The use of isozyme characters in systematic studies

Donald G. Buth*,*a and Robert W. Murphyb

*a Department of Biology, University of California (UCLA), Box 951606, 621 Circle Drive South, Los Angeles, CA 90095-1606, USA
b Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, 100 Queen’s Park, Toronto, Ont., Canada M5S 2C6

Received 16 January 1998; accepted 19 February 1998

Abstract

Several classes of isozyme characters are recognized including the number of genes that control a given enzyme system, intralocus and interlocus heteropolymer assembly, tissue-specific gene expression, developmental patterns of enzyme expression, and post-translational modification of gene products. All of these characters can be of value in systematic studies although few applications have been made to date; some of the more robust examples are recognized. Isozyme characters should be evaluated in terms of both their advantages and limitations. The level of universality at which they are used is group-specific. Tissue-specific gene expression data may be of the widest applicability in systematic studies. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Isozymes; Number of genes; Gene expression; Heteropolymer assembly; Post-translational modification

1. Introduction

In systematics, the information generated by the starch gel electrophoresis (SGE) of enzymes (including polyacrylamide PAGE and cellulose acetate CAGE; Richardson et al., 1986; Murphy et al., 1996) has become synonymous with allozymic data (e.g. Ayala, 1983). The historic pathway from the first development of this technology to applications in systematic biology has been indirect, and several subdisciplinary “filters” have narrowed the potential of the information that could be obtained. It is
our thesis that this additional information need not be overlooked, but rather applied in systematic studies when appropriate.

Electrophoretic technology facilitated the recognition that enzymes come in multiple forms. Markert and Moller (1959; p. 753) proposed the term *isozyme* “to describe the different molecular forms in which proteins may exist with the same enzymatic specificity”. This functional definition was intended to be broad, to cover all molecular forms of enzymes. This new wealth of variation would come to be of considerable evolutionary interest. However, it was the field of developmental biology that made the most of these early findings. The expression of various isozymes was found to differ both temporally (ontogenetic differentiation) and spatially (different tissues). When the study of gene activation became comparative, interspecific differences were noted. The potential was there to use these characteristics as markers for the study of both genic and organismic evolution (Markert, 1977; Whitt, 1981a). However, in the hands of developmental biologists, the study of genic evolution predominated, e.g. Markert et al. (1975), Fisher et al. (1980) and Whitt (1981a).

Population geneticists soon recognized that a subset of the isozyme variation being described would be ideal for studies of population structure, i.e. the level of gamma-taxonomy as defined by Mayr (1969). Allelic isozymes or *allozymes* (*sensu* Prakash et al., 1969) were recognized as the products of alleles of the same locus. The codominant nature of expression of allozymes allowed the unambiguous identification of heterozygous and homozygous individuals, which in turn allowed for the estimation of population structure, gene flow, and the overall assessment of agreement with Hardy–Weinberg equilibrium expectations. The resolution of products of many variable loci made population structure studies even more robust. Interspecific comparisons of allozymes that yielded measures of “genetic relationships” entered the level of Mayr’s (1969) beta-taxonomy and the bridge to the field of systematics had been built.

Avise (1974) provided a key review of the potential of these data in systematic biology. However, this review also contained many procedures that had become standard in population genetics but were unnecessary limitations to systematics. These included (1) using only the allozymic component of enzymatic variation, (2) summarizing particulate allozymic data into distance or similarity coefficients, and (3) employing only phenetic methods to estimate relationships. All three of these points were discussed in Buth’s (1984) updated review: (1) that isozyme characters could also be of systematic value, (2) that allozyme data could be treated in a particulate manner, although the coding of such data remains controversial (Mickevich and Mitter, 1981; Avise, 1983; Buth, 1984; Swofford and Berlocher, 1987; Murphy, 1993; Mabee and Humphries, 1993; Wiens, 1993; Berlocher and Swofford, 1997; Murphy and Doyle, 1998; Murphy and Lovejoy, 1998), and (3) obviously both allozymic and isozymic data can be analyzed cladistically. Thus, the acquisition of enzyme electrophoresis as a tool in systematic biology came largely through population genetics, which had filtered out much of the systematically useful potential that had been previously realized in developmental biology. These “filters” have limited the full potential of these data both in gathering the maximum amount of information often with no additional time, effort, or expense, as well as in their analysis.
1.1. What are isozyme characters?

