

Genetic markers in marine biology and aquaculture research : when to use what

Antonios Magoulas
Geneticist

Giorgos Kotoulas
Geneticist

Kostas Batargias
Geneticist

Eleftherios Zouros
Geneticist

Introduction

Looking thirty years back, one realizes that the word "molecular" in the titles of the 1966 seminal papers by HUBBY and LEWONTIN (1966) and LEWONTIN and HUBBY (1966) was both misleading and visionary. Misleading, because no molecular biologist would ever consider scoring allozyme variation as part of his/her trade. Visionary, because today's population biologists can collect their primary information at the most basic and decisively molecular level: the nucleotide. Indeed, they have a rather wide and diverse collection of tools in their disposal. But as this collection grows, so does the choice of tools become more difficult. It is not always the case that the appearance of a new method renders another method obsolete. Indeed, one may argue that the possibility of simultaneous use of two or more assays of scoring genetic variation may shed light on questions that neither assay could answer alone.

The purpose of this communication is to provide a comparative evaluation of the various techniques now widely used in population genetics. Although the subject of the use of genetic markers for the study of natural populations has been extensively reviewed in the last years (*e.g.* AVISE, 1994), the question of what markers are most appropriate at any given instance has not yet received, to our knowledge, a comprehensive treatment. It is hoped that the elementary and sketchy presentation attempted here will give the impetus for a more formal and complete analysis of the subject. We will first list the various types of markers together with what we consider to be their advantages and disadvantages and then present examples from our own work in a way of illustration.

Tools

Antibody-antigen reaction polymorphism

It may be considered as the first type of "molecular" detection of polymorphism. The technique never received wide application because of sampling difficulties, but also because it became evident from human data that most blood groups are basically monomorphic (HEDRICK and MURRAY, 1978). The HLA polymorphisms in humans have, however, played an important role in the study of human genetics.

Allozymes

The simplicity and general applicability of the technique have made this the most widely studied form of molecular variation. Any source of soluble proteins, from bacterial cultures to animal fluids, is in principle suitable for allozyme analysis and the protocols of electrophoretic separation and staining are easily adjustable from species to species. The genetic interpretation of allozyme profiles (zymograms) is also straightforward. One major drawback has

been the inability to read genotypes from small quantities of tissue, which makes allozymes inapplicable for small organisms or for the immature stages (*e.g.* larvae) of large ones. But the main disadvantage, one that appears to be intrinsically difficult to overcome, is that only a small fraction of enzyme loci appear to be allozymically polymorphic. With small variations, the same set of few loci has been used in allozyme surveys from prokaryotes to fungi, plants and animals. What we have learnt from these studies is that selection is not the sole force that determines the genetic makeup of natural populations and that the theories of Wright, Crow and Kimura have relevance for the real world, beside their mathematical elegance. The present awareness that what we see in natural populations is the result of the interplay between mutation rate, effective population size and selection we owe to the allozyme revolution of the seventies and eighties. The quest to evaluate the relative importance of these forces by looking at coding or regulatory parts of the genome will continue. It may soon become possible to record amino acid polymorphism in large scale surveys by direct sequencing rather than by electrophoretic separation of the peptide product. This will increase enormously the number of protein loci that could be surveyed. Because allozymes cannot be assumed to be selectively neutral and because the amount of their polymorphism is limited (these two aspects are obviously related), they are not the assay of choice for the study of the biogeography of a species or its present population structure, but they might be important for the study of local adaptation. They also appear to be of limited value for phylogenetic studies.

Anonymous nuclear DNA markers

Under this category we include assays that target a segment of DNA of unknown function. The segment can be amplified from individual specimens and the polymorphism scored as length difference of the PCR products. Alternatively, the product may be digested by a set of restriction enzymes and the polymorphism scored as restriction fragment length polymorphism (RFLP). The primers are usually designed from sequences originally obtained for other purposes. Another method is to use the PCR product from a reference individual as a probe against digested total DNA from the sampled individuals. This calls for a more cumbersome protocol,

but has the potential to uncover more polymorphism. cDNA probes represent a special version of this technique. Individual clones from a poly-A messenger RNA cDNA library are amplified by using vector primers and used as probes against Southern transfers. The resulting polymorphism could be due to either presence/absence of a restriction site or to variable number of tandem repeats (VNTR). Anonymous DNA markers comprise a large and diverse family. Some reveal a high level of polymorphism, others are mostly monomorphic. In contrast to mini or microsatellites (see below), there is no way of knowing if the assay will produce polymorphism before it is actually used. The interpretation of banding profiles can also be complicated and ambiguous, depending on the nature of the underlying polymorphism. Finally, the assumption of neutrality may not be justifiable in certain cases, as for example in the case of cDNA.

