A Mathematical Model for the Entities Species and Genus
Author(s): Pierre Legendre and Pierre Vaillancourt
Source: *Taxon*, Vol. 18, No. 3 (Jun., 1969), pp. 245-252
Published by: International Association for Plant Taxonomy (IAPT)
Stable URL: http://www.jstor.org/stable/1218824
Accessed: 18/09/2013 15:57

TAXON VOLUME 18 . JUNE 1969 pp. 245–356

A MATHEMATICAL MODEL FOR THE ENTITIES SPECIES AND GENUS

Pierre Legendre * *and Pierre Vaillancourt* ** (*Montréal*)

Summary

The taxonomic entities called genus and species are, though very much in use, still vaguely defined. Through the study of different modern authors one can approximate their meaning to a sufficient degree of precision to allow the formulation of a mathematical model. The species will be defined as the set of all genetically similar individuals, barring differences that reside at the allele level. We will then consider the "species such as in nature" to be the union of certain vital neighborhoods in a multi-dimensional space, the said intervals obviously being those that correspond to the given species. This concept will be used to define, with an evolutionary tree pattern, the genus. The genus will then be the set of species, linked by monophyletism, that form an evolutionary peak in the space of "conditions of life".

Résumé

Les entités taxonomiques que l'on appelle genre et espèce sont encore sujettes à l'arbitraire, quoique ces mots soient dans la bouche de tous les biologistes. En se basant sur les réflexions d'auteurs modernes, il est cependant possible d'en venir à des définitions de genre et d'espèce qui cernent d'assez près le problème, dans l'état actuel des connaissances, et de bâtir sur elles un modèle mathématique. L'espèce sera d'abord définie comme l'ensemble des individus génétiquement similaires, à des différences d'allèles près. Puis l'on considérera qu'écologiquement, l'espèce telle qu'elle existe dans la nature sera l'union des "voisinages vitaux", dans un espace multi-dimensionnel, de ces individus génétiquement similaires. Cette notion sera utile dans la définition du genre, qui sera basée sur le modèle de l'arbre évolutif. Le genre est défini comme l'ensemble des espèces, reliées par monophylétisme, qui, dans l'espace multi-dimensionnel des conditions de vie, forment un agglomérat appelé un sommet évolutif.

Introduction

When discussing the limitation of species and genera, which are the most important and distinct units in the biological hierarchy, scientists often disagree as to how these units should be defined. For those who think such taxa exist only in the mind of man who has created them for his convenience, definitions are of no significance, whereas evolutionists see in these units distinct evolutionary steps which have an existence in nature quite independently of man. The latter opinion, which seems to be predominant among modern zoologists but rarer among botanists, has demonstrated that the only way of defining what they call the biological species is by aid of its reproductive isolation; that this unit is real is not only supported by numerous biological observations but also by common logic (Lehmann, 1967). We cannot express modern evolutionary opinions on the natural biological hierarchy better than did Mayr (1965) when he said: "The grouping of organisms found in nature is the result of two evolutionary processes, the splitting of phyletic lines owing to speciation, and the unequal subse-

---

* Department of Zoology, McGill University, Montréal, Québec.
** Département de Mathématiques, Université de Montréal, Montréal, Québec.
Present address of both authors: Department of Biology, University of Colorado, Boulder, Colorado 80302, U.S.A.

quent divergence of such lines. Information on such past events is contained in the genetic program of all organisms. The classification of historically programmed organisms differs in principle from the classification of inanimate objects. The taxonomist, when grouping species into lower taxa and such lower taxa into higher taxa, must base his decision on the evaluation of the historically given information content of the organism to be classified".

In the age of computers, time has come to attempt to form models of the evolutionary units so that their mathematical background can be better clarified. We have tried to formalize the presently existing propositions of the structure of biologically definable taxa of biota whose reproduction is based on allogamy. Though some of our results may perhaps be generalized to be used for all organisms, we do not discuss such possibilities at present.

The generic concept

Generic concepts have recently been discussed from various points of view by several biologists, notably Mayr (1942), Simpson (1961), Tutin (1956), Davis & Heywood (1963), Anderson (1957), Löve (1963) and Kirpicznikov (1968). As to the difficulties of defining natural genera, Löve (1963) pointed out: "Morphological characters have always been and will always be the main criteria for the identification of a genus, but other methods may be found to be more reliable for definitions of its boundary. Thus, embryological distinctions and also anatomical characteristics have been used with considerable success in distinguishing critical genera, and so have some biochemical and serological methods. Lately, certain palynological peculiarities have been shown to be of great importance, and so have also some cytological features".

