Internodons as equivalence classes in genealogical networks: building-blocks for a rigorous species concept

D. J. Kornet[1], J. A. J. Metz[1] and H. A. J. M. Schellinx[2]

[1] Institute of Theoretical Biology, University of Leiden, Kaiserstraat 63, 2311 GP Leiden, The Netherlands. E-mail: Kornet@RULSFB.LeidenUniv.NL
[2] Équipe de Logique Mathématique, Université Paris VII, Tour 45-55, 5ème étage, 2 Place Jussieu, F-75251 Paris Cedex 05, France

Abstract. Among the options suggested in phylogenetic systematics to solve the species problem is the Hennigian or internodal species concept. This concept interprets species as parts of the genealogical network of individual organisms between two successive permanent splits or between a permanent split and an extinction event. Though this option is at present not favoured by phylogeneticists, we believe that, to solve the species problem, there is no alternative to finding a satisfactory partition of the genealogical network. In previous work a formal definition has been developed of Hennigian or internodal species (called internodons here), based on a logical relation between individual organisms. In this paper, we prove that this definition indeed partitions genealogical networks exhaustively into mutually exclusive entities, by showing that the defining relation is an equivalence relation. Although internodons should not themselves be seen as species, they are essential building-blocks for any satisfying species concept.

1 Introduction

Though the concept of species is a basic one in biology, there has never been agreement on its definition. This disagreement, which has persisted at least since the time of Darwin, is referred to as 'the species problem'. The controversy revives repeatedly, in particular in systematics, the discipline concerned with discovering, describing, and naming species. In recent decades, with the rise of phylogenetic systematics, founded by Hennig (1966), the species problem has received attention from a fresh perspective. This perspective has stressed to a greater extent the nature of species as historical entities, individuals even, which can arise and become extinct (Ghiselin, 1974; Hull, 1976).

Phylogenetic systematists have proposed several definitions of the species concept, but still none of these has appeared good enough to win general acceptance.

In the face of such deep disagreement, the way to a solution is surely to identify some tenets to which all participants can be expected to assent. One such is the proposition that the overwhelming majority of individual organisms of the past and present are members of one genealogical network. (Our reasoning, however, holds as well in the case of two or more separate networks.) This suggests that the species problem should be interpreted as the problem of demarcating groups of individual organisms within a genealogical network on particular criteria. Taking this view to its logical conclusion, we think that a species definition should and will find general acceptance only if it partitions genealogical networks of individual organisms exhaustively into mutually exclusive and historically continuous parts. The species definitions which have so far been put forward fail to meet this requirement. We first review how some well-known species definitions propose to demarcate species in the genealogical network.

Morphological species concepts, which define species by reference exclusively to the similarity of organisms, meet two principal problems. First, these definitions require the existence of clear gaps in similarity if they are to group organisms into mutually exclusive sets. This presupposition is vulnerable to challenge for organisms in any one time-slice, but is certainly untenable for organisms in different time-slices: it will generally be impossible to find gaps in morphological similarity among the organisms belonging to an ancestral line. Not surprisingly, adherents of the morphological species concept portray speciation as a lengthy process, during which it is difficult to tell, even in principle, whether organisms belong to the ancestor or the descendant species. Secondly, morphological species definitions run the risk of generating a partition of the network into parts which are not historically continuous, since the morphological similarity of organisms is not invariably evidence that they share a recent common ancestor. In the light of these problems, the morphological species concept has justifiably failed to win general acceptance among biologists.

The so-called biological species concept, advanced principally by Mayr (1940), though very popular, also violates the condition for general acceptance. In the biological species concept, the criterion on which organisms are grouped into species is their interbreeding ability. This criterion excludes all asexual organisms that may arise within the genealogical network. The biological species concept therefore fails to partition the whole of the genealogical network into species.

The biological species concept meets serious problems also in drawing the boundaries of species as parts of the genealogical network. The defining criterion 'interbreeding ability' cannot partition genealogical networks into mutually exclusive parts, because it is a non-transitive relation: suppose recent organisms a, b, and c belong to populations of a ring species which are just far enough apart that, while b can interbreed with both a and c, a cannot

interbreed with c. To accommodate the phenomenon of ring species, we could relax the criterion of interbreeding ability, and accept an indirect connection by interbreeding ability as a sufficient criterion for conspecificity. This option will however cause serious difficulties in the delimitation of biological species in the time dimension. It is likely that the criterion of indirect interbreeding ability will unite in the same species a present-day organism and an ancestor of it from early evolution.

