Lagascalia 19(1-2): 317-322 (1997)

PROTEINS IN PLANT SYSTEMATICS

GIOVANNI CRISTOFOLINI

Abstract

Protein analysis has been intensively used for about 40 years as a tool for elucidating systematic relationships. The rationale was that: (a) proteins are supposed to yield a first-hand information on the genome, (b) homology between corresponding proteins of different (even systematically very distant) organisms can be defined, and (c) the degree of similarity between homologous proteins of different organisms can be quantified. During the last decade an increasing mistrust was expressed by many authors about protein data, under the assumption that the information on nucleic acids overcomes and encompasses any phenotypic (incl. protein) information. Such a view is justified as long as the neo-darwinian theory on the genesis and inheritance of the biological diversity holds true. Some of the tenets of neo-darwinism are challenged nowadays, as more and more evidence is accumulating on alternative ways of genesis and transmission of heritable characters. Recent findings on Retroviral-like transposable elements, on epigenetic inherited systems, on inheritable acquired characters, on the function of symbiosis and mutualism in plant evolution, indicate that the organism as a whole is the real subject of evolution. The systematic value of phenotypic information , and in particular of the information stored in the protein structure, must be re-examined on the light of such findings. The necessity of a multiple approach to plant phylogenetic systematic is stressed.

The present report deals with the experience and the perspectives of protein analysis applied to plant phylogenetic systematics, with special regard to the problems related to macro-evolution. It will not cover the status and the achievements of isozyme analysis, a well established methodology in plant biosystematics, whose legitimacy is not questioned today by any biologist.

Macromolecular systematics has arisen and grown up for about 40 years based upon some basic assumptions. The main argument in favour of it has been that macromolecules, as a more or less direct product of the genome, are supposed to give a first-hand information on the genome itself.

A further argument was proposed in favour of considering macromolecules in biological systematics. Indeed, as it was appropriately pointed out (HILLIS, 1987), few morphological characters are shared among major groups of organisms (eukaryotes versus eubacteria, for instance); in contrast, biomolecules provide a phylogenetic record from very recent time to the origin of life on Earth, because of the size and the diversity in rates of change of different portions of the genome.

Early studies on amino acid sequences in functionally homologous proteins in different species seemed to indicate that structural differences were accumulating at constant rates (MCDONALD, 1990). One interpretation was that the accumulated nucleotide substitutions in homologous proteins were adaptively equivalent and thus selectively neutral. The theoretical basis of this interpretation was the mathematical demonstration by KIMURA & OHTA (1971) that the amino acid composition of a protein sequence would be expected to evolve at a pace determined by the mutation rate, and

thus be essentially constant. In this context, protein studies were largely used from the years sixties onward, as soon as effective techniques of protein purification and analysis were achieved.

Serological systematics also got some space within the frame of protein systematics, especially during the early times, because of its capacity of yielding information on many proteins extracted from many taxa within a short time, with an easy technology and at reduced costs. A further advantage of sero-systematics is that it requires reduced amounts of plant material: hence, even rare or endangered plant species can be studied without significantly affecting the consistence of natural populations. In contrast, other techniques (e.g. amino acid sequencing) usually require large amounts of plant tissue; this is one of the reasons why most plant protein data concern spinach, corn, and a few other vegetables. Otto Moritz in Kiel and David Fairbrothers in New Brunswick were among the pioneers of serological techniques applied to plant systematics (see e.g. FAIRBROTHERS, 1977; MORITZ, 1964).

From the years eighties onwards an increasing criticism was raised against the systematic use of protein data in general, and against serological data in particular. The main points of criticism against serological systematics were that:

The immunological reactions provides but an indirect knowledge on the proteins involved; the measure of the immune-precipitate is a rough figure, that does not give any qualitative information on the sequence involved.

There is no linear proportionality between the amount of immune-precipitate and the degree of homology in the amino acid sequence.

The results are not fully reliable, due to possible unforeseen reactions that occur in parallel with the main antigenic reaction.

Such criticisms were certainly soundly based, and should be carefully considered when using serological data for systematic-phylogenetic inferences. But a further criticism was directed toward the use of proteins in systematics in general: indeed, it was asserted that in any case the information on DNA would be more reliable, and systematically more significant, than any information on proteins. The theoretical basis for this opinion is that DNA sequence would be primary with respect to protein sequence, and therefore DNA-derived phylogenies would be more soundly based than phylogenies inferred from any other set of characters. Such a belief, in its turn, relies upon the tenet of the unidirectionality of the deterministic flow "DNA - RNA - proteins - phenotype". The protein sequence would be the mere result of the DNA sequence, plus a number of epigenetic events, which would act just as a "noise" over the transmission of the message.

Such extremely simplified scheme has been challenged for many years, and is inadequate, to say the least, if the complexity of genetic control and of evolution is considered.