This explains why one can easily question the definition of a particular genus: the norms being so vaguely defined, they can only be considered as one of the steps in the approximation of a "true" definition of this genus. In this way, Löve (1963) said that "there has been a tendency to regard the size of genera as some kind of an indication of their degree of artificiality, ... it certainly was true for the much-too-inclusive fern genus *Dryopteris* until its recent revision". The old *Dryopteris* genus has recently been split into several genera: *Dryopteris* ($n = 41$), *Phegopteris* ($n = 30$), *Thelypteris* ($n = 34$), *Lastrea* ($n = 35$), and *Carpogymnia* ($n = 40$), which are morphologically and cytologically distinct and well-defined, contrary to the older aggregate.

But what can be the criteria for the definition of a biologically strictly limited genus? In choosing modern taxonomical criteria, Cain (1956) pointed out: "The genus is the lowest obligate category for which invariably only comparative data, corrected as far as possible from all data bearing on actual phylogeny, are available. It is a *natural group of species, monophyletic* ... and *arbitrarily delimited* at any one moment in time when either it merges backward and forward into phyletically continuous forms, or phyletic lines exist half-way in their affinities between it and a contemporary related genus and could be equally well included in either. Since the entities it contains are so far being monotypic units, it is not surprising that some genera are known which are clearly natural groups yet cannot be diagnosed at all, since every character confined to them is lost or modified in one or more forms, the remainder of whose characters suffice to establish their membership in the group. ... The aims of (experimental taxonomy) are to identify evolutionary units, and by experiment to determine their genetical inter-relationship and the role of the environment in their formation".

In other words, the biologically distinct genus may be defined as a category including only species that are naturally related by monophyletism (genetic concept); this category may also be arbitrarily delimited so as to correspond to the intuitive idea that one has of an evolutionary peak that is distinct from any other closely related peak (evolutionary concept).

The species concept

For the understanding of the following formal definition of a genus, it will be useful to clearly understand what is meant by the species category. This can best be done if certain mechanisms of speciation are understood.

According to Löve (1964a), we may say that "each species is a reproductive community. It consists of populations, each of which is an expression of an integrated gene pool. ... Evolution below the species level is characterized by a continuum of variations and not by a succession of distinct types" as on the specific level. Each species "constitutes the stage of evolution when a major genetical system becomes closed and loses its ability for interbreeding and fusing with other such systems".

"The processes leading to reproductive isolation ... include all those differences which prevent two populations from exchanging genes through the formation of fertile hybrids, *actually and potentially*". And he continues (Löve, 1964b): "The processes leading to reproductive isolation are the processes of speciation in the strict sense. ... Reproductive isolation is brought into being either by changes on the genetical arrangement within the chromosomes or simply by ... the abrupt and instantaneous creation of a very effective barrier to reproductive miscibility by means of changes in the number of chromosomes".

We may conclude that a species is the smallest group of populations participating in a closed genetical system, reproductively isolated and morphologically somewhat different from any other species. Ecologically, the species consists of a group of individuals representing an adaptive peak, because of the success of their genetic composition in a certain ecological range.

Model of species

First we draft the model of the species as a genetic unit. We consider, in this part, the species as a relatively stable object. Since if we take into account all the possible variability of species, and the cases of hybridization, mutation, and sex differentiation, the model will be too elaborate for the purpose of the present work.

In this connection we consider the genetic unit that we call a species as an evolutionary peak, whereas the genus is a group of species related by monophyletism and forming together an evolutive peak.

Genetic definition of species. Before giving the mathematical model of the biological species and genus the following explanations must be given. The numbers in parentheses correspond to those in the following section on the mathematical form of the model.

(1) We define a Σ-gene as a gene considered as being the set of all its possible alleles, to which we add a naught allele, corresponding to the lack of this gene. Each allele of the i-th Σ-gene will be denoted $\Sigma_{i,}$ α_i etc. α_i varies from 0 to v_i where v_i is the maximum number of alleles possible for this gene. $\Sigma_{i,0}$ is the lack of the gene. The total number of existing genes is finite and is equal to K.

