TYPES OF DNA USED IN SPECIATION AND PHYLOGENY STUDIES

LAURA BUBURUZAN[1]*, LUCIAN D. GORGAN[1], ION I. BĂRA[1]

Key words: mtDNA, microsatellite DNA, scnpDNA, phylogeny

Abstract: The present paper represents a synthesis of the main types of molecular markers used in contemporary phylogeny and phylogeography studies. Our purpose is also to reveal the recent discovered role of nuclear DNA polymorphic loci in the studies of filiation.

INTRODUCTION

In most of the studies that use molecular markers, the predominant classes are mitochondrial DNA and microsatellite nuclear DNA. Analyzing the phylogeographical studies (Avise, 1998), it came out that 70 % from all these use the analysis of the animal DNA. In the same manner, the majority of the research projects that use molecular markers coming from the nuclear DNA, involved microsatellite DNA.

Zhang and Hewitt (2003) studied 1758 papers and notes that had been published in the Molecular Ecology journal. From all these 29.8% used mtDNA as a molecular marker and 42.5% the nuclear microsatellites. While mtDNA is most used in animal population evolution and genealogy studies, the microsatellite sequences have proved to be useful for the study of the interaction between the dynamics of population and his genetical structure.

These markers require certain limits, restricting the development of the research. This fact determined the exploration of other types of markers, such as scnp (single copy nuclear polymorphic DNA).

MITOCHONDRIAL DNA – STRUCTURE AND IMPORTANCE

Unlike the nuclear DNA which is situated in the nucleus, the mtDNA is situated in mitochondria. It has been established that the two types of DNA have a different evolution origin. MtDNA derives from the bacterial hereditary material, as a consequence of the incorporation processes of these microorganisms in the precursor eukaryotic cells, without being digested after that (the endosymbiotic theory).

MtDNA has an important advantage in all phylogenetical research, because it is inherited through the maternal line. This fact facilitates the monitoring of its transmission along the evolutive lines starting in the early evolution. In case that one individual is not available for a direct comparison with a biological sample, any sample which comes from the maternal genitor can be a good and usable one (San Mauro et al., 2006).

The reconstruction of a profile based on nuclear DNA is difficult to be elaborated, even for the first degree relatives, because of the meiotic recombination and the diploid inheritance of the nuclear DNA. The pattern of the maternal nuclear DNA inheritance can be also considered to have some deficiencies. Due to the fact that all individuals present the same sequence of mtDNA on maternal line this cannot be considered as an unique marker. Actually some individuals, apparently unrelated, can have a common and unknown maternal origin.

Another advantage is the one that mtDNA has a high level of variability and a high rate of mutation, comparatively with the nuclear DNA, in spite of the fact that it does not encode the information for the synthesis of many proteins. This fact makes it ideal for the phylogenetic and phylogeographical studies (Brown et al., 1979). The great number of polymorphism belonging to the nucleotide sequences from two hypervariable areas located in the region of noncoding mitochondrial control, can allow the discrimination between different individuals or different biological samples.

The probability of recovery of mtDNA from very small or degraded biological samples is higher than the one of nuclear DNA, because the mitochondrial DNA molecules exist in thousands of copies per cell, while nuclear DNA has only two copies per cell.

Another characteristic of mtDNA, generally accepted as a real advantage for the populational genetic studies, is the lack of recombination (Gillham, 1994; Rokas et al., 2003). The majority of the animals have a sexed reproduction. In this type of reproduction the genes are transmitted from both parents after the recombination process through crossing-over and the independent segregation of chromosomes during the first meiosis. But the genes of organelles coming from different lines of filiation cannot ever recombine because the genome of the organelles (such as mtDNA) is uniparental transmitted. Also, if it has been biparental inherited, the organelles from the two genitors wouldn't be able to merge in such a manner that it could be possible to combine their genomes. If two parental cells merge, one cannot find the recombinant genomes at mitochondrial level (Birky, 2001).

