

Strength and Limitations of DNA Barcode under the Multidimensional Species Perspective

Valerio Sbordoni

Abstract — DNA barcoding aims at providing an efficient method for species-level separation using a partial sequence of the mitochondrial COI gene. The efficiency of the barcode in separating species is based on the amount of genetic distance among samples. While in many taxa the species can be efficiently identified through the barcode, other situations cannot be treated by this approach. The causes for such discrepancy appear to be mostly related to the nature of speciation events and to the different roles of the genetic system, natural selection and evolutionary time. Thus, DNA barcode represent just one important descriptor in the framework of the multidimensional species approach.

Index Terms — DNA barcode, molecular systematics, species concepts, taxonomic procedures.



1 INTRODUCTION

Since 2003, many research groups started to accumulate molecular data with the aim of setting up a sort of inventory of life that might itemize biodiversity as a sequence of species-specific DNA. In particular, Paul Hebert from the Canadian University of Guelph [1], [2], proposed to use a sequence of the COI mitochondrial gene, coding for cytochrome oxidase 1, as a “molecular signature” to identify a species. The selection of this gene for exploring limits between species allows for the practical advantages of using mitochondrial DNA together with the previous wide use of this gene in a large variety of organisms. COI sequences are also currently used at different taxonomic levels, in phylogeny, phylogeography and population genetics studies, due to the great advantage offered by the availability of amplification protocols, as well as a large number of sequences ready for barcoding.

At present there are many ongoing DNA barcoding projects reported on the

website of the Barcode of Life Data Systems (www.boldsystems.org), an online workbench that supports collection, management, analysis, and use of DNA barcodes. An enormous bulk of barcode data for a wide array of organisms has already been made available to the scientific community. Thus, a sequence of the mitochondrial COI gene has become the most used mitochondrial marker, especially for animals. The same marker, preferably associated with nuclear DNA sequences, is commonly used also in more wide phylogenetic studies. The choice of a marker specific for plants and fungi is more problematic, but it currently seems oriented, at least for angiosperms, to the *trnH-psbA* sequence, an intergenic spacer of plastidic DNA [3], [4].

2 MOLECULAR SYSTEMATICS AND THE DNA BARCODE

By identifying genetic differentiation threshold values that would include individuals of the same species, the barcode approach allows investigating and analysing in some detail inter-species delimitation and its many related problems. A great advantage offered by DNA barcoding is the possibility of identifying cryptic species, that is to distinguish as belonging to different species individuals that, due to their similar morphology, were considered as belonging to a unique species. This would be feasible by identifying genetic distance limit values within which two individuals can be considered as belonging to the same species, while outside these limits they should be considered as belonging to different species. Although the limit seems to be somehow taxon dependent, it has been observed that a value of genetic distance between two DNA-barcode sequences equal to or higher than three per cent ($D \geq 0.03$) identifies distinct species.

However, it has become clear that the currently used genetic distance approaches by means of DNA barcodes have strong limitations, particularly when it comes to defining species boundaries [5]. One reason is that mtDNA rates of evolution vary substantially between and within species and between different groups of species, thus resulting in broad overlaps of intra- and interspecific distances [6], [7], [8]. But there are other reasons worth to be briefly discussed.

To what extent DNA barcode may solve the problem of identification of interspecific limits? And, will this approach be applicable to the different typologies of species? Based on the data produced up to day, it seems that DNA barcoding may solve problems of specific classification in a wide range of organisms and situations, but many taxa cannot be treated by this approach [9].

Let's try to understand why. Biologists and evolutionists know that species are the result of historical processes, mainly speciation and range dynamics. The knowledge of these processes is of help for inferring the essence of species whose properties reflect the "signature" of speciation where the following factors have more or less predominant roles:

1. The genetic system.
2. Natural selection.
3. Evolutionary time.

The genetic system, *i.e.* the organization of genes and of other chromosomal

genetic material may favour, by acceleration, the speciation process especially when it is based on Robertsonian translocations, or inversion polymorphisms or, as with the extreme case of instantaneous speciation, on allopolyploid mechanisms, so frequent in plants.

Natural selection may have a determinant role in those sympatric speciation phenomena connected to the shift of the trophic niche, that are frequent especially in phytophagous insects. In these cases selective pressure may significantly accelerate the adaptive diversification process of genotypes in their new evolutionary path.