(2) In the chromosome complement of an individual, there are two corresponding chromosomes, including two corresponding alleles for each gene. The Σ_i are the different genes of one of the series of chromosomes, and Σ^*_i the different genes of the other. Different alleles of Σ^*_i are Σ^*_i, β_i. The two corresponding alleles of an individual will form a pair (Σ_i, α_i, Σ^*_i, β_i) or an element of all the possible pairs of alleles. The set of all the different pairs of alleles of a single gene will be regarded as the cartesian product of the set Σ_i by the set Σ^*_i.

(3) The set of all possible combinations of genes that exist, each gene represented by a pair of two alleles, is the set of all the possible K-tuples, where each element of a K-tuple is a pair (Σ_i, a_i, Σ^*_i, β_i). This set will be called P.

(4) In this set, we take a subset $\overline{P}$ such that, in this subset, we eliminate all the combinations of alleles in which there is one naught allele ($\alpha_i = 0$) and one real allele ($\alpha_i \neq 0$).

(5) We define a genetic individual as an element of the set $\overline{P}$. The genetic individual represents the class of all the individuals that are genetically similar in all of their alleles. It is possible to find two individuals that are the same genetic individual, in the case of identical twins, or after vegetative or apomictic reproduction.

(6) Any individual will receive its genes from the major set of all the possible combinations of genes, taking two naught alleles for a gene that is absent. This means that for every individual there exists a corresponding and unique set of genes. We define G(x) as the set of genes that appear in the complement of the genetic individual x.

(7) Consequently, a species will be the set of all the individuals that have the same genetic load, including the possible variation of alleles in each gene. The species of the genetic individual x will be noted $\overline{G(x)}$. As mentioned in a preceding paragraph, this definition does not include chromosomal differences between males and females, etc.

Ecological notion: the species such as in nature. (8) A geometrical representation of the species, which corresponds to the genetic definition, might also serve as a definition; but we will regard it only as a representation. Hence the species is characterized as something that lives in a certain well delimited environment, or in other words as a little evolutionary peak as defined by Dobzhansky (1951, p. 9). First, the set of all the external conditions A_i that may influence the life of an individual needs to be defined. As examples, we may talk about temperature, time, pressure, concentration of water, etc. The units of the different A_i are also different units. Then, the cartesian product of all the A_i, elements of the set of conditions, is defined as the multi-dimensional space CD of the conditions of life of the individuals. Time is one of our coordinates, and the life of an individual corresponds to a continuous line in the multi-dimensional space CD.

(9) Here we define the vital neighborhood of the individual x along the i-th condition, as an interval in the line that may represent, in a graph, the condition A_i. As an example, the vital neighborhood of a Beetle along the Temperature condition, will be (0° C., 35° C.).

(10) Then we want to consider the range of all possible conditions under which a given genetic individual may exist. This will obviously be the cartesian product of all the vital neighborhoods corresponding to this individual (product taken over i). Geometrically this forms a sort of hypercube embedded in a multi-dimensional space.

Hence it is the maximal such hypercube in which the given genetic individual may exist. It is called the total vital neighborhood of x, and will be given the sign W_x.

(11) If we were to consider the union of all total vital neighborhoods for individuals that utilise the same genes as x, i.e., for all individuals element of $\overline{G(x)}$, we would have a sort of physical (though multi-dimensional) representation of the domain in which these individuals can exist. We may say that this new neighborhood represents the "species such as in nature" understanding that this is the species of x. We can say that the species of x is represented by $[x]$. We can differentiate the representations of the species by indexing them properly: $[x]_1, [x]_2, \ldots, [x]_n$. We will call Q the set of all the representations $[x]_i$.

Model of genera

(12) All the species that exist in nature are related by the evolutionary tree, which is a graphical representation of the pattern of evolution, joining all the species by phyletic lines. The last common branching point of two species is their last ancestor. Let us introduce of few notations. Given n species $\overline{\{G(x)_1}, \ldots, \overline{G(x)_n}\}$ we shall designate their last common ancestor by $\overline{(G(x)_1}, \ldots, \overline{G(x)_n)}$. Given a species $\overline{G(x)_i}$, we designate the set of all its descendants (species above it) by $\overline{[G(x)_i)}$. With this we can introduce the important concept of phyletic closure of a set of species $\overline{\{G(x)_1}, \ldots, \overline{G(x)_n\}}$. It will be $\overline{[(G(x)_1}, \ldots, \overline{G(x)_n)})$. This corresponds to the branch of the phyletic tree above the last common ancestor of the set of species under consideration.