Whitt (1981a; p. 271) noted “...the use of allozyme data exploits only a portion of the genetic information available through electrophoresis”. He proceeded to discuss two isozyme characters: the actual number of genes encoded in multiple loci and the patterns of tissue expression manifested by these gene products. Buth’s (1984) outline of isozyme characters used these two and added the structural character of heteropolymer formation. The distinction was made between intralocus and interlocus heteropolymer formation; the former is possible among multimeric enzymes as expressed in heterozygotes whereas the latter could form among multimeric products coded by different loci. Ferguson (1988) partitioned isozyme characters into those that are of a structural nature (number of loci, heteropolymer formation) and those of a regulatory nature. The latter would include the previously noted tissue expression character but also another, developmental expression. Thus the spatial (tissue) and temporal (ontogeny) aspects of expression important to developmental biologists (Whitt, 1981a) were now clarified as potentially informative characters in phylogenetic inference. Finally, tissue-specific post-translational modifications were identified as another category of isozyme character (Murphy and Crabtree, 1985; Murphy et al., 1996). Isozyme characters are summarized in Table 1.

1.2. Kinds of isozyme characters

1.2.1. Number of loci

The number of loci that controls the production of a particular enzyme may increase or decrease in response to gene duplication or gene silencing events, respectively. When these conditions are maintained through subsequent speciation events, they may be treated as synapomorphies.

The duplication of genes is an important factor in the evolution of many groups of organisms (Ohno, 1970). Genes can be duplicated individually via tandem duplication, a process that could yield multiple structural loci that could remain under single, non-duplicated regulatory control. Buth (1984) outlined the following as an example of how a tandem duplication could become a synapomorphy: (1) appearance as a polymorphism among individuals within a species (e.g. a GPI locus in *Cheirodon axelrodi*; Kuhl et al., 1976), (2) fixation of the duplication within a species via evolutionary processes such as natural selection and/or genetic drift (e.g. a GPI locus in *Acanthopsis choirorhynchus*; Ferris and Whitt, 1977), and (3) sharing the duplication among daughter species (e.g. a third mitochondrial aspartate transaminase locus in certain lineages of tetraploid catostomid fishes; Buth, 1979; Crabtree and Buth, 1981). Alternatively, entire genomes can be duplicated via polyploidy.

1.2.2. Heteropolymer formation

Over 70% of commonly studied enzymes are multimeric in structure (Harris and Hopkinson, 1976). In the heterozygous state, two or more polypeptide subunits usually combine randomly in vivo to produce zones of activity when subjected to electrophoresis in starch gels. The resultant patterns are easily interpreted because the
Table 1
Isozyme characters, and the nature of their character states. A character state is expressed in all members of a taxonomic unit

<table>
<thead>
<tr>
<th>Characters</th>
<th>Character states</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Structural characters</td>
<td>Presence or absence of a particular gene</td>
</tr>
<tr>
<td>A. Number of genes</td>
<td>Presence or absence of heteropolymers formed between different allelic forms of an orthologous gene encoding a multimeric enzyme; potential expression only in the heterozygous condition</td>
</tr>
<tr>
<td>B. Heteropolymer assembly</td>
<td>Presence or absence of heteropolymers formed between products of paralogous genes encoding a multimeric enzyme; potential expression only in multilocus enzyme systems</td>
</tr>
<tr>
<td>1. Intralocus</td>
<td></td>
</tr>
<tr>
<td>2. Interlocus</td>
<td></td>
</tr>
<tr>
<td>II. Regulatory characters</td>
<td></td>
</tr>
<tr>
<td>A. Tissue-specific expression</td>
<td>Presence or absence of products of a particular gene in a particular tissue; can also be coded as the array of tissues expressing a particular gene product</td>
</tr>
<tr>
<td>1. Qualitative</td>
<td>Relative or actual measure of a particular gene product expressed in a particular tissue</td>
</tr>
<tr>
<td>2. Quantitative</td>
<td>Relative or actual temporal stage at which the products of a particular gene are expressed in a particular tissue</td>
</tr>
<tr>
<td>B. Developmental expression</td>
<td></td>
</tr>
<tr>
<td>III. Post-translational modification</td>
<td>Presence or absence of a modified gene product; may or may not be tissue-specific in its expression</td>
</tr>
</tbody>
</table>