To recall only some of the achievements that have enriched our knowledge in recent times, mention should be made of the Retroviral Like Elements, that seem to be a channel of information transfer from the cellular environment to DNA: such elements may have a function in bringing about rapid and dramatic changes in gene regulation and development (MCDONALD, 1990, and elsewhere).

A further elements concerning the complexity of the evolutionary process is given by recent discoveries on the Epigenetic Inherited Systems: TORDERA & al. (1993) discussed the molecular basis of inheritable epigenetic elements, and suggested that histones may play a primary role. For an up-dated review see JABLONKA (1995).

The heritability of acquired changes has been experimentally studied in recent years; exhaustive reviews of the present knowledge have been published by Landman (1991, 1993) and by Jablonka & al. (1992).

CAMPBELL (1988), and more recently BENCI & GALLENI (1995), have discussed how sudden changes may occur in biological systems, due to the instability which is intrinsic in their complexity. This aspect has been also treated by SARA' (1989, 1993), who stressed the ìholisticî character of the evolutionary process. The role of self-organisation in generating ordered structures in complex systems has been the object of speculations for many years; this most promising field has been extensively treated by KAUFFMAN (1993), who discussed its theoretical bases and perspectives.

Last but not least, the relevance of symbiosis in determining sharp evolutionary changes has been highlighted by MARGULIS (1981). The concept of symbiosis as the key-step in the origin of eukaryotic cells has been accepted by most biologists, and yet the theoretical implications of such a fact do not seem to have been fully understood.

What we know today is that evolution is a tremendously complex process, that cannot be reduced to the trivial mechanism of an unidirectional flow of information from the DNA to the phenotype, with the environment just acting as phenotype selector.

It is universally accepted that the evolution of an ecosystem through time is the result of an indefinitely complex network of interactions among biotic and abiotic elements. Perhaps, the organismic evolution should be regarded also as the evolution of a system, where the macromolecular species interact among themselves and with the cellular environment in a complex system of actions and re-actions.

Some conclusion of systematic relevance can be derived from the premises discussed above. First, if genic, genomic and chromosomal mutations are not *the only origin* of biodiversity, and the inheritable information *is not limited* to DNA sequence, then any macromolecular information is systematically and phyletically significant; moreover, if nuclear DNA, chloroplast and mitochondrial DNA, other nucleic acids in the cell, nuclear and cytoplasmic proteins, are *interacting systems*, and the extent to which each of these systems is inheritable is not fully known, then no one type of data definitely overcomes all others; no one type of data includes and explains (neither actually nor potentially) all others.

Within this frame, protein systematics, and even serological systematics, can still give some relevant contribution to the understanding of phyletic relationships. As it was recalled at the beginning, a major advantage of protein systematics relies in this, that conservative proteins can be detected in organisms belonging to different systematic families, orders, or even classes, and provide information on the relative distance. A case study is that of a class of major reserve proteins, the so-called Legumin-like proteins (L-l proteins). The L-l proteins are 11s globulins whose molecule is formed by six units (monomers), each of them being composed by two polypeptides, the one acidic and the other basic. After L-l proteins were first detected and described in the cotyledons of Leguminosae (DERBYSHIRE & al., 1976), proteins with the same

characteristics were described in a great many families of Angiosperms. Even proteins that had previously described under different names proved to share the same structural characteristics (see, for instance, JENSEN, 1984). JENSEN & BERCHTOLD (1989) could demonstrate that a non-soluble protein isolated from the endosperm of Conifers has the same structure as the L-l proteins, and CONTE (1994a) gave evidence of the presence of a protein with the same properties in the endosperm of *Gnetopsida*. Obviously, the rough demonstration of the presence of a protein, whose mass, charge, and subunit composition corresponds to those of the Legumin could not be assumed as an evidence for the presence of an homologous protein. The serological cross-reactivity proved useful in this case to detect homology, and even to give a rough estimate of the degree of correspondence. So, CONTE (1994b) could demonstrate the homology of the reserve protein of Gnetopsida with that of Angiosperms, and prove the absence of the same protein from Cycads, and JENSEN & al. (1994) use the cross reactivity of the same protein to elucidate the systematic relationships of *Euphorbiaceae*.

The presence and the degree of similarity of an homologous protein in different taxa is the result of genetic and epigenetic factors; therefore, it does not necessarily reflect an exactly corresponding similarity of the underlying genes, and is not less informative. Information derived from nucleic acids analysis and from protein analysis are not necessarily tautological: ALBERT & al. (1994) discussed a case study, where the consensus phylogenetic tree of land plants resulting from the parsimony analysis of rbcL nucleotide sequence substantially diverges in its topology from the consensus tree resulting from the parsimony analysis of the amino acid sequences of the corresponding protein.