Randomly amplified polymorphic DNA (RAPD)

The method uses single short primers of arbitrary sequence to amplify anonymous regions of genomic DNA. It is a fast and cheap assay, but the penalty for this convenience is poor reproducibility and ambiguity in the interpretation of results. The profiles are usually multibanded and polymorphism is scored as presence/absence of specific bands. Homozygosity for presence of a band cannot be distinguished from heterozygosity. As a result, most population genetics models cannot be applied, and analysis is based on phenotype rather than allelic frequencies. RAPDs are more suitable for species and subspecies comparisons, than for intra-species population studies.

Minisatellites

Discovered serendipitously, this variation forms the basis of DNA fingerprinting. A minisatellite "locus" consists of tandemly repeated "units" each of which contains a "core" sequence of around 12-16 nucleotide bases and two sequences flanking the core. Loci with the same core but different flanking sequences are scattered around the genome. Thus, probing with the core sequence results in

a multibanded profile unique to each individual. If the probe is the repeat unit of a specific minisatellite locus (*i.e.* the core and the two flanking sequences), rather than the core alone, the hybridization may produce two-banded profiles under stringent conditions, so that single-locus allele frequencies can be scored. Minisatellites mutate at a very high rate (as high as 15%), which makes them ideal for individual identification, but reduces their utility for deducing genetic relation among randomly selected individuals from a population (the short-memory concept of identity by descent).

Microsatellites

Although similar by name, this class of markers is quite different from minisatellites. The repeat unit is very simple (mostly two, but also three or more nucleotides), the flanking sequences of each repeat locus are unique and the total length of the "locus" is much smaller than in minisatellites. Most importantly, microsatellites are much more numerous in the genome (particularly of vertebrates) and have a mutation rate between 10^{-3} and 10^{-4} . They are ideal for mapping "causal" genes, whether these are responsible for single factor conditions (*e.g.* muscular dystrophy in humans) or for multifactorial traits (quantitative trait loci, QTL). They are also the best markers for determining parenthood in mass-crosses, tracing escapes from contained to wild populations and estimating coefficients of kinship among individuals drawn from a natural population. Their basic drawback remains the high cost and labor-intensiveness of the first phase of the technique, *i.e.* the development of primers. This is to some extent counter-balanced by the usually good crossability of primers in related species.

Mitochondrial DNA (mtDNA) variation

Three properties of mtDNA set it apart from nuclear DNA: it occurs in multiple copies in each cell (in contrast to two copies for a "single copy" nuclear locus), it is transmitted uniparentally, and it does not recombine. Presence of multiple copies does not, however, translate into a large variety of copies within the cell. For reasons not fully understood, the speed with which the maternal lineage of

a heteroplasmically conceived individual becomes homoplasmic is rather high. As a result, we can speak of the "mitotype" of an individual in the same way as we speak of its (nuclear) genotype. In addition (or as a result of this), the number of different mitotypes in a species is not huge, as originally suspected. If heteroplasmy rather than homoplasmy was the rule in nature, mtDNA would have been useless for population genetics. One consequence of uniparental transmission, which applies also to plasmid DNA, is that the effective population size for mtDNA is smaller than that of nuclear DNA, so that mtDNA variation is more exposed to the vagaries of random drift. Sex-specific differences in gene flow could also be revealed by contrasting nuclear with mtDNA. In a species in which mtDNA is maternally transmitted but gene flow occurs mainly or exclusively through males, divergence among populations is expected to be much higher for mtDNA than nuclear DNA. Incidental biparental transmission (when a small percentage of paternal mtDNA "leaks" into progeny) apparently is not a major concern in the use of mtDNA for phylogenetic or biogeographic studies, but biparentally induced heteroplasmy may signal the coexistence of subspecies or of highly diverged conspecific lineages in the population. From the point of view of population genetics, absence of recombination reduces the mtDNA molecule to a single gene. Coupled with uniparental inheritance, this means that the dynamics of mtDNA evolution are similar to those of a locus in a haploid asexually reproducing species. Such systems are prone to strong selective sweeps: an advantageous mutation anywhere in the genome will drive to fixation the type of molecule in which it occurred (or a deleterious mutation will drive it to extinction). At any given time, a population under selective sweeps will have a lower level of mtDNA variation, but viewed over long periods of time the rate of substitution at neutral sites would not be affected. That variation patterns compatible with selective sweeps were observed suggests that many sites in the mtDNA are under purifying selection. The notion that mtDNA evolves faster than nuclear single copy DNA is no longer accepted in its general form. It is evident that the rate varies both among species and among parts of the molecule within the same species. These differences have not reduced the importance of mtDNA as the molecule of choice for biogeographic and phylogenetic studies, but they must be considered when we are in need to provide explanations for observed patterns of variation.