Finally, in the absence of diverging selective pressure, even the simple allopatric condition of two originally conspecific populations may lead to speciation in the long run if the genic flow barrier persists. This is where the evolutionary time finds its role. The longer the isolation time, the higher the number of different nucleotidic substitutions in the DNA sequences of the two genic pool involved. Accumulation of diverging mutations in the whole genome will inevitably lead to speciation by genetic drift, and the speed of this process will be inversely proportional to the population effective size.

This brief analysis of times and modes of speciation is of help for interpreting the biological meaning of genetic distance estimates based on the DNA barcode concept. A limit of 3% genetic divergence is relatively well working to separate species that are the result of geographic speciation events driven by gradual accumulation of diverging mutations. Emblematic examples are diverging populations and species of animals adapted to cave life [10], with special regard to the thoroughly studied *Dolichopodacave* crickets [11], [12], [13].

On the other hand, much smaller values, up to a ten factor, do not allow the discrimination by barcoding of species differing by one or more chromosomal inversions, as it is the case with some *Anopheles* [14], or of recent species originated sympatrically by adaptive shift, as it has been reported in fruit flies of the genus *Rhagoletis* [15]. Most probably, in the case of *Dolichopoda*, the divergence process may have involved most of the genome and, consequently, mitochondrial genes as well. Conversely, in the other two examples, the divergence should have involved, in relatively short times, only a few genes target of selection, or particular combinations of these genes, showing no effects in mitochondrial DNA sequences. Indeed, mitochondrial DNA is often used as a molecular clock [16], [17].

3 THE MULTIDIMENSIONAL APPROACH

The debate on species' properties shows many different aspects: a more classical one opposing the "biological concept of species, BSC" to the "typological, TSC" one [18], and a more modern one that seeks a conflict between BSC and the "phylogenetic concept" [19].

But when analysing the nature of these contrasts it must be agreed that the three concepts focus on properties of the species that do not contrast and which are coherent with a vision of the species as the *in itinere* product of the evolutionary process. Therefore it should be agreed that a species, beside representing "*the smallest monophyletic group of common ancestry*"

[20], expresses its distinct gene pool (BSC) by having eventually acquired the distinctive characters emphasized in the typological species concept.

Actually, in routine work, many taxonomists tend to use operational approaches toward species, based on morphological characters that are unique and shared or their hierarchies. Although this is not explicitly declared, these species are based on philosophies ranging from the classic typological concept up to the phylogenetic and the phenetic ones. Traditional phenetic definitions of species are based essentially on the numeric recognition of intervals separating clusters of phenetically similar individuals [21], [22]. For this evaluation many different, "not weighed" types of taxonomic characters are examined, but their biological meaning is not evaluated.

The phenetic "concept" has been strongly criticized because it was not considered as sufficient to describe the complex interrelations existing among clusters of similar populations. Nonetheless, this kind of approach has operational advantages of some practical value and consequently it is widely adopted and used by systematists.

Being an intrinsically complex entity, the species requires a multidimensional approach taking into account the whole set of taxonomic characters. Nowadays, the huge progress in multivariate analysis, together with the wide choice of technologies to measure parameters related to the ecology of niche, or to sexual behaviour, and/or to many other crucial features of species-specificity, have richly endowed the kit of characters available to the modern taxonomist. The technical and conceptual progress in the field of molecular biology has made relatively easy and rapid both the acquisition and the phylogenetic interpretation of sequence data containing an enormous quantity of information. The routine use of these characters has substantially empowered the understanding of the species' genetic structure and has brought to the discovery of the existence of cryptic species.

However, as previously discussed, DNA sequences are not the only depositories of evolutionary history. Any other kind of character is potentially suitable to give its own contribution to the species definition. For instance, much of the ecological role (niche) of an organism, is written in its morphology, although there is no assurance that the ecological divergence between two similar species corresponds to a genetic gap, even though this coincidence shows up in the majority of cases. The diagnostic value of each character varies by taxon or specific evolutionary, geographical and ecological situations. For most taxa specialists know very well which characters are more kin to represent the species' biological properties.