Subunits usually have different electrical charges (thus different electrophoretic properties) and combine in predictable ratios (see Figs. 3–5 of Buth, 1990). If many differences accrue among subunits, the polypeptides cannot form stable heteropolymers, i.e. their combination is no longer random. Reduction in expected quantity or cessation of production of intralocus heteropolymers is a pattern of variation that can be exploited in systematic studies.

Among the diploid vertebrates, approximately 50% of commonly studied enzyme systems are under multilocus control (Harris and Hopkinson, 1976). When different polypeptide subunits are produced by different (i.e. paralogous) loci, interlocus heteropolymers may or may not be formed. The resultant electrophoretic patterns are not as easily interpreted because the gene products are usually unequal in quantity in a given tissue and do not combine in the symmetrical ratios expected with allelic products of the same locus. Nevertheless, multilocus multimeric enzyme systems can vary with regard to the production of interlocus heteropolymers and this variation can be of systematic value.

1.2.3. Tissue expression

Structural gene loci are responsible for the production of a given enzyme but a separate set of regulatory genes determine where and how much, if any, of a product is expressed. Kettler et al. (1986; p. 485) determined that “differences in activity for
a given enzyme among tissues, as well as for different enzymes in the same tissue, were sufficiently independent to permit each locus-tissue expression of a species to be treated as a separate character”. The strategy for the development of a robust tissue expression data set is, therefore, the opposite of that used in allozyme studies in which a single optimal tissue is usually used as the source for a particular enzyme. Tissue expression data sets include information on a particular enzyme’s activity across several tissues, with such activity scored as presence-absence (i.e. qualitative coding) or quantified in each tissue. Such gene expression differences are assumed to be steady-state levels of activity in differentiated adult tissues (Kettler et al., 1986).

1.2.4. Developmental expression

As discussed by Ferguson (1988), interspecific differences in ontogenetic patterns of isozyme expression reflect the activities of regulatory genes (Whitt, 1981b). The ontogenetic changes in enzyme expression are known for many enzyme systems in many species, e.g. the LDH system in the heart of Mus musculus (Fig. 3 of Markert, 1983). However, we are unaware of any robust data set of ontogenetic patterns of enzyme expression that was generated for the specific purpose of addressing a systematic problem. Ferguson (1988) suggested that the study of developmental expression might be facilitated via the perturbation of normal development as seen in hybrids (e.g. Philipp et al., 1983).

1.2.5. Post-translational modification

Murphy et al. (1990) reviewed some occurrences of post-translational modifications of isozymes. There appears to be two classes of these changes; (1) those that are environmentally mediated or are physiological functions (e.g. van Tets and Cowan, 1966; McGovern and Tracy, 1981; Hillis and Patton, 1982; Fields et al., 1989), and those that have restricted tissue expressions (Murphy and Crabtree, 1985; Murphy et al., 1990). The tissue restricted enzyme expression in rattlesnakes (Murphy and Crabtree, 1985) appears to be autapomorphic (Murphy, unpublished data), and has no phylogenetic significance. Although the expression of the modified mMdh-A isozyme occurs in several species of amblastomatid salamanders (e.g., Mdh-3 in Bogart et al., 1987, and references therein), the isozyme data have never been used as systematic characters because expression of the modification usually occurs in liver tissue only, and is absent in larvae, which are sometimes used as tissue sources in the comparisons (J. Bogart, pers. comm.).

2. Applications of isozyme characters

2.1. Selected examples

Whereas the arguments for the recognition of isozyme characters and their application in systematics have been made before (Whitt, 1981a, 1983, 1987; Avise, 1983; Buth, 1984), relatively few studies have actually applied these data in systematic studies. More commonly, isozyme characters were investigated for their own sake
(e.g. gene evolution) rather than being applied to an organismic genealogical estimation. Isozyme characters are often mapped onto existing phylogenies to aid in their evolutionary interpretation (e.g. Sites et al., 1986; Murphy, 1988; Pierce and Crawford, 1997). Reports of autapomorphic isozyme characters are common (e.g. Rainboth et al., 1986).