RFLP, SSCP, Sequencing

All types of DNA variation are ultimately viewed as zones (bands). In most cases the different position of bands reflects difference in the length of DNA fragments. When the difference in length is caused by restriction enzymes, the conventional term for the method of detection is Restriction Fragment Length Polymorphism (RFLP). Clearly, sequencing represents the ultimate level of scoring of DNA variation. But it remains expensive and labor-intensive for large scale surveys. Single Strand Conformation Polymorphism (SSCP) is a relatively simple technique whereby the two strands are separated and forced to migrate independently. The speed of migration is affected by the conformation of the single strand. The utility of the technique depends on its discriminatory power. It is not yet known if single nucleotide differences can be detected. Another limitation is that at present it is restricted to small DNA fragments (normally less than 250 nucleotides).

▮ Applications

It is obvious from this short evaluation of the various types of genetic markers that the marker of choice will depend on the particular question one wants to ask. There is no such a thing as an all-purpose marker. In this section we present examples from our own work which illustrate this point.

Heterozygosity and fitness

Even before the advent of genetic markers, evolutionists were divided in two schools. The classical school supports the view that there can be only one best allele at a time and that homozygosity for that allele is the best genotype. The balanced-polymorphism school supports the view that there can be no best allele, but only good or bad combinations of alleles and that most of the time

heterozygosity represents a better combination than homozygosity. The matter remains unsolved to this date. The critical test is to distinguish between homozygote fitness depression caused from homozygosity for the marked locus itself and depression resulting from homozygosity by descent in a linked but unscorable locus. POGSON and ZOUROS (1994) attempted to answer this question by correlating fitness (shell size in individual scallops) with degree of heterozygosity for two sets of loci, allozymes and anonymous cDNA's. These two sets of markers appear to be the best for this purpose. Because both come from transcribed parts of the genome the possibility that they are linked to a hidden deleterious gene is the same. For the classical school this means that allozymes and cDNAs should behave the same way. The balanced school, by placing the emphasis on the scored locus, predicts that allozyme heterozygotes would outperform allozyme homozygotes, but that this difference would not be observed between homozygotes and heterozygotes for cDNA variants which are assumed to be neutral. The results were compatible with the balanced-polymorphism school. It remains to be seen if the same will be observed in other organisms.

Selection versus random drift in natural populations

Typically, ecologists and, particularly, managers of natural biological resources are not interested in the genetic mechanisms of selection. Rather, they want to know if there are populations within a species that can be considered as separate units from the point of view of reproduction and interaction with the environment and the ecosystem. This proved to be an especially difficult issue to resolve by means of populations genetics. If two populations are found to be genetically homogeneous for the markers used, the answer can be that both populations are under the same regime of selection which forcefully maintains the same allele frequencies, or that the two populations are homogeneous because of extensive exchange of migrants. If the populations are found to be different, the explanation can be that they are under different selection pressures or that the populations have been isolated from each other for so long that random drift has established different allele frequencies.

The best way to answer this is to examine two or more types of variation simultaneously. Indeed, allozymes, mtDNA and nuclear markers that can be assumed to be selectively neutral might be the best combination for this purpose. When this was attempted by KARL and AVISE (1992) in the American oyster, the result was unexpected: there was no geographic differentiation for allozymes, but there was a strong dichotomy between Atlantic and Gulf populations for both mtDNA and anonymous nuclear genes. The authors argued that allozymes are under the same type of selection across the distribution of the species. As a result they carry no information about the species past history or its present breeding structure. These could, however, be inferred from the distribution of mtDNA and nuclear variation. Later studies of the same populations (MCDONALD *et al.*, 1996) suggested that the actual situation may not be as simple.

We are currently applying this strategy to the European anchovy and the swordfish, and to a lesser extent to several species of the family Sparidae. A comprehensive mtDNA survey of the anchovy (MAGOULAS *et al.*, 1996) revealed a sharp distinction between Black Sea and the rest of the species distribution that can be best explained by historical events going back to at least the last ice age and by current one-way gene flow from Black Sea into the Aegean. This is one of the best examples of a unidirectional gene flow in the marine environment and how it can preserve the footprints of the population history of the species. As usually happens with observations of special interest, it raises new questions and offers an opportunity to readdress unsettled issues. Is the huge Black Sea anchovy population genetically uniform? Do species that inhabit both the Black and the Aegean Seas exhibit one of three distinct types of genetic heterogeneity determined by the pattern of gene flow: those in which migration is restricted in either way, those that experience unidirectional flow and those in which gene flow occurs in both directions? The favored explanation for the uniformity of allozyme frequencies that has been seen in many species distributed over vast and diverse geographic areas is some form of balancing selection that is a characteristic of the species itself rather than to the environment of any specific population. If this is true, the Black Sea and the Bay of Biscay populations of anchovy ought to be allozymically similar in spite of the fact that the first population has not received genes from the second for the last 10,000 years. If the two populations were found to be different for allozymes, the

selection hypothesis will be considerably weakened and allozyme uniformity, where observed, must be explained by gene flow. Microsatellites are useful here for yet another reason. It is hard to provide direct evidence that an individual with a Black Sea mitotype found in the Northern Aegean is indigenous or an immigrant from the Black Sea. This could be answered if microsatellites turned out to be sufficiently different between the two seas. With a large number of microsatellites it might even be possible to trace immigrants from the Black Sea or their immediate progeny in the Central Aegean and the Ionian Seas.

