Entomology Serving Society: Emerging Technologies and Challenges

Entomological Society of America
Centennial National Symposium

S. BRADLEIGH VINSON, Editor
Department of Entomology, Texas A&M University,
College Station, Tex.

ROBERT L. METCALF, Editor
Department of Entomology, University of Illinois,
Urbana, Ill.


Entomological Society of America, Lanham, MD 20706
Contents

Preface
   S. Bradleigh Vinson .............................................. vii

Introduction to the Symposium "Entomology
Serving Society: Emerging Technologies and Challenges"
   Robert Metcalf ................................................. 1

The Land Use Controversy:
A Resolvable Dilemma
   Michael E. Irwin .............................................. 7

Semiochemists and Allelochemists: A Wondrous
Diversity of Relationships with Insects
   Murray S. Blum ............................................ 11

The Significance of Diversity:
New Challenges for the Entomologist
   Terry L. Erwin .............................................. 25

The Conservation Status of Insects:
Mass Extinction, Scientific Interest,
and Statutory Protection
   William J. Boecklen ..................................... 40

Molecular Baculovirology:
From Genes to Strategies
   Lois K. Miller ........................................... 58

Molecular Genetics Applied to Systematics
   Guy L. Bush ............................................... 86
Molecular Genetics Applied to Systematics

Guy L. Bush
Department of Zoology, Michigan State University, East Lansing, MI 48824–1115

SYSTEMATIC BIOLOGY HAS COME a long way since the early days of classification which culminated in the eighteenth century with Carl Linnaeus's (1707–1778) insistence on the immutability and clear-cut distinction of species. Modern systematists and evolutionary biologists now accept the fact that species are in a constant state of evolutionary flux and that it is often impossible to establish species boundaries. Our goal as systematists is to provide the taxonomic and evolutionary framework on which all biological science relies for making order out of the diversity of life and for comprehending the relationships among organisms. As Simpson (1961) noted, "systematics is the scientific study of the kinds and diversity of organisms and of any and all relationships among them."

Modern Systematics, therefore, draws on all biological disciplines to achieve this goal. Such diverse fields as paleontology, morphology, development, physiology, behavior, ecology and genetics have contributed in one way or another to many good systematic studies. In recent years molecular biology, a field that focuses on the structure and function of proteins and nucleic acids such as DNA and RNA, has brought a new perspective to systematics and our perception of the evolution of life. Its power
as a systematic tool stems from the fact that molecular evolution is a historic process whereby genes accumulate changes as a result of stochastic and selective processes. Protein and DNA sequences document these changes and thus, in a sense, represent "fossil records" from which past evolutionary change and phylogenetic affinities can be reconstructed with knowledge of only their present-day structure.

Early attempts to reconstruct the evolutionary history of organisms compared protein sequences or involved techniques, such as microcomplement fixation, to indirectly assess differences in protein structure (Wilson et al. 1977). Although comparing amino acid sequences provides a useful phylogenetic perspective, they do not have the resolving power of sequences derived directly from DNA coding for the same proteins themselves. Not only are most amino acids redundantly represented by two to four triplet DNA codons, but the genes themselves include introns and regulatory regions that are not expressed as a gene product. For this reason DNA has become the molecule of choice for most systematic and evolutionary studies. I will, therefore, focus the rest of my talk today on the ways DNA can be put to use in insect systematics.

DNA has become the quintessential tool for resolving systematic and evolutionary problems in insects for diverse reasons (Powell in press):

1) The size of the data set in each species is large so information content is very high, with the number of independent characters limited only by the number of nucleotide base pairs in its DNA.

2) Nucleotide sequences are linearly arranged; organization is simple rather than multidimensional so analysis is objective and the need for weighting is minimized.

3) Genetic differences between taxa can be established without recourse to Mendelian genetics. Genetic comparisons can therefore be made between distantly related organisms.

4) DNA is less subject to direct environmental effects than other traits such as those based on morphology and physiology. Thus, the extent of nonheritable variation is minimized or nonexistent. Sequence data can therefore be easily compared between laboratories.

5) Of particular importance to the systematist is the fact that DNA is constantly changing by mutation; sequences appear
to evolve in a clocklike fashion, albeit a very sloppy one, but one that if used judiciously provides estimates of relative divergence times between taxa which lack a fossil record.

6) Many DNA sequences are also phylogenetically continuous through time. This lack of phylogenetic limits provides a record of evolutionary history from very recent to ancient times (Hill 1987).

7) Finally, specific genes and sequences vary greatly in their rate of base-pair evolution; conserved areas can be used for comparing distantly related taxa. Variable regions can be used to compare more closely related taxa (Simon et al. 1990).