The most robust applications of gene number characters have come from studies of allotetraploid fishes in which gene silencing (“functional diploidization”) has become an important phyletic event (Buth, 1978, 1979; Ferris and Whitt, 1978a; Ferris, 1984). However unlikely (Avise, 1983), maximum parsimony treatments of these data must allow for reversals (Buth, 1982). Gene duplication changes have been important in plant systematics (Gottlieb, 1977, 1982; Gottlieb and Weeden, 1979).

Restrictions on heterotetramer formation in L-lactate dehydrogenase isozymes were used to define reptilian families Lacertidae (Gorman, 1971) and Scincidae (Murphy, 1984). See Murphy (1998) for an expanded evaluation of this character. Restriction of intralocus heterodimer formation in creatine kinase isozymes (Ck-A locus) appears to be limited to teleost fishes (Ferris and Whitt, 1978b; Buth et al., 1985).

Robust sets of tissue expression characters have been applied to elucidate phylogenetic relationships among umbrid fishes (Kettler et al., 1986), Hawaiian “picture-winged” Drosophila (Thorpe and Dickinson, 1988), and chelid turtles (Lieb et al., 1998). These three studies should be consulted for coding alternatives for these data.

Synapomorphic post-translational modifications have yet to be recognized and applied in systematic studies.

2.2. Advantages

Several attributes of isozyme characters facilitate their study. Like many other biochemical-molecular data, the lack of environmental influence, independence of characters, and relatively simple genetic control can be viewed as positive attributes. Isozyme characters can be coded in a particulate fashion and do not need to be summarized as distance coefficients (Buth, 1984; Murphy, 1993). No additional technology is required for the resolution of isozyme characters beyond what is already extant in a laboratory equipped for allozyme study (Murphy et al., 1996). Indeed, some characters such as heteropolymer formation are essentially free as they may be scored from gels otherwise stained for allozymic information. In cases where intact frozen specimens were collected, relevant tissue arrays are already in hand.

Because electrophoretic mobility is the feature compared among allozymes, such comparisons must be contemporaneous (aka “side-by-side comparisons”). Many isozyme characters are free from such temporal constraints and can be generated as soon as any relevant taxon becomes available, similar to the acquisition of most DNA sequence data. In fact, such an approach is necessary for tissue expression data wherein different storage times may insert an unwanted variable.

Allozymic variation is taxonomically limited, usually to comparisons at and below the generic level (Bush and Kitto, 1978). Higher-level comparisons usually yield no
shared alleles so synapomorphies are unable to be recognized (e.g. Sites et al., 1984). However, for several decades, allozymic data were the characters of choice for population-level comparisons. Isozyme data may be useful at the population level as well. Along with allozymes, Aguirre et al. (1998) collected gene duplication and silencing data for tree lizards (\textit{Urosaurus}) from the Baja California region of Mexico. They found that the isozyme expression of duplicated genes had patterns of geographic distribution that correlated exceptionally with peninsular paleostratigraphy. While isozyme characters remain to be fully explored at the population-level, their potential range of utility extends to much higher taxonomic categories compared to allozymes.

Perhaps the greatest value of isozyme characters lies in their utility as alternative data sets. The quest for additional data seems never-ending regardless of whether the investigator desires independent comparisons for congruence (Mickevich and Johnson, 1976; Hillis, 1987) or a “total evidence” approach (Kluge, 1989).

2.3. Limitations

2.3.1. Number of potentially-informative characters

One limitation in using isozyme characters in genealogical estimation is the low number of characters and alternative states that can be derived from these sorts of data. In only two situations, e.g. broad-based tissue expression comparisons (Kettler et al., 1986; Lieb et al., 1998) and gene number differences among older polyploids (Ferris and Whitt, 1978a), would an isozyme character data set be sufficiently robust to “stand alone” for the evaluation of congruence with phylogeny estimates based on other data sets. Whereas the characters of these alternative data sets might be numbered in the dozens, the most common means of coding such data reduces each character to just two states; presence vs absence of an enzyme in a given tissue, or duplicated vs functionally diploid loci, respectively. In both cases, an alternative state could be the result of a loss event (see homoplasy discussion). Isozyme characters such as the number of genes among diploids, heteropolymer formation and post-translational modifications are expected to be few in number and supplemental to other characters.

