Thus $\phi(x,t)$ is on the orbit of ψ through x whenever $|t| \leqq T$.

References

1. L. S. Pontryagin, Ordinary differential equations, Pergamon, Long Island City, N. Y., 1962.
2. J. L. Kelley, General topology, Van Nostrand, Princeton, N. J., 1955, 221–225.

RECONSTRUCTING THE HISTORY AND GEOGRAPHY OF AN EVOLUTIONARY TREE

DAVID SANKOFF, Université de Montréal

1. Introduction. In the process of phylogenesis a species splits into two or more populations which evolve independently into distinct varieties. Later, any of these may in turn split. As time progresses, current populations which stem from different branches of an earlier split may constitute distinct species, genera, families, etc. Biologists have traditionally represented this process in terms of tree diagrams, as in Figure 1a. At each time $t \in [-T, 0]$ where $-T$ is the date of the first split, and the present is time zero, a tree consists of a number of populations, each of which is the forerunner or ancestor of a certain subset of the present-day populations (e.g., Figure 1b).

DEFINITION 1. An evolutionary tree on a finite set S is a family $\{\mathscr{P}_t\}_{-T}^{0}$ of partitions of S, where

$$\mathscr{P}_{-T} = \{S\},\ \mathscr{P}_0 = \{\{X\} \mid X \in S\},$$

$$-T \leqq t \leqq u \leqq 0 \Leftrightarrow \mathscr{P}_u \text{ is a refinement of } \mathscr{P}_t$$

and $\lim_{t \uparrow u} \mathscr{P}_t = \mathscr{P}_u$.

DEFINITION 2. Let $\{\mathscr{P}_t\}_{-T}^{0}$ be an evolutionary tree on S. Every subset $X \subseteq S$ where $X \in \mathscr{P}_t$ for some $t \in [-T, 0]$, denotes a **population** in the tree. We shall have occasion to distinguish X_t, population X at time t, from X_u, the same population at time u, for $X \in \mathscr{P}_t \cap \mathscr{P}_u$. If $t < u$, we say X_t is **ancestral** to X_u. A population X is ancestral to a population Y if $Y \subset X$, and then we may also say X_t is ancestral to Y_u for all $\mathscr{P}_t$, $\mathscr{P}_u$ where $X \in \mathscr{P}_t$, $Y \in \mathscr{P}_u$.

The major problem in genetic taxonomy is as follows. Given a set S of genetically

---

David Sankoff received his McGill Univ. Ph.D. in 1969 under D. A. Dawson. His unusually wide background in statistics, biology, social sciences, and linguistics includes research assistantships in mathematics, sociology, anthropology, and anatomy at McGill and field work over several years in New Guinea. He is a member of the Univ. de Montréal Centre de recherches mathématiques. *Editor.*

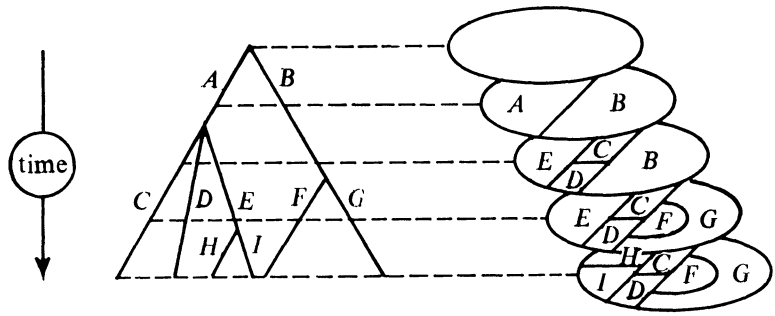

FIG. 1a FIG. 1b

related, currently existing (at time $t = 0$), populations, how can their evolutionary tree be deduced? In the next section we study a model of genetic divergence where, once split apart, populations evolve completely independently of one another. In this case reconstruction of the evolutionary tree from data on the existing populations is quite easy. This model is appropriate for trees which contain different genera, families, classes, etc., which *do* evolve relatively independently.

For evolutionary trees of populations which all belong to the same *species*, however, the problem is much more difficult since there may be interactions, i.e., interbreeding, between the various branches. In Section III we develop a model for this more interesting genetic divergence process, in terms of which we can solve the reconstruction problem.