Unlike anchovy, present day patterns in the genetic differentiation of swordfish populations cannot be attributed to historic accidents, but rather to active migration for feeding or spawning. The observed differences between samples from the Mediterranean and the Gulf of Guinea (KOTOULAS *et al.*, 1995) can be most easily explained by some kind of homing behavior. Again microsatellites are the best genetic markers available to provide support for this hypothesis. If adults from different stocks mix but remain reproductively isolated by returning to different spawning grounds, there must be a strong correlation between mtDNA and microsatellites variants. This correlation, if it exists, could provide the basis for decomposing a sample of individuals caught in a given area into "natives" and "visitors", or for identifying the geographical origin of individuals converging on the same feeding grounds from different parts of the world.

Genetic markers in aquaculture research

We are using microsatellites in two aquacultured species, the gilthead sea bream and the Japanese oyster. In both cases, the primary purpose has been to identify the parents of an offspring from a mass mating. The ability to do this has revolutionized the practice of genetic breeding. It allows for a quick, inexpensive and more accurate estimation of heritabilities, and for evaluating the genetic quality of parents through the performance of their offspring. In the sea bream we have demonstrated additive genetic variance for growth, and identified individual parents that have consistently produced faster growing progeny. The same technique can be used to determine genetic differences in viability and

to estimate inbreeding depression. But in this case the statistical power of the assay is much lower, so that much larger samples are required. Microsatellites are also ideal to trace and verify cytogenetic treatments, such as induction of polyploidy or gynogenesis and have been used for this purpose in the Japanese oyster.

Conclusion

Assuming that obtaining the full nucleotide sequence of long pieces of DNA from thousands of individuals will continue to be an expensive practice for the immediate future, genetic markers will remain the main tool in population genetics. From the variety of markers presently available, it can be said that microsatellites, mtDNA and allozymes will maintain their eminence (more or less in that order). Studies whose the main objective is to trace an individual or deduce the genetic relationship between two individuals may rely exclusively on microsatellites. Similarly, mapping unknown genes affecting a quantitative or qualitative character may need no other markers except microsatellites. One can expect that soon there will be a data bank from which one would be able to retrieve microsatellite primer sequences for an ever increasing number of species.

It is, however, doubtful whether the more elusive questions about the history of natural populations and evolution would be answered by the use of one or the other kind of genetic markers. The most successful studies in this field will be those that would use two or more types of markers. Microsatellites and mtDNA will be part of the arsenal, which must also include a type of variation sensitive to selection. While allozymes serve this purpose at present, there is no reason why they could not be replaced with methods that would be able to survey a wider range of loci more likely to be under selection, such as regulatory, developmental or behavioral genes.

References

- AVISE (J.C.), 1994 —
Molecular Markers. Natural History and Evolution. Chapman & Hall, N.Y.
- HEDRICK (P.W.), MURRAY (E.), 1978 —
 Heterozygosity revisited.
Amer. J. Hum. Gent., 30: 377-382.
- HUBBY (J.L.), LEWONTIN (R.C.), 1966 —
 A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54: 577-94.
- KARL (S.A), AVISE (J.C.), 1992 —
 Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. *Science*, 256: 100-102.
- KOTOULAS (G.), MAGOULAS (A.), TSIMENIDES (N.), ZOUROS (E.), 1995 —
 Marked mitochondrial DNA differences between Mediterranean and Atlantic populations of the swordfish *Xiphias gladius*.
Mol. Ecol., 4: 201-210.
- LEWONTIN (R.C.), HUBBY (J.L.), 1966 —
 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics*, 54: 595-609.
- MAGOULAS (A.), TSIMENIDES (N.), ZOUROS (E.), 1996 —
 Mitochondrial phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*).
Mol. Biol. Evol., 13: 178-190.
- MCDONALD (J.H.), VERRELLI (B.C.), GEYER (L.B.), 1996 —
 Lack of geographic variation in anonymous nuclear polymorphisms in the American oyster *Crassostera virginica*.
Mol. Biol. Evol., 13: 1114-1118.
- POGSON (G.H.), ZOUROS (E.), 1994 —
 Allozyme and RFLP heterozygosities as correlates of growth rate in the scallop *Placopecten magellanicus*: A test of the associative overdominance hypothesis.
Genetics, 37:221-231.