(13) Theoretically the space of conditions CD is a metric space, i.e., a space in which we can define a distance between points or clusters of points. In the case which concerns us a metric space is a finite dimensional vector space along with a function d which assigns a real number to each pair of points of the space, and is such that

1) $d(x, y) \geqslant 0 \; (= 0 \longleftrightarrow x = y)$

2) $d(x, y) = d(y, x)$

3) $d(x, y) \leqslant d(x, z) + d(z, y)$

where x, y and z denote points of the space. For a given space, different functions d may serve as metrics, even if they do not assign the same number to a given pair of points. One can for instance, through the use of a proper metric, attach a weight to certain coordinates. If two species differ only by their time of occurrence, their distance can be very small. A given species might also occupy a very large territory, though it's diameter ($\delta([x]_i) = \max\limits_{a, b \,\varepsilon\, [x]_i} d(a, b)$; $d(a, b)$ = diameter of $[x]_i$) is still quite small, again by use of the proper metric. Hence if we consider the clusters formed by the species as in nature, we can say that they are more or less distant from each other. A group of species R will be said to form an evolutionary peak when for every object $[x]_i$ corresponding to the species $\overline{G(x)_i}$, given a representation $[x]_j$ that corresponds to a species that is not in R, we can find another representation $[x]_k$, corresponding this time to a species in R, such that the distance in CD between $[x]_i$ and $[x]_j$ is less than the distance between $[x]_i$ and $[x]_k$, $[x]_k$ being different from $[x]_i$.

(14) We want a genus to be formed of species phyletically related. In other words, we want them to have a good deal of their genetic load in common. Hence, they should also be close in CD. In order to test if a given set of species S form a genus,

the following must be done. We commence by isolating a given set of species S^+ containing S and having the two following properties:

1) they are phyletically close;

2) their representations seem to form an evolutionary peak.

We now consider the phyletic closure of S^+ and apply to it a test to determine if it is an evolutionary peak small enough to be naturally considered a genus, and not a family or phylum. If not, we try again after having removed some of the elements of S. The set of species obtained in the last step of this process will be called the genus of S. We can now say that a genus is a given set of species S, such that S is equal to its phyletic closure and such that the set of representatives of S form an evolutionary peak small enough to be arbitrarily acceptable as a genus.

MATHEMATICAL FORM OF THE MODEL

Notations:

$c =$ subset

$\varepsilon =$ element of, in

$\Theta, / =$ such that

$\longleftrightarrow =$ if and only if

$\Xi =$ there exist

$\Psi =$ for every

$\Pi =$ cartesian product

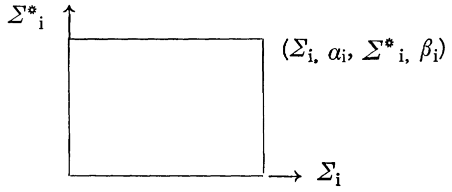
$U =$ union

1. Definition: Let us define Σ_i, the i-th Σ-gene.

2. Definition: Let us define Σ^*_i the gene corresponding to Σ_i (equal to Σ_i). $\Sigma^*_{i,} = \{\Sigma^* i, \beta_i\}$. $\Sigma_i = \{\Sigma_{i,}\ \alpha_i\} = \{\Sigma_{i,\ o}, \Sigma_{i,\ 1}, \ldots, \Sigma_i\ v_i\}$.

   The two different alleles of a single gene that we find in an individual form a pair $(\Sigma_i, \alpha_i, \Sigma^*_{i,} \beta_i)\ \varepsilon\ \Sigma_i \times \Sigma^*_i$

Example of a cartesian product:



3. Definition: Let us define $P = \prod_{i=1}^{K} (\Sigma_i \times \Sigma^*_i)$.

4. Definition: Let us define $\overline{P}\ c\ P\ \Theta\ x\ \varepsilon\ \overline{P} \longleftrightarrow (\Psi i = 1, \ldots, K)$ $(\Sigma_{i,}\ \alpha_i = \Sigma_{i,\ o} \longleftrightarrow \Sigma^*_{i,}\ \beta_i = \Sigma^*_{i,\ o})$.