The applicability of the studies in mtDNA is reduced only to animals. The animals posses a mtDNA with a set of features that make it an almost perfect molecular marker for the evolution and populational genetics studies (Avise et al., 1987; Moritz et al., 1987; Harrison 1989; Avise 1991; Simon 1991). Its contribution to the understanding of evolution and natural history is crucial (Avise, 1994).

The rate of mtDNA evolution is very slow in plants, excepting a few small groups (Palmer et al., 2000). It results that the plant mtDNA is of a very little help for the studies of populational genetics. In fact, the only big obstacle for the study of plant phylogeny is the difficulty in finding the proper mitochondrial molecular markers (Schaal et al., 1998).

A somatic cell has about 500 to 1000 mitochondria, each of them with a few DNA molecules. These mtDNA molecules are fixed in the interior of the mitochondrion because they are attached to the interior of the mitochondrial membrane, forming together with the proteins the nucleoids (Satoh and Kuroiwa, 1991; Newman et al. 1996).

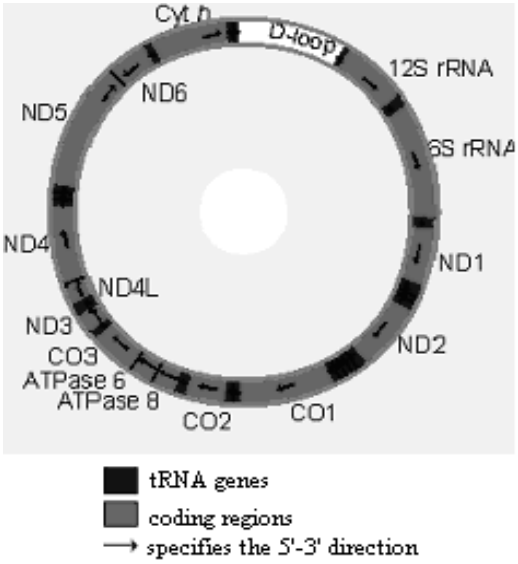

Figure 1 – Animal mitochondrial genome structure (after Gray, 1989)

The animal mtDNA is very compact. Some differences have been found, regarding the detailed organization of the genes for some groups of animals, but the main features are the same. The animal mitochondrial genome is very small, usually circular and it encodes for a limited

amount of functions. There are no introns and some of the genes are even overlapping, so that almost all the base pairs belong to one gene or another (Etienne, 1999). Most of the studied animal mitochondrial genomes have 37 genes (Boore, 1999; Jameson et al., 2003).

There are 13 protein coding regions (figure 1). All these proteins are part of the breathing system (cytochrome b, three subunits of cytochrome oxidase, an ATPase subunit and seven NADH dehydrogenase subunits), (Gray, 1989).

The important differences between the animal mitochondrial genome and the *S. cerevisiae* genome for example, prove a difference in the genetical organization in spite of the common functions.

Figure 2 presents the organization of the mitochondrial genome in *S. cerevisie.* The most important difference is involving the dispersion of the loci on the genetical map.

The most extended regions are the interrupted genes box (which encode for b cytochrome) and oxi3 (which encode for the first subunit of the cytochrome oxidase). Together these two genes, with their large introns, are almost as long as the whole animal mitochondrial genome.

The rest of the genes are uninterrupted and they correspond to another two subunits of the cytochrome oxidase encoded by the mitochondrial genome, to ATPase subunits and to the mitochondrial ribosomal protein. The number of genes that form the yeast mitochondrial genome is 25.