There are basically two kinds of DNA used in most insect systematic research, mitochondrial DNA (mtDNA), found only in mitochondria and nuclear DNA (nDNA) of the nucleus. The mtDNA is a small (15kb–20kb), maternally inherited, single stranded circular molecule which is easily extracted and cloned. Its gene content, which is highly conserved across all animal groups, codes for only a very limited number of genes. Although the 13 proteins, 22 tRNAs and 2 rRNAs found in the Drosophila yakuba mitochondrion (Wolstenholme and Clary 1985) and other insects represents only a miniscule portion of the entire insect genome, it provides a unique and extremely useful source of information for taxonomic and evolutionary studies. Moritz et al. (1987) have discussed the uses (outlined below) and limitations of mtDNA in systematic and evolutionary studies. Because it is maternally inherited it provides information on any sexual bias in dispersal and serves as a source of genetic markers for establishing the pattern of matrilineal gene flow.

MtDNA is widely employed to analyze the population structure and biogeography of insects and other eukaryotic organisms. It has been particularly useful for studying the dynamics of hybrid zones. From a systematic standpoint mtDNA has proved invaluable in the reconstruction of phylogenetic relationships among closely related and clonally reproducing taxa to date historical events. However, it is generally inappropriate for phylogenetic analysis of categories above the genus level as most regions of mtDNA evolve rapidly. Also introgression of mtDNA from one species to another presents problems as it has been found to flow across closely related species boundaries.

With the exception of cases which take advantage of the unique maternally inherited qualities of mtDNA, most of these
examples also pertain to nuclear DNA (nDNA) as well. Nuclear DNA, which represents the bulk of the genome, is a much more complicated long, double stranded and open-ended molecule which contains a continuum of nonrepetitive, moderately repetitive and highly repetitive DNA sequences. Because of the number and complexity of sequence arrangements in eukaryotic chromosomes, nuclear DNA offers a much wider range of possibilities for resolving phylogenetic and evolutionary problems at every level of divergence from the level of the population to the class and even phylum. As in mtDNA conserved regions can be used for studying relationships among distantly related taxa while variable regions are suitable for studying relationships among closely related taxa.

There are three basic techniques for assessing DNA sequence divergence among taxa: (1) DNA/DNA hybridization, (2) restriction site mapping with restriction fragment length polymorphisms (RFLP’s), and (3) DNA sequencing. These techniques differ in the nature of the information they provide and in the portion of the genome that can be assessed. The first is an indirect method for assessing the degree of sequence evolution while the others provide direct, unambiguous sequence data.

**DNA/DNA Hybridization.** DNA hybridization was one of the first methods used to assess quantitatively divergence at the DNA level between taxa. It is based on the fact that the thermal stability of DNA duplexes is determined by proper base pairing, i.e. A with T and G with C. Single copy DNA (scDNA) is now used exclusively for DNA hybridization studies. By raising and lowering the temperature double-stranded scDNA can be made to melt into single strands or reanneal into a duplex. The melting temperature ($T_m$) will decrease as the mismatching between base-pairs increases. When DNA from one taxon is hybridized with that of another and the $T_m$ recorded, the percent of base-pair mismatch can be determined from the change in median melting temperature or $DT_m$. Caccone et al. (1988b) have found that a $1^\circ$ $DT_m$ corresponds to 1.7% bp mismatch in *Drosophila*. For the most part the relationship is linear so that the $DT_m$ is a valid overall measure of scDNA divergence.

Some sections of single copy DNA in insect genomes appear to be evolving at extremely high rates with the overall rate estimated to be some five to ten times faster than found in most vertebrates (Caccone and Powell 1990). Quite high levels of scDNA divergence may even exist between taxa that are morphologically
and chromosomally very similar and that may be able to form interspecific hybrids. Caccone et al. (1988a) have derived a phylogeny for the Drosophila melanogaster subgroup which provides an example of the results to be expected from this technique. The UPGMA dendrogram, based on scDNA, is congruent with chromosomal data, morphology, and behavior.

There are several advantages as well as problems in using DNA hybridization to assess relationships among taxa (Springer and Krajewski 1989). Because of the high rates of DNA evolution in insect genomes, the technique can be used to establish relationships from the interspecific to the ordinal level. Usually only one sample from a single population is sufficient to establish accurate estimate of species relationships. ScDNA comparisons are based on millions of nucleotide base-pairs. Thus a measure of differentiation is taken using a very large number of sites across the entire genome. Rates of divergence appear to be more clocklike than estimates derived from other methods (RFLP's and sequencing).

Heterozygosity, however, may be sufficiently high to obscure differences between diverging populations, races, semispecies and sibling species. The technique also reveals nothing about population dynamics or the process of speciation as changes in individual loci involved in the adaptive process can not be examined. Another drawback stems from the fact that this approach does not employ discrete character data. Only a single measure of total genetic differentiation between taxa is obtained so it is inappropriate for analysis by cladistic parsimony. Gene duplications and deletions may also cause problems in interpretation and biased sequence sampling and homoplasy may distort distances unless corrections can be made. Finally, DNA hybridization cannot index sequences that are more than 20 to 25% mismatched; this may distort distances as taxa become distantly related unless corrections can be made.