The difficulty it meets in determining species boundaries, particularly in the time dimension, as well as its limited operationality, have prevented the general acceptance of the biological species concept. Mayr was well aware of the limitations of the biological species concept, and consequently came to suggest (1978: 85) that it should be interpreted as a "non-dimensional concept", to be applied only at one particular time and place. This is of course equivalent to admitting that the biological species definition is unable to deal with the genealogical network of individual organisms as a whole.

The same defect besets the so-called phylogenetic species concept (Cracraft, 1983; Nelson and Platnick, 1981; Nixon and Wheeler, 1990), in which species are defined as the smallest groups of organisms living at a particular time that are recognizable by a unique combination of properties shared by their members. Since this species concept was designed to apply only to organisms of a particular time-slice, it not surprisingly lacks the ability to find boundaries in the genealogical network in the time dimension.

One may wonder, at this stage, whether it is at all possible to partition the genealogical network exhaustively into mutually exclusive and historically continuous entities. If not, the principle that a species definition should and will find general acceptance only if it meets this criterion implies that no satisfactory species definition will ever be found. This may mean that we will end up with pluralism: we will find ourselves adopting a different species concept in each context (taxonomy, paleontology, ecology, etc.), rendering statements about species which transcend these contexts worthless, thereby impeding communication between workers in different fields.

But this conclusion is unwarranted. There is a group of species concepts which explicitly refer to the structure of the genealogical network of individual organisms. Hennig stated that speciation takes place by the splitting up of parts of the genealogical network; he illustrated his view with diagrams representing a part of the network containing a split (Hennig, 1966: 19, Fig. 4). Hennig does not further qualify the properties of splits that constitute speciation events; for example, he did not distinguish between temporary and permanent splits. Hennig's view suggests that we can interpret species as parts of the genealogical network contained between two successive splits. Nixon and Wheeler (1990: 213) introduced the name 'internodal species concept' for species concepts of this kind. The evolutionary species concept of Wiley (1981), which builds upon Simpson (1961), also appears to construe speciation events as splits in the genealogical network, though this is not stated explicitly. We can infer this view from, for instance, a diagram of Wiley in which he relates character evolution to speciation events (Wiley, 1981: 125, Fig. 5.3).

Ridley (1989) elaborated Hennig's definition, allowing temporary splits to be speciation events.

In fact, if Hennig's approach is to be successful, only splits that are permanent can plausibly be interpreted as speciation events, if only because temporary splits are very numerous in the network, opening up for instance between any two siblings that do not interbreed (Kornet, 1993). Under this qualification, the approach of partitioning the genealogical network into entities bounded by two successive splits, or between a split and an extinction, seems a promising procedure to cut up the network exhaustively into mutually exclusive parts. The entities delimited this way may indeed meet the criterion for general acceptance of a species concept.

To investigate whether this suggestion could be pursued further, Kornet (1993) replaced the informal definition of internodal species as suggested by Hennig by a relation employing formal logic and set theory. (This approach is not unprecedented in biology: many tools were developed by Woodger 1952 and Gregg 1954, and applied in evolutionary biology by e.g. Williams 1970.) It appeared possible to define internodal 'species' by a relation **INT**, which groups all organisms that are connected to each other by this relation into internodal 'species'. This paper provides the proof that **INT** is an equivalence relation. Being an equivalence relation, **INT** partitions the genealogical network exhaustively into mutually exclusive sets.

As explained by Kornet and McAllister (subm.), the equivalence classes delimited by **INT** in the genealogical network – that is, informally speaking, the entities contained between two successive permanent splits in the network – do not meet our intuitions about the life span of species. Because smaller branches of the genealogical network are continually becoming extinct, permanent splits occur very frequently. Each of these splits brings to a close one entity and gives rise to a new one. Therefore, the life span of these entities may well often be limited to a few generations of their member-organisms. For this reason, Kornet and McAllister declined to consider these entities as species, and introduced the name 'internodons' for them instead.

However, the partition generated by the **INT** relation retains its importance for a solution of the species problem. Though they cannot reasonably be interpreted as species, internodons have proved to be valuable building-blocks out of which, on the basis of a further criterion, longer-lived entities can be formed. These entities are moreover recognizable in practice by the application of morphological criteria. Kornet and McAllister (subm.) interpret these entities as species, introducing the concept of *composite species*.

It may be that the relation **INT** does not apply to the universal set U of all organisms. Organisms are ordinarily connected to other organisms by parental relationships which makes them members of a supra-organismal structure. The domain of **INT** is a subset of U consisting of **GN** organisms. Organisms with the property **GN** are members of a genealogical network, a structure of organisms connected by parental relationships which includes at least one occurrence of sexual reproduction. If life has arisen several times, the universe may contain more than one such genealogical network.