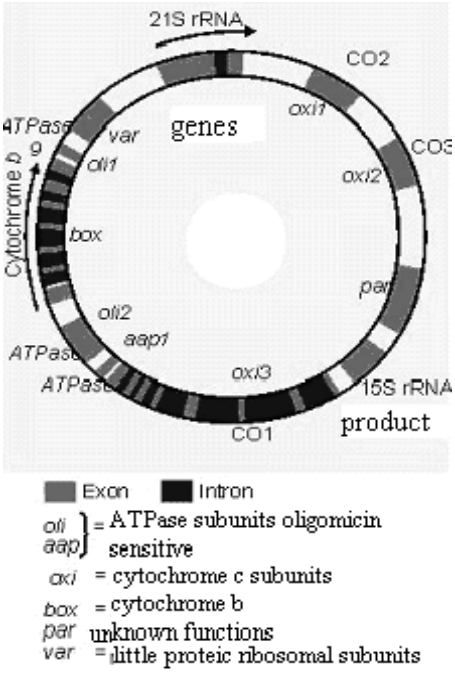

Figure 2 – *S. cerevisiae* mitochondrial genome structure (after Gray, 1989)

There are some disadvantages regarding the use of mtDNA in different studies, such as the recent discovery of mitochondrial pseudogenes inside the genome of a large part of organisms (Zhang and Hewitt, 1996a). Despite all the methods and techniques that inhibit the inclusion of pseudogenes in test samples, this brings a limited help for solving the problems. It has been

proven that these types of sequences similar to those belonging to mtDNA are found in a large amount of copies and with a few differences for a lot of organisms (Zhang and Hewitt, 1996b).

The study of only mtDNA without its combination with information coming from other molecular sources has a limited value. This brings information only about the history of maternal inheritance.

NUCLEAR DNA – MOLECULAR MARKERS AND THEIR IMPORTANCE

The slow rate of nuclear DNA evolution was often considered an element that limited its use in the intraspecific studies. It is known the fact that in many vertebrates the rate of evolution of scnp nuclear sequences is slower than the one of mtDNA (Brown et al., 1979). But this is not a permanent rule. In plants, the nuclear DNA is the one with the fastest evolution from the three genomes that they posses (Wolfe et al. 1987). In animals, except the mammals, the evolutive rate of the nuclear DNA is usually slower than the one of mtDNA. For example, the rate of substitution for mtDNA and scnpDNA in *Drosophila* does not differ highly (Caccone et al., 1988), because of the growth of substitution rate in the nuclear genome, which is eight times higher than in primates (Moriyama and Gojobori, 1992).

Microsatellite DNA markers are the most often used among all nuclear markers. Microsatellite markers have overtaken the mitochondrial the other DNA markers in the race for intraspecific analysis.

The microsatellites are short nucleotide sequences, tandem repeatable (ex. ACn , where n>8). These tandem units can be di-, tri- or tetranucleotids. The mutational processes appears because of the replication, process in which the two three or four sister nucleotide units unite with their complementaries through the excision or addition of other repeatable sequences of the tandem. This type of replication through sliding takes place in microsatellites more often than the point mutations. That is why nuclear microsatellites tend to be hypervariable areas, fact that makes them useful in molecular phyligenetic studies (Strassmann et al., 1996).

These types of simple repetitive sequences are wide spread in the eukaryotic DNA, having a high level of polymorphism due to their high mutational rate. Their inclusion in populational genetics studies, determined the appearance of important progresses in the field of detecting the genetical structure of different populations, of testing the interrelations between individuals and of studying the recent history of different populations.

Although the level of knowledge regarding these sequences is higher than 14 years ago, when the potential of their use in the field of populational genetics has been put into discussion (Bruford and Wayne, 1993), the advantages and disadvantages of their use are not known completely. This happens because the potential these molecular markers have, has not been studied enough. Their models of molecular evolution and their mutational mechanisms could be better known only through systematic studies of the sequences.

The studies of mtDNA and nuclear microsatellites complete only a part of the scientific evaluation of the phylogenetic and phylogeographic relations. There also is a third big category of nuclear molecular markers – scnp markers (single copy nuclear polymorphic sequences).

Analyzing the diversity of genomic projects one can demonstrate the existence of scnp markers in nuclear genomes. The nucleotidic diversity of the human nuclear genome is of 0.1% (International SNP Map Working Group 2001). This means that there are about 3 million nucleotidic differences between any two individuals. Considering the relatively young evolutive history of the human population, this variability rate is encouraging for the confirmation of the availability of scnp loci in the nuclear genome of any eukaryotic organism.