2. Genetic divergence; independent populations. The similarity between two populations,

$$s(X_t, Y_u) = s(Y_u, X_t) \geqq 0,$$

is measured by the proportion of gene **types** they have in common. More specifically, there is some fixed set Γ of genetic **sites**, and at each site the two populations either have the same gene type or two completely different types. (We ignore the small proportion of gene sites for which there may be different types within a single population.) We assume Γ sufficiently large that we can neglect statistical fluctuation in the dynamic models we shall discuss.

Note that

$$s(X_t, X_t) = 1. \tag{1}$$

The simplest quantitative model of evolutionary divergence posits that in Γ, each site has a constant probability r per unit time of undergoing a replacement event. Then the probability of a type remaining unreplaced over a time interval of length $u - t$ satisfies the differential equation

$$\frac{d\Pr(u-t)}{du} = -r\Pr(u-t), \tag{2}$$

(see Feller, [1] Chapter XVII). The assumption that Γ is large may be rephrased mathematically as an assumption that the proportion of sites escaping replacement will also satisfy this equation. (Were Γ small, (2) would hold only for the expected value of the proportion.) Under the hypothesis that once replaced, a type can never recur, it follows that similarity also obeys (2). In other words, for X ancestral to Y (including the case when $X = Y$ but $u \geqq t$),

$$\frac{ds(X_t, Y_u)}{du} = -rs(X_t, Y_u), \tag{3}$$

from which we immediately derive, for initial condition (1):

PROPOSITION 1. *For X ancestral to Y, $s(X_t, Y_u) = \exp[-r(u-t)]$.*

Under the further hypothesis that a new type cannot occur as an innovation in two or more populations, and interpreting independent evolution in terms of probabilistic independence, we have the following more general statement:

PROPOSITION 2. *For all X and Y*

$$s(X_t, Y_u) = \exp[-r(v-t)]\exp[-r(v-u)],$$

where v is the latest point of time at which there exists a population ancestral to both X_t and Y_u.

Proof. For X_t ancestral to Y_u, it is clear that $v = t$, in which case we use Proposition 1. Likewise for Y_u ancestral to X_t.

In all other cases there will be a most recent population Z ancestral to both X and Y. Let

$$v = \max\{\tau \mid Z \in \mathscr{P}_\tau\}.$$

The maximum exists because of the limit assumption in Definition 1. Then

$$s(Z_v, X_t) = \exp[-r(v-t)],$$

$$s(Z_v, Y_u) = \exp[-r(v-u)].$$

By independence, the probability of a site being unaffected by replacement both between Z_v and X_t, and between Z_v and Y_u, is the product of the probabilities for the individual events. The same product relation holds for proportions of types unreplaced, by our assumption about Γ. The hypothesis of uniqueness of innovation ensures that the coefficient of similarity between X_t and Y_u will be precisely the proportion of sites unaffected by replacement in both evolutionary branches. Hence

$$s(X_t, Y_u) = s(Z_v, X_t)s(Z_v, Y_u)$$

which proves the proposition.

An **ultrametric** space (S, d) is a metric space where, for $W, X, Y \in S$,

$$d(X, Y) \leqq \max\{d(X, W), d(Y, W)\}. \tag{4}$$

At time zero, i.e., the present, let S be the set of populations currently representative of a given evolutionary tree. Without ambiguity, we can write X for $X_0 = \{X\}$.

For $X, Y \in S$ let $v(X, Y)$ be the time of the most recent common ancestor of X and Y as defined in Proposition 2.

PROPOSITION 3. *The pair $(S, -v)$ is an ultrametric space.*

Proof. Clearly $-v(X, Y) = 0$ if and only if $X = Y$; and $-v(X, Y) = -v(Y, X)$. It remains to prove (4), the ultrametric inequality (which implies the triangle inequality required of a metric). Suppose it does not hold and for some $W, X, Y \in S$

$$-v(X, Y) > \max\{-v(X, W), -v(Y, W)\}.$$

Then X and W have a more recent common ancestor population $Z^{(1)}$ than do X and Y, and Y and W have a more recent common ancestor $Z^{(2)}$ than do X and Y. But by Definitions 1 and 2, the ancestors of W form a nested sequence of subsets of S. Therefore one of $Z^{(1)}$ or $Z^{(2)}$ must be a common ancestor to both X and Y, contrary to our supposition. Hence the ultrametric inequality holds.