It is out of any doubt that macromolecules - mainly DNA, but RNA and Proteins as well - are invaluable tools in biological systematics, because they allow homologous characters to be detected and compared, and similarities to be quantified. Yet, their intrinsic value in reconstructing phylogenies is by no means higher than that of any other genetically fixed character, as long as we do not know more about the biology of evolution. Macromolecular data should be used to corroborate or to refuse phyletic hypotheses, rather than to construct self-sufficient phyletic trees. No one single type of data can produce directly phylogenetic trees, nor phylogenetic classificatory systems.

Evolutionary biology is at a phase of its own evolution, where many tenets are challenged. There is not, at the present moment, a generally accepted theory to explain all aspects of biological evolution. Consequently, phylogenetic systematics is deprived of the sound theoretical basis on which it rested twenty years ago. Under such conditions, a pragmatic approach to the problems is necessary: different sources of information and different techniques should be chosen depending on the particular problem, on the source material available, on the practical opportunities. The more polymorphic the approaches, the more informative shall be the resulting synthesis.

References

ALBERT, V. A. & al. (1994). Functional constraints and rbcL evidence for land plant phylogeny. *Ann. Missouri Bot. Gard.* **81(3)**: 534-555.

BENCI, V. & L. GALLENI (1995). Stability and instability in evolution. *Quad. Istit. Matem. Appl. Univ. Pisa* **1**: 1-11.

CAMPBELL, J. H. (1988). Evolution as non-equilibrium thermodynamics. In B. H. WEBER & al. (eds.), *Entropy, Information and Evolution*. Harvard Univ. Press, Cambridge, Mass.

CONTE, L. (1994a). Storage proteins in Gnetopsida. 1. Recognition of legumin-like proteins. *Giorn. Bot. Ital.* **128**: 839-843.

—— (1994b). Storage proteins in Gnetopsida. 2. A serotaxonomic study of legumin-like proteins. *Giorn. Bot. Ital.* **128**: 845-850.

DERBYSHIRE, E., D. J. WRIGHT & D. BOULTER (1976). Legumin and Vicilin, storage proteins of Legume seeds. *Phytochemistry* **15**: 3-24.

FAIRBROTHERS, D. E. (1977). Perspectives in plant serotaxonomy. *Ann. Missouri Bot. Garden* **64**: 147-160.

HILLIS, D. M. (1987). Molecular versus morphological approaches to systematics. *Ann. Rev. Ecol. Syst.* **18**: 23-42.

JABLONKA, E. (1995). *Epigenetic inheritance and evolution*. Oxford University Press, New York.

——, M. LACHMAN & M. J. LAMB (1992). Evidence, mechanism and models for the inheritance of acquired characters. *J. Theor. Biol.* **158**: 245-268.

JENSEN, U. (1984). Legumin-like and Vicilin-like storage proteins in Nigella damascena (Ranunculaceae) and six other dicotyledonous species. *J. Plant Physiol.* **115**: 161-170.

—— & H. BERCHTOLD (1989). Legumin-like proteins in Gymnosperms. *Phytochemistry* 28: 1389-1394.

——, I. VOGEL-BAUER & M. NITSCHKE (1994). Legumin-like proteins and the systematics of Euphorbiaceae. *Ann. Missouri Bot. Gard.* **81**: 160-179.

KAUFFMAN, S. A. (1993). *The origins of order. Self-organization and selection in evolution*. Oxford University Press, New York.

KIMURA, M. & T. OHTA (1971). Protein polymorphism as a phase of molecular evolution. *Nature* **229**: 467-469.

LANDMAN, O. E. (1991). The inheritance of acquired characteristics. *Ann. Rev. Genet.* **25**: 1-20

—— (1993). Inheritance of acquired characteristics revisited. *Bioscience* **43**: 696-705.

MARGULIS, L. (1981). Symbiosis in Cell Evolution. W. H. Freeman, San Francisco.

MC DONALD, J. F. (1990). Macroevolution and retroviral elements. *Bioscience* **40(3)**: 183-191.

MORITZ, O. (1964). Some special features of serobotanical work. In C. A. LEONE (ed.) *Taxonomic biochemistry and serology*: 275-290. Ronald Press, New York.

SARA', M. (1989). The problem of adaptation: an holistic aproach. *Riv. Biol.* **82**: 75-101.

—— (1993). L'evoluzione biologica: un approccio olistico centrato sull'organismo. *Riv. Biol.* **86**: 281-289.

TORDERA, V., SENDRA, R. & E. PEREZ-ORTIN (1993). The role of histones and their modifications in the informative content of chromatin. *Experientia* **49**: 780-788.

Address of the Author:

Prof. G. Cristofolini, Dipartimento di Biología, Università di Bologna, Via Irnerio 42, 40126 Bologna, Italy.