This discussion makes us re-consider, although with due caution and adjustments, the usefulness of the phenetic approach. Since different descriptors have different values for a given organism, one cannot rely on automatic discriminating procedures, while one can rely on the availability of the whole set of algorithms and multivariate procedures set up in the systematic and ecological fields. Yet, the responsibility of the final decision will inevitably have to rely on the competence and experience of the specialist. It must be stressed that different taxonomic characters do not necessarily vary in a coordinated way, yet they are often conflicting. Both evaluation and weighing of characters

have always had and continue to have a fundamental role in the process of species' delimitation [11].

Based on this logic and premises, species are considered as "clusters of individuals that are effectively separated from other clusters in the space defined by their descriptors" [23]. Alike the preceding phenetic definitions, species are seen as clouds of probability in an hyperspace. Here though characters are weighed and a value is assigned to genetic, and inter-reproductive descriptors, *i.e.* exactly those characterising the species as a monophyletic cluster, as a cluster of genotypes and as a cluster of individuals sharing a special relation with their environment. An "ad hoc" reduction of this hyperspace makes this multidimensional concept operative. For instance, the typical biological species becomes a particular case where intra-population genetic and reproductive relationships are quantified and analysed as a sub-set of a wider set of descriptors. Yet, the use of the multidimensional approach should be particularly useful for organisms with asexual or uniparental reproduction, including bacteria, protists, fungi, rotifera and many parthenogenetic taxa, to which it is traditionally difficult or impossible to apply the biological concept of species, thus overcoming the tie of amphigonic reproduction and allowing, not only in theory, the evaluation of clusters defined by appropriate descriptors. The literature on taxonomy, and not only the recent one, offers many examples of this approach, adopted with success in cave crickets [11], butterflies [24], fishes [25], fossil Ostracoda [26], Rotifera[27], etc.

Many species' definitions, privileging either properties, can be accommodated within this approach, but I want here to recall in particular a somewhat unknown definition by Alfred Russell Wallace, incidentally quoted in one of his writings where he disputes with Galton: "*A species ... is a group of living organisms, separated from all other such groups by a set of distinctive characteristics, having relations to the environment not identical with those of any other group of organisms, and having the power of continuously reproducing its like*" [28].

This definition, dated many years before the Synthetic Theory, refers to all the emerging properties of species: a set of distinctive characters (highlighted by the typological concept), the relationship with the environment (ecological concept), and finally the power of reproducing its own characteristics, which implies the properties of the hereditary material. Compared with many others, Wallace's definition has certainly the merit of stressing multidimensionality, a concept that expresses the best operational solution to the problem of the delimitation of species.

REFERENCES

- [1] P. D. N. Hebert, A. Cywinska, S. L. Ball and J. R. DeWaard, "Biological Identifications Through DNA Barcodes", *Proc. R. Soc. Lond. B.*, vol. 270, pp. 313-321, 2003.
- [2] S. Ratnasingham and P. D. N. Hebert, "Barcoding Bold: The Barcode of Life Data System", *Molecular Ecology Notes*, doi: 10.1111/j.1471-8286.2006.01678.x, 2007.
- [3] W. J. Kress, K. J. Wurdack, E. A. Zimmer, L. A. Weigt and D. H. Janzen, "Use of DNA Barcodes to Identify Flowering Plants", *PNAS*, vol. 102, pp. 8369-8374, 2005.
- [4] M. W. Chase and M. F. Fay, "Barcoding of Plants and Fungi", *Science*, vol. 325, pp. 682-683, 2009.

