Journal of Heredity 2003:94(3):265–270

Computer Notes

Web Database of Molecular Genetic Data From Fish Stocks

A. Imsiridou, H. Hardy, N. Maudling, G. Amoutzias, and J.-M. Zaldívar Comenges

From 80B Ethnikis Antistaseos, GR-46100 Igoumenitsa, Greece (Imsiridou); 177 Laburnum Grove, Portsmouth, PO2 0HF, United Kingdom (Hardy); Bioinformatics Group, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester, M13 9PT, UK (Maudling and Amoutzias); and European Commission, Joint Research Center, Institute for Environment and Sustainability, TP 280, 21020 Ispra (VA), Italy (Zaldívar Comenges).

Address correspondence to A. Imsiridou at the address above, or e-mail: imsiri@otenet.gr.

Much research at the national and international level has been devoted to the development of several genetic methods for use in the characterization of fish stocks. We have developed a database that collates data from population genetic studies of fish. This database is accessible to researchers and control authorities on the Internet and should serve as a repository for genetic information of commercially important fish species. The prototype database has been developed in such a way that the new information can easily be updated by researchers in the field, who can submit their own data. The site can be found at http://fishgen.jrc.it.

Several genetic methodologies have been developed in order to identify the origin of fish stocks. There is a need to gather this data in a database which is easily accessible to researchers and control authorities. The main objective of our work was to develop a prototype database for the collation of data from population genetic studies of fish. This database is accessible to all members of the scientific community through the Internet and should serve as a repository for information about genetic differentiation of stocks of commercially important fish species. It includes data from fish studies involving various methodologies and techniques applied in fish genetics: allozyme electrophoresis, restriction fragment length polymorphism (RFLP), DNA sequencing, and DNA microsatellites. Furthermore, this database will allow molecular biologists working on genetic fish identification to rapidly and efficiently access the existing information about materials, methods, and obtained results for the desired species.

Systems and Methods

Type and Quality of Information

The use of molecular genetic techniques in fisheries research has grown over the past several years due to an increased awareness of the value of genetic data. We have chosen data for four types of analysis applied in fish genetics (allozyme, RFLP, sequencing, and microsatellite analysis).

Allozyme Analysis

Isozymes are functionally similar but separable forms of enzymes encoded by one or more gene loci. The isozymes that are products of different alleles at the same gene locus are termed allozymes (Markert and Moller 1959). Isozymes and allozymes are usually run on starch gels. Alternative forms of a given protein are separated on the basis of differences in their net charge (a function of their amino acid sequence), size, and shape. After the end of electrophoresis, isozymes are stained with a wide variety of stains, depending on the locus visualized. Different alleles (polymorphism) are viewed as separate bands on the starch gel. This technique can be useful for defining genetic markers for fish stock identification, as evidenced by numerous studies that document differences in protein allele frequencies or different diagnostic alleles, between stocks.

RFLP Analysis

The development of DNA amplification using the polymerase chain reaction (PCR) (Saiki et al. 1988) resulted in an increasing number of molecular techniques which are now being applied, such as RFLP analysis. Restriction enzymes are used to digest DNA strands at a specific location (recognition site), resulting in a number of DNA fragments. These fragments can be separated by gel electrophoresis, and differences in the pattern of the fragments between individuals are called RFLPs (Dowling et al. 1990). An RFLP may result from a base substitution that causes the gain or loss of a restriction site, or from an insertion/deletion mutation.

A distinct restriction pattern for each enzyme is denoted with a capital letter, and each fish is assigned a multiletter code that describes its composite mtDNA genotype. This method has been widely used to detect DNA variation between populations, as well as for the discrimination of different fish stocks.

Sequencing Analysis

The most common method of DNA sequencing is the dideoxy-termination method (Sanger et al. 1977). In this procedure, a primer is used to initiate synthesis of new strands of DNA; only one primer is used and consequently

only one of the two strands is synthesized. In addition to the regular nucleotides used in a normal synthesis reaction, modified (dideoxy) nucleotides are added to a sequencing reaction. This modification prevents the formation of the phosphate bond that would link the next nucleotide of the chain. Thus synthesis is terminated on the addition of a modified nucleotide, and the result is a population of fragments of different sizes. These fragments are labeled by either radioactive or fluorescent detection methods and then separated on a high-resolution electrophoretic gel; different genotypes are denoted as nucleotide sequences with the variable (polymorphic) nucleotide positions. This technique provides an unusually sensitive method for exploring differences among populations within species.