Genome projects and populational genetics studies on other organisms such as rice, *Drosophila* and *Arabidopsis* confirm these discoveries.

After sequencing the rice genome it has been observed that the level pf polymorphism for this plant in 0.67% (Yu et al., 2002).

In the case of fruit fly, Moriyama and Powell (1996) concluded that the nucleotidic diversity is so large that each diploid individual is a heterozygote for each locus.

The distribution of the polymorphic sites is not a random one either in the case of the genome or at a gene level. The existence of the polymorphisms is associated with the different rates of recombination, gene density in different regions of the genome, path of transmission and power of selection.

The genomic regions with a low rate of recombination have mostly a low level of polymorphism (Lercher and Hurst, 2002). Nucleotide diversity and gene density for some regions are correlated in inverse proportion. The regions under a strong selection such as immune system genes or the resistance for diseases loci have the highest diversity (Noël et al., 1999).

The populational genetics studies proved that the rate of substitution at the level of some synonym sequences for different organisms is often higher than the one of introns. The rate of substitution of introns is in the same manner higher than the one of the flanking regions.

CONCLUSIONS

In phylogenetical studies the most used molecular marker is DNA, the nuclear DNA for plants and the mtDNA for animals.

The advantages mtDNA has in the phylogenetical research are the fact that it is inherited through maternal line, has a high level of variability and a high rate of mutation, the lack of recombination and a high probability of recovery.

Recent discovery of mitochondrial pseudogenes inside the genome is a disadvantage for the use of mtDNA in different studies.

Microsatellite DNA markers are the most often used among all markers.

There also is a third big category of nuclear molecular markers – scnp markers (single copy nuclear polymorphic sequences) whose existence depend on the different rates of recombination, gene density in different regions of the genome, path of transmission and power of selection.

REFERENCES

Avise, J. C., 1991. *Ten unorthodox perspectives on evolution prompted by comparative population-genetic findings on mitochondrial DNA*, *Annual Review of Genetics*, 25, 45-69.

Avise, J. C., 1994. *Molecular Markers*, *Natural History and Evolution*, Chapman & Hall, New York.

Avise, J. C., 1998. *The history and purview of phylogeography: personal reflection*, *Molecular Ecology*, 7, 371-379.

Avise, J. C., Arnold J., Ball R.M., Bermingham E., Lamb T., Neigel J.E., Reeb C.A., Saunders N.C., 1987. *Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics*, *Annual Review of Ecology and Systematics*, 18, 489-522.

Birky W. C., 2001. *The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms and models*, *Annu. Rev. Genet.*, 35, 125 – 148.

Boore, J. L., 1999. *Animal mitochondrial genomes*. *Nucleic Acids Res.*, 27, 1767–1780.

Brown, W. M., George M., Wilson A.C., 1979 – *Rapid evolution of animal mitochondrial DNA*, *Proc. Natl. Acad. Sci*., 76/4, 1967-1971.

Bruford, M.W., Wayne, R.K., 1993. *Microsatellites and their application to population-genetic studies*, *Current Opinion in Genetics and Development,* 3, 939-943.

Caccone, A., Amato G.D., Powell J.R., 1988. *Rates and patterns of scnDNA and mtDNA divergence within the Drosophila melanogaster subgroup*, *Genetics*, 118, 671-683.

Etienne, Jacqueline, 1999. *Biochimie genetique. Biologie moleculaire*., Masson, Paris, 22.