2.3.2. Predictably higher levels of homoplasy

Isozyme data have predictably high levels of homoplasy compared to tracking allelic mutations at a locus. In many respects, they can suffer from the same problems associated with coding alleles as being present or absent (Murphy, 1993; Murphy and Lovejoy, 1998) in that the loss of isozyme expression (or allelic array) can be considered synapomorphic. For example, with most isozyme characters, if two terminal non-sister taxa have lost the expression of an enzyme locus independently, then the probability of a homoplastic loss of the plesiomorphic state is 100% (Murphy, 1993). A similar situation results from the independent duplication of loci (Sites and Murphy, 1991).

Unlike isozyme character states, one allele can potentially change into a relatively large number of alternative alleles, figuratively represented as a “starburst.” For
example, Highton and Larson (1979) reported the resolution of 26 alleles at an esterase locus, 23 alleles for the transferrin locus, and 17 alleles for Ldh-B in salamanders of the genus *Plethodon*. Similarly, Singh et al. (1976) resolved 37 alleles (electromorphs) at the Xdh-A locus of *Drosophila pseudoobscura*, making the probability of parallel change in allozyme data at least 1/37 or \( P = 0.027 \). Thus, unlike the 100% likelihood of parallel loss of isozymes in a tissue, or the loss of heteropolymers (Murphy, 1988), there is a relatively small chance of homoplasy resulting from parallel mutation events for allozymes (assuming that a multi-buffer sequential approach is sufficient to resolve most electromorphs at a highly polyallelic locus). Therefore, we predict that isozyme characters will have lower consistency or retention indices compared to allozymic data but not necessarily lower than DNA sequence data that also must deal with a severely limited number of states.

### 2.3.3. Intraspecific variation; high or low?

There has been little documentation of the extent and nature of intraspecific variation in isozyme characters. In general, intraspecific variation is rarely reported and has been assumed to be an insignificant factor. If so, isozyme characters will not be useful in population-level studies. On the other hand, we are aware of at least one example of such variation that must serve both as a caveat and an indication of potential: In a survey of isozyme characters among 60 prairie rattlesnakes, *Crotalus viridis viridis*, some mutually exclusive tissue-specific gene expression patterns were observed (Murphy, Aguirre, Morafka, and Scott, unpublished data). In some specimens, a product of a given locus might be expressed in kidney whereas in other specimens it was expressed only in liver. Only through the comparison of large numbers of individuals could such variation be uncovered. However the data yield per unit invested time drops rapidly when such laboratory experimentation and fieldwork (when tissue samples must be prepared and maintained in isolation from one another) are required.

### 2.3.4. The need for comparable semaphoronts sensu stricto

Wiley (1981) defined comparable semaphoronts as “individual specimens at similar stages in their life history” and discussed the need for limiting comparisons to semaphoronts in systematic studies. For isozyme characters, comparability must be strictly observed. The reduction of enzyme activity in prolonged frozen storage limits valid comparisons to those specimens treated similarly. Frozen tissue collections at various museums may be too vulnerable to this variable, as well as too limited in available tissue array, to be of value in such comparisons.

Varying physiological states, specifically pathology, can affect the results of certain forms of isozyme data. The effect of mild starvation on tissue-specific gene expression in prairie rattlesnakes was evaluated (Murphy, Aguirre, Morafka, and Scott, unpublished data). Three groups of animals were compared: (1) animals immediately euthanized upon receipt from the field, (2) those maintained in captivity and frequently fed mice, and (3) those provided with water only for a few weeks. Among the maintained snakes, all fed snakes gained weight, and all starved snakes lost weight. Using serial dilutions to quantify isozyme expression (Kettler et al., 1986), Murphy et al. observed
statistically significant different levels of tissue-specific enzyme expression among the three groups of snakes for many enzyme systems. Specimens euthanized upon arrival had the highest levels of isozyme expression. The fed snakes fared second best. The starved snakes expressed little or no discernible activity for several enzyme systems. These findings bring into question the validity of results obtained from organisms not treated in an identical manner.