PROPOSITION 4. *Let S represent a finite set of populations existing at time zero. An ultrametric d on S determines a unique evolutionary tree where, if $X, Y \in S$, then $-d(X, Y)$ is the date of the most recent population ancestral to both X and Y.*

Proof. There are a finite number of different values of d, say $0 < d_1 < \cdots < d_m = -T$. For each $X \in S$, consider the nested sequence of sets

$$X_0 = \{X\} \subseteq \cdots \subseteq X_{-d_i} = \{Y \in S \mid d(X, Y) \leqq d_i\} \subseteq \cdots \subseteq X_{-T} = S.$$

The ultrametric inequality (4) assures, for any two such sequences $X_0, \cdots, X_{-T}$ and $Y_0, \cdots, Y_{-T}$, there is an integer p satisfying

$$\begin{aligned} X_t \cap Y_t &= \varnothing \quad \text{for } t = 0, -d_1, \cdots, -d_p \\ X_t &= Y_t \qquad \text{for } t = -d_{p+1}, \cdots, -d_m. \end{aligned} \tag{5}$$

Let $\mathscr{P}_0 = \{\{X\}\ X \in S\}$, and, for $t = d_i$, let $\mathscr{P}_t$ be the set of distinct X_t. For $-d_i < t \leqq -d_{i-1} = u$, let $\mathscr{P}_t = \mathscr{P}_u$, for $i = 1, \cdots, m$. From (5) it follows that $\{\mathscr{P}_t\}_{-T}^{0}$ satisfies Definition 1. For any $X, Y \in S$, our construction assures that X and Y are in the same element of $\mathscr{P}_t$ for t up to and including $t = -d(X, Y)$. This is precisely the ancestry condition required by the theorem, and it uniquely determines $\{\mathscr{P}_t\}_{-T}^{0}$.

Propositions 2–4 provide a solution to the reconstruction problem. The biologist first measures the similarities between the populations in S. Using the special case of Proposition 2 where $u = t = 0$, he solves

$$v(X, Y) = \frac{1}{2r} \log s(X, Y), \quad \text{for all } X, Y \in \boldsymbol{S}.$$

By Proposition 3, $(\boldsymbol{S}, -v)$ is an ultrametric space which, by Proposition 4, uniquely determines the evolutionary tree of $\boldsymbol{S}$. In fact, the proof of this latter proposition includes a construction of the tree.

3. Divergence with interaction. To be able to treat the case where changes in one population can be influenced by another, we add a geographical dimension to our hitherto purely historical considerations. At any point $t \in [-T, 0]$, each population in $\mathscr{P}_t$ will be associated with a face of a planar graph $\mathscr{M}_t$. This is illustrated in Figure 2.

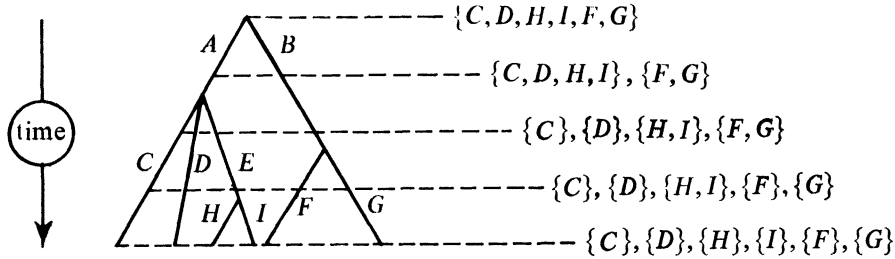

FIG. 2

For any tree $\mathscr{M}_{-T}$ is a loop, or degenerate graph consisting of one edge, one face and no vertices, as in Figure 2. If the split at time $-T$ is into two populations, then $\mathscr{M}_t$, for t immediately after $-T$, is a graph consisting of two faces, three edges (two exterior, one interior) and two vertices. Whenever a population splits into n fragments, the face corresponding to it is subdivided into two portions, then one of these two is further subdivided, then one of the resulting three is chosen for further subdivision, and so on, until an n-way fragmentation is achieved. The subdivision of a face is accomplished by choosing any two distinct edges bordering that face, placing a new vertex midway along each of these edges and joining the two vertices with a new edge. Alternatively, if the face has an exterior border (the ocean!), two new vertices on this single edge may be joined by a new edge.