5. Definition: x is a genetic individual $\longleftrightarrow x\ \varepsilon\ \overline{P}$.

6. Definition: $\Psi\ x\ \varepsilon\ \overline{P}$, let us define $G(x) = \{\Sigma_i / (\Sigma_{i,} \alpha_i, \Sigma^*_{i,} \beta_i) \neq (0,0) \text{ in } x\}$.

7. Definition: The species of x $= \overline{G(x)} = \{y/G(y) = G(x)\}$.

8. Definition: Let us define $\{A_i\}_{i=1}^{r}$ the set of external conditions, and let us define

$$CD = \prod_{i=1}^{r} A_i.$$

9. Definition: The vital neighborhood of the genetic individual x according to the i-th coordinate is the interval $V_i^x = (a_i, b_i) \; c \; A_i$ in which this individual can theoretically exist.

10. Definition: The total vital neighborhood of a genetic individual x is $\prod_{i=1}^{r} V_i^x = W_x \; c \; \prod_{i=1}^{r} A_i$.

11. $[x] = \bigcup_{\alpha \varepsilon \overline{G(x)}} W_\alpha$ represents the species of x such as in nature. Let $Q = \{[x]_i\}_{i=1}^{n}$.

12. Definition: $\overline{(G(x)}_1, \ldots, \overline{G(x)}_n) =$ last common ancestor of the set of species $\{\overline{G(x)}_1, \ldots, \overline{G(x)}_n\}$.

Definition: $[\overline{G(x)}_i) =$ the set of all species descending from $\overline{G(x)}_i$.

Definition: $[(\overline{G(x)}_1, \ldots, \overline{G(x)}_n)) =$ phyletic closure of the set of species $\{\overline{G(x)}_1, \ldots, G(x)_n\}$.

13. Definition: a function d: $X \times X \rightarrow R$ (reals) is called a metric on the vector space X if
    1. $d(x,y) \geqslant O \; (= 0 \longleftrightarrow x = y)$
    2. $d(x,y) = d(y,x)$
    3. $d(x,y) \leqslant d(x,z) + d(z,y)$

    where x, y and z denote points of the space.

    Definition: a subset A of the metric space (X,d) is said to be of diameter $\delta(A) = \max_{x,y \, \varepsilon \, A} d(x,y)$.

    Definition: a group of species R having as representatives the set $R^* = \{[x]_i\}$, $R^* \; c \; Q$, is said to form an evolutionary peak when $\Psi [x]_i \; \varepsilon \; R^* \; \Psi \; [x]_j \; \varepsilon \; R^* \; \Xi [x]_k \; \varepsilon \; R^* \; \Xi \; d([x]_i, [x]_j) < d([x]_i, [x]_k)$.

14. Definition: a genus is a set of species S such that
    1. $S = [(S))$
    2. the set of representatives of S form an evolutionary peak
    3. the diameter of this set of representatives is small enough to correspond to the intuitive concept of a genus.

References

Anderson, E. 1957 – An experimental investigation of judgements concerning genera and species. Evolution, 11: 260–261.

Cain, A. J. 1965 – The genus in evolutionary taxonomy. Systematic Zoology, 5: 97–109.

Davis, P. H. & V. H. Heywood 1963 – Principles of angiosperm taxonomy. Edinburgh.

Dobzhansky, T. 1951 – Genetics and the origin of species. Columbia Univ. Press, New York. 364 pages.

Kirpicznikov, M. E. 1968 – On the concept of genus in flowering plants. Bot. Zhurnal, 53: 190–201.

Lehmann, H. 1967 – Are biological species real? Philos. of Science, 34: 157–167.

Löve, Áskell 1963 – Cytotaxonomy and generic delimitation. Regnum Veget., 27: 45–51.

—— 1964a – The evolutionary framework of the biological species concept. Genetics Today, Proceedings of the XI International Congress of Genetics, p. 409–415.

—— 1964b – The biological species concept and its evolutionary structure. Taxon, 13: 33–45.

Mayr, Ernst 1942 – Systematics and the origin of species. New York.

—— 1965 – Numerical phenetics and taxonomic theory. Systematic Zoology, 14(2): 73–97.

Simpson, G. G. 1961 – Principles of animal taxonomy. New York.

Tutin, T. G. 1956 – Generic criteria in flowering plants. Watsonia, 3: 317–323.