**Restriction Enzyme Mapping.** Restriction enzyme mapping is a second, relatively simple, method for examining sequence variation which provides more precise estimates of divergence between taxa within specific regions of the insect genome than DNA hybridization. Different restriction endonucleases which cleave double-stranded DNA at specific palindromic sequence sites are used to examine variation within a random or specific subset of nuclear or mitochondrial DNA (see
Kreitman and Aguadro 1986; Harrison et al. 1987; Simon 1988; Wheeler 1989 for examples). A battery of restriction enzymes having 4- or 6-base recognizing sequences is used to digest or cut genomic DNA at specific sites. The objective is to identify a sufficient amount of nucleotide variation to specify the identity of a specific fragment and to make reasonable inference about phylogenetic relationships and population structure. Once restriction sites are identified they can be mapped and the information used to derive a phylogenetic estimate using various parsimony methods (Simon 1988).

Restriction site analysis allows one to choose an appropriate sequence or gene to resolve a particular problem. Some regions of the genome are conservative and nucleotide substitutions occur so slowly that they can be used to reconstruct truly ancient events while other regions, such as intergenic spacers, may evolve so rapidly that individuals are polymorphic. An example of how this technique can be used to probe the genetic structure of a hybrid zone and to establish the status of closely related species of crickets, *Gryllus pennsylvanicus* and *G. firmus*, is exemplified in a genetic study of a mosaic hybrid zone by Rand and Harrison (1989). They combined results from restriction mapping of mtDNA with allozyme data and hybrid crosses to demonstrate that the species boundary is permeable, but the degree of permeability varies from one genetic marker to the next as well as with the ecological and geographic context of species interaction.

In the Connecticut portion of the hybrid zone *G. pennsylvanicus* is found inland and *G. firmus* along the coastal plain. The hybrid zone is a mosaic rather than a cline of strikingly differentiated populations which combine genotypes of the two species in disparate proportions, which demonstrates the semipermeable nature of the species boundary between these two taxa. Gene flow in this case appears to be strongly influenced by habitat preference as *G. firmus* genotypes predominate in the sandy areas and *G. pennsylvanicus* in loamy soils.

Restriction site mapping provides sufficient resolving power of sequence variation to reconstruct the evolutionary history of specific alleles or segment of a chromosome within a species. Aquadro et al. (1986), have used variation in the restriction map of a 13-kb region of chromosome II which includes the alcohol dehydrogenase structural gene (Adh) in a phylogenetic analysis of
29 restriction site haplotypes observed among 49 lines of *D. melanogaster*.

**DNA Sequencing.** Although restriction site mapping is a relatively recent addition to the molecular systematists' tool box, its current wide use is rapidly being replaced or supplemented by a new approach that permits rapid and easy sequencing of DNA segments (Wilson et al. 1989). Complete nucleotide sequences for specific segments, genes or longer stretches of the genome have always been regarded as the most reliable source of information for resolving a wide range of systematic and evolutionary problems. Until recently obtaining complete sequences was extremely time consuming and very costly. Conventional sequencing required the construction and screening of clone libraries from many individuals. Comparative sequencing on a large scale therefore was not a practical alternative to restriction analysis and only a few attempts were made to put sequence data to use in phylogenetic studies prior to 1986.

This situation has dramatically changed. The recent development of the polymerase chain reaction (PCR) makes it possible to sequence directly subsections of a DNA molecule after amplification without resorting to clone libraries (Saiki et al. 1985; Mullis et al. 1986; Wrischnik et al. 1987). Amplification of DNA segments is carried out using two synthetic oligonucleotide primers, each about 25 bases long, the thermostable *Thermophilus aquaticus* (*Taq*) DNA polymerase, and four deoxyribonucleotide triphosphates. The first primer matches part of a DNA strand at one end while the second primer matches the sister strand at the other end. Since the 3' ends of the two primers point towards each other, repeated cycles of heating and cooling lead to a chain reaction and an exponential synthesis of many copies of the specific segment bounded by the two primers (see White et al. for a review). Only picogram amounts or even single molecules of DNA are needed to obtain microgram quantities of a highly specific region of genomic DNA which can be sequenced directly or subjected to RFLP analysis on a scale necessary for phylogenetic analysis. Direct sequencing data tend to be less ambiguous and more informative than data from restriction mapping (Kocher et al. 1989).