- [5] J. D. Witt, D. L. Threlkoff and P. D. Hebert, "DNA Barcoding Reveals Extraordinary Cryptic Diversity in an Amphipod Genus: Implications for Desert Spring Conservation", *Mol. Ecol.*, vol. 15, pp. 3073–3082, 2006.
- [6] W. W. Kipling and D. Rubinoff, "Myth of the Molecule: DNA Barcodes for Species Cannot Replace Morphology for Identification and Classification", *Cladistics*, vol. 20, pp. 47–55, 2004.
- [7] D. Rubinoff, "Utility of Mitochondrial DNA Barcodes in Species Conservation", *Conserv. Biol.*, vol. 20, pp. 1026–1033, 2006.
- [8] D. Rubinoff, S. Cameron and K. Will, "A Genomic Perspective on the Shortcomings of Mitochondrial DNA for "Barcoding" Identification", *J. Hered.*, vol. 97, pp. 581–594, 2006.
- [9] J. Waugh, "DNA Barcoding in Animal Species: Progress, Potential and Pitfalls", *BioEssays*, vol. 29, pp. 188–197, 2007.
- [10] V. Sbordoni, "Advances in Speciation of Cave Animals". In: C. Barigozzi (ed.), *Mechanisms of Speciation*, New York, A.R. Liss Inc., pp. 219–240, 1982.
- [11] V. Sbordoni, G. Allegrucci and D. Cesaroni, "A Multidimensional Approach to the Evolution and Systematics of *Dolichopoda* Cave Crickets". In: G. M. Hewitt et al. (eds.), *Molecular Techniques in Taxonomy*, NATO ASI Series, vol. H57, Berlin, Springer, pp. 171–199, 1991.
- [12] G. Allegrucci, M. Rampini, P. Gratton, V. Todisco and V. Sbordoni, "Testing Phylogenetic Hypotheses for Reconstructing the Evolutionary History of *Dolichopoda* Cave Crickets in the Eastern Mediterranean", *J. Biogeography*, vol. 36, pp. 1785–1797, 2009.
- [13] L. Martinsen, F. Venanzetti and A. Johnsen, V. Sbordoni and L. Bachmann "Molecular Evolution of the pDo500 Satellite DNA Family in Dolichopoda Cave Crickets (Raphidophoridae)", *BMC Evolutionary Biology*, vol. 9, 301 (14 pp.), 2009.
- [14] M. Coluzzi, A. Sabatini, A. della Torre, M.A. Di Deco and V. Petrarca, "A Polytene Chromosome Analysis of the *Anopheles gambiae* Species Complex", *Science*, vol. 298, pp. 1415–1418, 2002.
- [15] J. J. Smith and G.L. Bush, "Phylogeny of the Genus *Rhagoletis* (Diptera: Tephritidae) from DNA Sequences of Mitochondrial Cytochrome Oxidase II", *Mol. Phylog. Evol.*, pp. 33–43, 1997.
- [16] L. Bromha and D. Penny, "The Modern Molecular Clock", *Nature Rev. Genet.*, vol. 4, pp. 216–224, 2003.
- [17] A. Caccon and V. Sbordoni, "Molecular Biogeography of Cave Life: a Study Using Mitochondrial DNA from Bathysciine Beetles", *Evolution*, vol. 55, pp. 122–130, 2001.
- [18] E. Mayr, *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*, New York, Columbia University Press, 1942.
- [19] M. J. Donoghue, "A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative", *Bryologist*, vol. 88, pp. 172–181, 1985.
- [20] K. de Queiro and M. J. Donoghue, "Phylogenetic Systematics and Species Revisited", *Cladistic*, vol. 6, pp. 83–90, 1990.
- [21] C. D. Michener, "Diverse Approaches to Systematics", *Evol. Biol.*, vol. 4, pp. 1–38, 1970.
- [22] E. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy*, W. H. Freeman, San Francisco, 1973.
- [23] V. Sbordoni, "Molecular Systematics and the Multidimensional Concept of Species", *Biochemical Systematics and Ecology*, vol. 21, pp. 39–42, 1993.
- [24] D. Cesaroni, M. Lucarelli, P. Allori, F. Russ and V. Sbordoni, "Patterns of Evolution and Multidimensional Systematics in Graylings (Lepidoptera: *Hipparchia*)", *Biol.J.Linn.Soc.*, vol. 52, pp. 101–119, 1994.
- [25] M. Barluenga, K. N. Stolting, W. Salzburger, M. Muschic and A. Meyer, "Sympatric Speciation in Nicaraguan Crater Lake Cichlid Fish", *Nature*, vol. 439, pp. 719–723, 2006.
- [26] M. Gross, K. Minati, D.L. Danielopo and W. E. Piller, "Environmental Changes and Diversification of *Cyprideis* in the Late Miocene of the Styrian Basin (Lake Pannon, Austria)", *Palaeobiodiversity and Palaeoenvironments*, vol. 88, pp. 161–181, 2008.
- [27] D. Fontaneto, E. A. Herniou, C. Boschetti, M. Caprioli, G. Melone, C. Ricc and T. G. Barraclough, "Independently Evolving Species in Asexual Bdelloid Rotifers", *Plos Biology*, doi: 10.1371/journal.pbio.0050087, 2007.
- [28] A. R. Wallace, "The Method for Organic Evolution", *Fortnightly Review (N.S.)*, vol. 57, pp. 435–445, London, 1895.