., Vogelstein, B., Kinzler, K.W., 2000. Analysing uncharted transcriptomes with SAGE. Trends Genet. 16, 423–425.

Wise, J.P., Winn, R.N., Renfro, J.L., 2002. Generating new marine cell lines and transgenic species – conference summary. J. Exp. Zool. 15, 217–220.