DEFINITION 3. A geography associated with an evolutionary tree $\{\mathscr{P}_t\}_{-T}^{0}$, is a family of planar graphs $\{\mathscr{M}_t\}_{-T}^{0}$ where there is 1-1 correspondence between the populations of $\mathscr{P}_t$ and the faces of $\mathscr{M}_t$, satisfying

(a) $\mathscr{M}_{-T}$ is a loop,

(b) for two successive refinements $\mathscr{P}_t$, $\mathscr{P}_u$

$$\boldsymbol{X} \in \mathscr{P}_t,\ \boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(n)} \in \mathscr{P}_u,\ \boldsymbol{X} = \bigcup_{i=1}^{n} \boldsymbol{X}^{(i)}$$

$\Rightarrow \mathscr{M}_u$ is derived from $\mathscr{M}_t$ by the subdivision of the face corresponding to $\boldsymbol{X}$ into the faces corresponding to $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(n)}$.

This is the simplest way of constructing planar graphs by extension and hence is the simplest model of the evolution of territorial configurations of related populations.

How do populations interact? Instead of just replacing types at sites in Γ with completely new types, we now allow, in addition, the adoption of types from neighboring populations. Two neighboring populations are, of course, populations whose corresponding faces share an edge.

If X and Y are neighbors, we write

$$X \in N_Y \Leftrightarrow Y \in N_X.$$

We can construct models where the total replacement rate is constant but the proposition of adoptions depends on the number of neighbors, other models where new replacements occur at a constant rate but the adoption rate depends on the number of neighbors, or models where the adoption rate is constant. Mathematically speaking, these all lead to the same type of problem, and so we study just the last one.

We shall describe the genetic divergence process between two successive splits. In this interval $\mathscr{M}_t$ and $\mathscr{P}_t$ are fixed, so we can suppress the time subscripts on populations without risking ambiguity.

For each population X, we assume a probability rate r for new replacements as before, and probability rate $a/k(X)$ for adoptions from each of its $k(X)$ neighbors. Suppose $X \in N_Y$. Then $ds(X_t, Y_t)/dt$, the rate of change in the similarity between the two simultaneously evolving populations X and Y, has several components. There is the change due to new replacements, into X and into Y; the change due to adoptions from X into Y and vice-versa; and finally the change due to adoptions into X and Y from their other neighbors. For the first component, the same arguments which justify (2) and (3) in the case of a single evolutionary line, also imply that the change rate due to new replacements into the *two* populations is $-2rs(X_t, Y_t)$. (Were all other components zero, this could also be derived directly from Proposition 2, where $t = u$.) For the next component, the total adoption rate between X and Y is $a(1/k(X) + 1/k(Y))$ but a proportion $s(X_t, Y_t)$ of types adopted are already identical in the two populations so that the change rate due to this process will be $(1 - s(X_t, Y_t))\, a(1/k(X) + 1/k(Y))$. In addition we must take into account adoptions from the remaining $k(X) - 1$ neighbors of X and the remaining $k(Y) - 1$ neighbors of Y. Adoptions from neighbors of X change the similarity at a rate

$$\frac{a}{k(X)-1} \sum_{Z \in N_X - Y} [(1 - s(X_t, Y_t))s(Y_t, Z_t) - s(X_t, Y_t)(1 - s(X_t, Z_t))], \tag{6}$$

following the same line of reasoning, and adoptions from neighbors of Y have an analogous effect, but with Y and X interchanged in (6).

Collecting terms, we find that

$$\frac{ds(X_t, Y_t)}{dt} = -\beta(X, Y)s(X_t, Y_t) + \alpha(X, Y), \tag{7}$$

where

$$\alpha(X, Y) = a\left\{\frac{1}{k(X)} + \frac{1}{k(Y)}\right\} + \frac{a}{k(X)-1} \sum_{Z \in N_X - Y} s(Y_t, Z_t) + \frac{a}{k(Y)-1} \sum_{Z \in N_Y - X} s(X_t, Z_t),$$

$$\beta(X, Y) = 2r + \alpha(X, Y) + \frac{a}{k(X)-1} \sum_{Z \in N_X - Y} (1 - s(X_t, Z_t)) + \frac{a}{k(Y)-1} \sum_{Z \in N_Y - X} (1 - s(Y_t, Z_t)). \tag{8}$$

For two populations X and Y which are not neighbors, coefficients $\alpha(X, Y)$ and $\beta(X, Y)$ are as in (8) but without the term $a(1/k(X) + 1/k(Y))$.