Microsatellite Analysis

Over the past 10 years a great number of fish studies have been published involving single-locus minisatellites or microsatellites, known as variable number tandem repeats (VNTR) loci. Microsatellites have a unit length of 1–6 bp, repeated up to about 100 times at each locus of nuclear DNA (Tautz 1989). Mutation rates are high, and heterozygosity is also high. Microsatellite loci can be studied by developing primers specific to unique flanking domains of individual microsatellite loci, allowing amplification and description of separate alleles.

Individual alleles at a locus differ in the number of tandem repeats of the unit sequence, and for this reason they can be differentiated by gel electrophoresis according to their size. Microsatellite markers have proved particularly valuable for fish stock discrimination because of the high levels of polymorphism compared with other genetic markers.

Data Collection

Genetic data from the above methodologies have been gathered for 11 fish species and stored in the prototype database (*Clupea harengus*, *Diplodus sargus*, *Esox lucius*, *Gadus morhua*, *Leuciscus cephalus*, *Merlangius merlangus*, *Merluccius merluccius*, *Oncorhynchus keta*, *Oncorhynchus tshawytscha*, *Salmo salar*, *Salmo trutta*): allozyme data (Bernatchez and Osinov 1995; Imsiridou and Triantaphyllidis 2001; Imsiridou et al. 1997; Lenfant and Planes 1996); RFLP data (Apostolidis et al. 1996; Cronin et al. 1993; Hansen and Loeschcke 1996; Imsiridou et al. 1998; Kornfield and Bogdanowicz 1987); sequencing data (Apostolidis et al. 1997; Bernatchez et al. 1992; Carr and Marshall 1991; Carr et al. 1995; McVeigh et al. 1991); and microsatellite data (Estoup et al. 1993; McConnell et al. 1995; Miller and Kapuscinski 1996; Rico et al. 1997; Tessier et al. 1997).

Data Organization

The database stores information on fish species, sample locations, analysis methods and results, and the authors of each study. The specifications of how these data are incorporated and represented in the database are described in the "glossary of terms and characteristics of fields in the database," which the user can access by clicking on the Information window.

Data Management

We have given considerable thought to how the data should be input. Access to the database is divided in to three different categories: database administrator, users with insertion privileges (accredited experts), and read-only users. Initially, because the staff responsible for database maintenance will input data, quality control will be less problematic. At present, because there is no validation mechanism in the database and the Web interface, only a small number of accredited experts will be allowed to manage the data in the database. All other users of the database will have read-only privileges. Two methods of user-validated data entry are under development: by using HTML forms and PHP, and by using Java applets on the client side connected by a TCP socket to a server-side application accessing the database via JDBC.

Database Development and Implementation

During the analysis phase we developed an off-line version of the database using the standard relational database application, Microsoft Access, which can be used on a PC. The main purpose was to check the validity of the data structure emerging from the entity-relationship diagram. The Access version of the database is now available on a CD-ROM. We then established a Web-accessible version using open source software Postgre SQL, which includes an object relational database management system.

Web Interface

The Web-database prototype was implemented according to the design considerations. The site URL is http://fishgen.jrc.it, and the first thing the user sees on the site is a welcome page. Once the user has entered the valid login (jrc) and password (guest), the user can access the main page with its three different navigation buttons (species, location, author).

Discussion

This is the first database to draw together a wide variety of genetic data (allozyme, RFLP, sequencing, microsatellites) from fish stock populations. These data can assist researchers and control authorities in identifying the origin of a fish stock by providing a collation of genetic data with which they can do calculations and make comparisons with their own data. Suggestions on how a researcher can use the database to find the most probable origin of a fish stock are described in the "user document," which can be accessed by clicking on the Information window link.

This database will assist researchers and scientific officers in making initial estimates of stock structure and differentiation. It will also provide researchers with a centralized

repository for their results, allowing them to share data more effectively with others. The database will facilitate the development of scientific and management programs on the study and conservation of commercially important fish stocks.

Acknowledgments

The authors are indebted to Mrs. Vasiliki Baka, Mr. Iain Shepherd, and Dr. Costas Triantaphyllidis for their valuable help in correcting the manuscript.

References

Apostolidis AP, Karakousis Y, and Triantaphyllidis C, 1996. Genetic differentiation and phylogenetic relationships among Greek *Salmo trutta* L. (brown trout) populations as revealed by RFLP analysis of PCR amplified mitochondrial DNA segments. Heredity 77:608–618.