The PCR can be adapted to help resolve a wide variety of systematic and evolutionary problems. Primers corresponding to highly conserved areas of the genome can be used to initiate the
sequencing of unstudied distantly related taxa that have no previous sequence information (Kocher et al. 1989; Simon et al. 1990). Using highly conserved PCR primers from humans and *Drosophila*, Simon et al. (1990) have been able to amplify and sequence a 326 bp segment of the small mitochondrial ribosomal RNA gene (12S, Domain III) of the 13-year periodical cicada (*Megicicada tredecium*). These 12S primers, as illustrated in Figure 1, have sequences which are conserved across eukaryotic, bacterial, plasmid, and mitochondrial small rRNAs in 103 out of 106 species examined. Many such "universal" priming regions exist, so that it is possible to amplify almost any segment of a mtDNA or nDNA (Kocher et al. 1989). The only requirements for the construction of such primers are reference sequences from two or more widely diverged organisms. Because the intervening sequences between the primers are variable to a greater or lesser degree, depending on the sequence being examined, they are versatile sources of phylogenetic information. Furthermore, once one section of the genome has been sequenced it provides a means of sequencing neighboring regions by employing the inverse PCR technique (Ochman et al. 1988). Wheeler's (1989) use of ribosomal DNA (r-DNA) to resolve the evolutionary status of higher taxonomic categories within the Paleoptera, Eumetabola, and Holometabola, provides an excellent example of how the PCR can be applied in conjunction with other molecular methods, such as restriction site mapping, for phylogenetic studies.

Probably the most significant contribution the PCR technique has given to the systematic world is that the extensive collections residing in our museums represent an incredibly rich source of material for molecular systematic and evolutionary research. Specimens that are hundreds of years old can be sequenced, thus making it possible to follow gene frequency changes through time (Kocher et al. 1989). Museums are often the repository and the major source of organisms that have long gone extinct or that are extremely rare or unusual. They represent a veritable gold mine of genetic material for the modern systematist and evolutionary biologist as well as for the animal and plant breeders in search of unique genotypes. Museums therefore have taken on a new and extremely important role in conserving the genetic material of organisms that either are or soon will be extinct. Because such a small amount of DNA is needed to initiate the PCR genetic and evolutionary studies can be made on
San Antonio Centennial Proceedings

12SA Primer

Hs 5' A A A C T G G G A T T A G A T A C C C C A C T A T T 3'
Dy . . . A . . . . . . . . . . . . . . . . . . . . . . . . T . T . 
Ec C . . A A A . . . . . . . . . . . . . . . . . . . . . . . . G

12SB Primer

Hs 5' G A G G G T G A C G G G C G G T G T G T 3'
Dy A . A . A . . . . . . . . . . . . . . . . . . . . . . . . A . 
Ec T G . T . . . . . . . . . . . . . . . . . . . . . . . . . .

Figure 1. Sequences of 12SA and 12SB primers, based on conserved regions, modeled after Homo sapiens mtDNA (Hs) and compared to homologous positions in Drosophila yakuba (Dy) and Escherichia coli (Ec) (from Simon et al. 1990).

exceedingly small organisms or on organisms that can not be reared in the laboratory. Finally, it will have a profound effect on the fields of ecological, behavioral, and physiological genetics.

I have said nothing about how sequence data may be treated to test phylogenetic and evolutionary hypotheses. Inherent limitations in using DNA markers for phylogenetic and evolutionary studies, such as the imprecision of a molecular clock, the existence of transpecific polymorphisms, and the difficulty in comparing certain sequences because of misalignment, insertions, and deletions are currently yielding to advances in data collection, processing and tree construction (Felsenstein 1989; Moritz et al. 1987). There are also increasingly sophisticated programs available for the personal computer like Dave Swofford’s PAUP (Platnick 1988) which greatly facilitate the analysis of sequence data for phylogenetic reconstruction and hypothesis testing.

Breakthroughs in sequence technology, such as selective amplification of specific DNA sequences using the polymerase chain reaction or direct sequencing of ribosomal RNA, are bringing these powerful molecular tools well within the reach of many systematists. Molecular systematics has also begun to reopen and resolve old controversies on the evolutionary affinities of organisms sometimes in unexpected ways that only a few years ago seemed unresolvable. One of the surprises has been that divergence between several closely related species may occur almost synchronously in time so that it is impossible to resolve phylogenetic relationships and the order of speciation. Phylogenetic
reconstructions look more like a bush than a tree with bifurcating branches (Felsenstein 1988). In other cases a taxon which was thought to be a single species turns out to represent a complex of morphologically indistinguishable species, some with genetic distances that suggest an ancient separation. Undoubtedly we are in store for many more such surprises. There is no question that molecular biology is breathing new life into one of the oldest and most important fields of biological study. Nevertheless, I would like to emphasize that although these advances in molecular biology represent a new and higher level of analytical sophistication they are, in fact, only tools that may help resolve interesting biological problems uncovered as a result of fieldwork, natural history studies, and laboratory research by dedicated systematists and evolutionary biologists.

References Cited