PROPOSITION 5. *Let $\{\mathscr{P}_t\}_{-T}^{0}$ be an evolutionary tree with associated geography $\{\mathscr{M}_t\}_{-T}^{0}$. If genetic divergence proceeds according to* (7), *then $\mathscr{P}_0$, $s(X_0, Y_0)$ for all X_0, $Y_0 \in \mathscr{P}_0$, and $\mathscr{M}_0$ uniquely determine the tree and its geography.*

Proof. The graph $\mathscr{M}_0$ summarizes all neighboring relations between populations in $\mathscr{P}_0$. These relationships are fixed as far back as $\mathscr{P}_t$ remains unchanged. Therefore we can write down equation (7) explicitly, with initial conditions $s(X_0, Y_0)$. The system of first-order equations so obtained satisfies conditions for a unique solution and can be solved by successive approximation. We write the solution as s'.

Suppose the most recent population split was at time v, when populations W and Z were formed from population $\{W, Z\}$. Immediately after v, and any time $s(W_t, Z_t)$ is close to 1,

$$\frac{ds(W_t, Z_t)}{dt} < 0,$$

as can be seen in (6) or (8). Thus, $s(W_t, Z_t) < 1$ on $(v, 0]$. But, by condition (1) at v,

$$\lim_{t \downarrow v} s(W_t, Z_t) = 1.$$

Then time v, and the populations W and Z can be found as

$$v = \max\{\tau \mid \exists X, Y \in \mathscr{P}_0, s'(X_\tau, Y_\tau) = 1\},$$

and $s' = s$ on $(v, 0]$.

The graph $\mathscr{M}_v$ is then constructed by deleting the edge between the faces corresponding to W and Z. Note that by continuity

$$\lim_{t \downarrow v} s(X_t, W_t) = \lim_{t \downarrow v} s(X_t, Z_t) \text{ for all } X \in S,$$

since s measures proportions of types shared by two populations.

We now have $\mathscr{P}_v$, $s(X_v, Y_v)$ for all X, $Y \in \mathscr{P}_v$, and $\mathscr{M}_v$. We can then set up a new system of equations (7) with initial conditions $s(X_v, Y_v)$ and solve as before. The new solution s'' will be valid as far back as the second most recent split, and so on. The generalization to n-way splits, and the case where more than one population splits at the same instant, are obvious. We continue the solution procedure until we have deleted the last non-exterior line of the graph, which gives us $-T$ and $\mathscr{M}_{-T}$. This construction, for which each step is uniquely determined, proves the proposition.

This last proposition means that a biologist, equipped with similarity data as well as a knowledge of the geographical configuration of a number of currently existing related populations, can reconstruct the entire evolutionary tree of the populations, as well as the geographical configuration at all times in $[-T, 0]$.

4. Discussion. There are a number of practical problems associated with the theory of both Section II and Section III. One is that Γ is too small to ignore statistical fluctuation. Another is that r and a are not universal constants but may change somewhat from site to site in Γ and from population to population. The hypotheses about non-recurrence of innovation are not always justified. When these factors are taken into account, the reconstruction methods we have described must be bolstered by search algorithms and statistical estimation. Some useful references are: Dayhoff [2], Sokal and Sneath [3], Lerman [4], and Jardine and Sibson [5].

References

1. William Feller, An Introduction to Probability Theory and Its Applications, 2nd ed, Wiley, New York, 1957.

2. M. O. Dayhoff, Computer analysis of protein evolution, Scientific American, 221 (1969) 86–95.

3. Robert R. Sokal and Peter H. Sneath, Principles of Numerical Taxonomy, Freeman, San Francisco, 1963, (2nd ed. forthcoming).

4. I. C. Lerman, Les Bases de la Classification Automatique, Gauthier-Villars, Paris, 1970.

5. N. Jardine and R. Sibson, Mathematical Taxonomy, Wiley, New York, 1971.

LIPSCHITZIAN POINTS

E. M. BEESLEY, University of Nevada, Reno, A. P. MORSE, University of California, Berkeley, and D. C. PFAFF, University of Nevada, Reno

1. Introduction. Notations are explained in the next two paragraphs and the first paragraph of Section 2.

Throughout we understand: that R is the set of real finite numbers; that Rp is the set of real finite positive numbers; that ω is the set of nonnegative integers; with fractions in mind, that F is the set of rational numbers; that J is the open