Apostolidis AP, Triantaphyllidis C, Kouvatsi A, and Economidis PS, 1997. Mitochondrial DNA sequencing variation and phylogeography among *Salmo trutta* L. (Greek brown trout) populations. Mol Ecol 6:531–542.

Bernatchez L, Guyomard R, and Bonhomme F, 1992. DNA sequence variation of the mitochondrial control region among geographically and morphologically remote European brown trout *Salmo trutta* populations. Mol Ecol 1:161–173.

Bernatchez L and Osinov A, 1995. Genetic diversity of trout (genus *Salmo*) from its most eastern native range based on mitochondrial DNA and nuclear gene variation. Mol Ecol 4:285–297.

Carr SM and Marshall DH, 1991. Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. Can J Fish Aquat Sci 48:48–52.

Carr SM, Snellen AJ, Howse KA, and Wroblewski JS, 1995. Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus morhua*) from bay and offshore locations on the Newfoundland continental shelf. Mol Ecol 4:79–88.

Cronin MA, Spearman WJ, Wilmot RL, Patton JC, and Bickham JW, 1993. Mitochondrial DNA variation in chinook (*Oncorhynchus tshawytscha*) and chum salmon (*O. keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Can J Fish Aquat Sci 50:708–715.

Dowling TE, Moritz C, and Palmer JD, 1990. Nucleic acids II: restriction site analysis. In: Molecular systematics (Hillis DM and Moritz C, eds). Sunderland, MA: Sinauer Associates; 250–317.

Estoup A, Presa P, Krieg F, Vaiman D, and Guyomard R, 1993. (CT)n and (GT)n microsatellites: a new class of genetic markers for *Salmo trutta* L. (brown trout). Heredity 71:488–496.

Hansen MM and Loeschcke V, 1996. Genetic differentiation among Danish brown trout populations, as detected by RFLP analysis of PCR amplified mitochondrial DNA segments. J Fish Biol 48:422–436.

Imsiridou A, Apostolidis A, Durand JD, Briolay J, Bouvet Y, and Triantaphyllidis C, 1998. Genetic differentiation and phylogenetic relationships among Greek chub *Leuciscus cephalus* L. (Pisces, Cyprinidae) populations as revealed by RFLP analysis of mitochondrial DNA. Biochem Syst Ecol 26:415–429.

Imsiridou A, Karakousis Y, and Triantaphyllidis C, 1997. Genetic polymorphism and differentiation among chub, *Leuciscus cephalus* L. (Pisces, Cyprinidae) populations of Greece. Biochem Syst Ecol 25:537–546.

Imsiridou A and Triantaphyllidis C, 2001. Allozyme electrophoresis studies in European hake *Merluccius merluccius* (Gadidae) populations. Biol Bratislava 56/5:545–547.

Kornfield I and Bogdanowicz SM, 1987. Differentiation of mitochondrial DNA in Atlantic herring, *Clupea harengus*. Fish Bull 85:121–128.

Lenfant P and Planes S, 1996. Genetic differentiation of white sea bream within the Lion's Gulf and the Ligurian Sea (Mediterranean Sea). J Fish Biol 49:613–621.

Markert CL and Moller F, 1959. Multiple forms of enzymes. Tissue ontogenetic and species specific patterns. Proc Natl Acad Sci USA 45:753–763.

McConnell SKJ, O'Reilly P, Hamilton L, Wright JM, and Bentzen P, 1985. Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): genetic differentiation of North American and European populations. Can J Fish Aquat Sci 52:1863–1872.

McVeigh HP, Bartlett SE, and Davidson WS, 1991. Polymerase chain reaction/direct sequence analysis of the cytochrome *b* gene in *Salmo salar*. Aquaculture 95:225–233.

Miller LM and Kapuscinski AR, 1996. Microsatellite DNA markers reveal new levels of genetic variation in northern pike. Trans Am Fish Soc 125:971–977.

Rico C, Ibrahim KM, Rico I, and Hewitt GM, 1997. Stock composition in North Atlantic populations of whiting using microsatellite markers. J Fish Biol 51:462–457.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, and Erlich HA, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–490.

Sanger F, Nicklen S, and Coulson AR, 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467.

Tautz D, 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471.

Tessier N, Bernatchez L, and Wright JM, 1997. Population structure and impact of supportive breeding inferred from mitochondrial and microsatellite DNA analyses in land-locked Atlantic salmon *Salmo salar* L. Mol Ecol 6:735–750.

Received December 10, 2002
Accepted January 27, 2003
Corresponding Editor: Sudhir Kumar

DOI: 10.1093/jhered/esg057