2.3.5. A labor-intensive approach

Whereas in terms of expendable supplies, isozyme characters might seem to be a fiscal bargain, but it is rather the case of spending more time than money to resolve these traits. Additional primary fieldwork should be expected because the desired complete array of different tissues will not have been deposited in museums’ frozen tissue collections by other investigators. As more of these collections become geared toward DNA studies, the frozen tissue arrays can be expected to become less diverse. In an era when the resolution of products of more than 50–100 enzyme systems can be considered “standard” in an allozyme laboratory, multiplying the lab work times 6–10 tissues can increase data-generation time considerably. If one adds the dimension of different developmental stages, such isozyme characters become a truly labor-intensive endeavor.

3. Conclusions, speculations, and recommendations

Isozyme characters can contribute to systematic studies but it is very important to recognize the level of universality at which they might be appropriately applied. Unacceptable levels of homoplasy should be expected when one exceeds the taxonomic limits of application within particular groups of organisms (e.g. Matson, 1998). At the very least, isozyme characters can contribute to the biochemical identification of organisms.

We anticipate that some isozyme characters will ultimately find much greater utility than others (see Crother et al., 1998). For example, even among the diploid squamate reptiles, gene duplications have been reported for the families Phrynosomatidae, Eublepharidae, Scincidae, Xenosauridae, Varanidae, Typhlopidae, Colubridae, Elapidae, and Viperidae (Lawson and Dessauer, 1979; Murphy and Ottley, 1980; Murphy and Crabtree, 1985; Sites and Murphy, 1991; Mink and Sites, 1996; Murphy, 1998). In some cases, these duplications appear to be shared among several members of families (e.g. Lawson and Dessauer, 1979), whereas in other cases they appear to be restricted to single species or species groups (e.g. Sites and Murphy, 1991; Mink and Sites, 1996; Mendoza-Quijano et al., 1998). Although such characters appear homoplastic (Sites and Murphy, 1991), they may be so conserved as to serve as unambiguous synapomorphies. Given that these isozyme characters are easily gathered in standard electrophoretic surveys, their application should be recognized and continued. We must, however, insert a caveat regarding the use of the literature in such comparative studies. Most allozyme surveys did not have as their goal the resolution of all loci in a given system, nor did they include a rigorous investigation of
tissue-specific expression. Thus, reports of the expression of a single locus in a particular enzyme system should not be taken to mean that other paralogous loci are not expressed in that organism. Only studies specifically designed to address gene expression should be used for comparison.

The tissue-specific gene expression characters have the greatest potential to offer large amounts of isozyme data (e.g. Buth and Rainboth, 1998; Crother et al., 1998). We highly recommend that future studies that use such characters employ the serial dilution method of Kettler et al. (1986) to quantify levels of enzyme activity for comparison.

Post-translational modification and heteropolymer formation may be characters that are encountered fortuitously in allozyme surveys. As such, these characters are “free for the taking” if noticed by the investigators. However, neither of these isozyme characters should be expected to routinely contribute as a separate, robust data set (but see Lieb et al., 1998).

The future? Gene expression in ontogeny may be the most fertile area for future systematic applications. Additional isozyme characters may be recognized. Various coding schemes for current isozyme characters need to be tested. Although some investigators may feel that the direct probing of transcripts from specific tissues results in a more precise assessment of gene expression and that direct sequencing of DNA/RNA results in a more accurate determination of genetic variation, allozyme and isozyme characters will likely continue to make significant contributions to systematic investigations. The relative low cost and ease of resolution of allozyme and isozyme characters, plus the biparental nature of allozyme inheritance, make these characters especially appropriate for many kinds of studies.

Acknowledgements

The first author credits Dr. Gregory S. Whitt for introducing him to the concept of isozyme characters over 25 years ago. Many discussions over the years with Dr. Whitt and Dr. Stephen D. Ferris have proven to be quite beneficial. This study was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada Grant A3148 and the National Institutes of Health Grant RR08156-10 to R.W.M. This is contribution 133 from the Centre for Biodiversity and Conservation Biology, Royal Ontario Museum.

References