Gillham, N.W., 1994. *Organelle Genes and Genomes*, Oxford University Press.
Gray, M. W., 1989. *Origin and evolution of mitochondrial DNA*, *Ann. Rev. Cell Biol.*, 5, 25-50.
Harrison, R.G, 1989. *Animal mitochondrial DNA as a genetic marker in population and evolutionary biology*, *Trends in Ecology and Evolution*, 4, 6-11.
International SNP Map Working Group, 2001. *A map of humangenome sequence variation containing 1.42 million single nucleotide polymorphisms*, *Nature*, 409, 928-933.
Jameson, D., Gibson A.P., Hudelot C., Higgs P.G., 2003. *OGRe: a relational database for comparative analyses of mitochondrial genomes*, *Nucleic Acids Res*., 31, 202–206.
Lercher, M.J., Hurst, L.D., 2002. *Human SNP variability and mutation rate are higher in regions of high recombination*, *Trends in Genetics*, 18, 337-340.
Moritz, C., Dowling T.E., Brown W.M., 1987. *Evolution of animal mitochondrial DNA: relevance for population biology and systematics*, *Annual Review of Ecology and Systematics*, 18, 269-292.
Moriyama, E.N., Gojobori, T., 1992. *Rates of synonymous substitution and base composition of nuclear genes in Drosophila*, *Genetics*, 130, 855-864.
Moriyama, E.N., Powell, J.R., 1996. *Intraspecific nuclear DNA variation in Drosophila*, *Molecular Biology and Evolution*, 13, 261-277.
Newman, S.M., Zelenaya-Troitskaya O., Perlman P.S., Butow R.A., 1996. *Analysis of mitochondrial DNA nucleoids in wild-type and a mutant strain of Saccharomyces cerevisiae that lacks the mitochondrial HMG box protein Abf2p*, *Nucleic Acids Res*., 24, 386–393.
Noël L., Moores L., van Der Biezen A., Parniske M., Daniels M.J., Parker J.E., Jones J., 1999. *Pronounced intraspecific haplotype divergence at the RPP5 complex disease resistance locus of Arabidopsis*, *Plant Cell*, 11, 2099-2111.
Palmer, J. D., Adams K.l., Cho Y., Qiu Y.L., Song K., 2000. *Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates*, *Proceedings of the National Academy of Sciences USA*, 97, 6960-6966.
Rokas, A., Ladoukakis E., Zouros E., 2003. *Animal mitochondrial DNA recombination revisited*, *Trends in Ecology and Evolution*, 18/8, 411-417.
San Mauro, D., Gower D.G., Zardoya R., Wilkinson M., 2006. *A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome*, *Molecular Biology and Evolution*, 23, 227–234.
Satoh, M., Kuroiwa, T., 1991,*Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell*, *Exp. Cell Res*., 196, 137–140.
Schaal, B. A., Hayworth D.A., Olsen K.M., Rauscher J.T., Smith W.A., 1998. *Phylogeographic studies in plants: problems and prospects*, *Molecular Ecology*, 7, 465-474.
Simon, C., 1991. *Molecular systematics at species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA*, *Molecular Techniques in Taxonomy*, Springer-Verlag, Berlin, 33-71.
Strassmann, J.E., 1996. *Strategies for finding and using highly polymorphic DNA microsatellite loci for studies of genetic relatedness and pedigrees*, *Molecular Zoology: Advances, Strategies and Protocols*, John Wiley and Sons, New York, 163-180.
Yu et al., 2002. *A draft sequence of the rice genome (Oryza sativa L. ssp. indica*), *Science*, 296, 79-92.
Wolfe K. H., Li W.H., Sharp P.M., 1987. *Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs*, *Proceedings of the National Academy of Sciences USA*, 84, 9054-9058.
Zhang, D.-X., Hewitt, M., 1996a. *Nuclear integrations: Challenges for mitochondrial DNA markers, Trends in Ecology and Evolution,* 11, 247-251.
Zhang, D.-X., Hewitt, M., 1996b. *Highly conserved nuclear copies of the mitochondrial control region in the desert locust Schistocerca gregaria: some implications for population studies, Molecular Ecology*, 5, 295-300.

1 - "Alexandru Ioan Cuza" University of Iaşi, Faculty of Biology, Department of Genetics
* - laura_sv2002@yahoo.com