Any organisms of U that are not **GN** organisms are outside the domain of **INT**. These organisms may be members of so-called trees, independently arisen ancestor-descendant sequences of organisms in which no sexual reproduction has even taken place. While such trees may have existed in early evolution, it is unlikely that any remains today, so the exclusion of organisms forming such trees has minor impact on the definition of species. Finally, there may be organisms which neither have nor are parents. Possible candidates for that category are 'shoats' or 'geeps' which are artificially produced by mixing the embryos of a sheep and a goat, and which do not themselves reproduce. If it is denied that such organisms have parental relationships, they too are outside the domain of **INT** (see also Kornet, 1993).

The **INT** relation is based on an **INTSD** relation which applies only to a subset of **GN** organisms, the so-called **SD** organisms: organisms that either have interbred, or reproduced asexually but have at least one descendant that reproduced sexually. In the next two sections we will define **INTSD** and prove that it is an equivalence relation. Section 4 gives the definition of the **INT** relation, accommodating **INTSD** for **GN** organisms which are not **SD** **organisms, and shows that INT** too is an equivalence relation.

2 Definition of the INTSD relation

The route chosen differs somewhat from the one taken in Kornet (1993), which presents more of the biological background.

First, some notational conventions. Bold upper-case letters denote predicates, either single-argument or relational ones. If one of the arguments in a relational predicate is considered as fixed, and the predicate is thus transformed into a predicate having one argument fewer, this argument is indicated between brackets. Outlined letters denote the extensions of the single-argument predicates indicated by the same letters in bold type, i.e. the sets of all organisms to which those predicates apply. These single-argument predicates may of course themselves be derived from a relational predicate by fixing all but one of its arguments.

2.1 Primitive terms

x**P**y reads as 'x is parent of y'.

$x \geqq y$ reads as 'x was born at the same time as or after y'. By reasonable assumptions on the nature of time, $\geqq$ totally orders our universe: $\geqq$ is reflexive and transitive, and for all x, y it holds that $x \geqq y$ or $y \geqq x$. We will write $\leqq$ for the converse of this order: i.e. $x \leqq y$ reads as 'x was born at the same time as or before y'. Similarly we use $x < y$ for 'x was born before y', and $x \equiv y$ for 'x was born at the same time as y'.

For more discussion of the semantics of the primitive terms, see Kornet (1993).

2.2 Definitions and lemma

$$\text{The relation } \mathbf{A} \text{ is the transitive closure of } \mathbf{P} \qquad \text{(D1)}$$

$x \mathbf{A} y$ reads as 'x is ancestor of y'.

$$\text{The relation } \mathbf{C} \text{ is the symmetric closure of } \mathbf{P} \qquad \text{(D2)}$$

$x \mathbf{C} y$ reads as 'x is immediately connected to y'.

$$x(\mathbf{C}_{\geq p})y : \Leftrightarrow x \mathbf{C} y \wedge x \geq p \wedge y \geq p \qquad \text{(D3)}$$

$x(\mathbf{C}_{\geq p})y$ reads 'x and y are immediately connected and both are born after or at the same time as p'. Observe that relation $\mathbf{C}_{\geq p}$ is symmetric.

$$\text{The relation } \mathbf{PC}_{\geq p} \text{ is the reflexive and transitive closure of } \mathbf{C}_{\geq p} \quad \text{(D4)}$$

$x(\mathbf{PC}_{\geq p})y$ reads as 'x is path-connected to y so that no organism on the path is born before p'. Relation $\mathbf{PC}_{\geq p}$ clearly inherits symmetry from $\mathbf{C}_{\geq p}$. Observe that by definition, if $x(\mathbf{PC}_{\geq p})y$ and $q \leq p$, then $x(\mathbf{PC}_{\geq q})y$.

$$\mathbb{GDYN}(y) = \{x \mid x(\mathbf{PC}_{\geq y})y\} \qquad \text{(D5)}$$

$\mathbb{GDYN}(y)$ reads as 'the gross dynasty set of y'. An organism x belongs to the gross dynasty of y if x is path-connected to y such that no organism on the path is born before y. As $y(\mathbf{PC}_{\geq y})y$, obviously $y \in \mathbb{GDYN}(y)$.

$$x \mathbf{I} y : \Leftrightarrow (x \neq y) \wedge \exists q \{(x \mathbf{P} q \wedge y \mathbf{P} q)\} \qquad \text{(D6)}$$

$x \mathbf{I} y$ reads as 'x has interbred (i.e. shares offspring) with y'.

$$\mathbf{SD}x : \Leftrightarrow \exists p \, \exists q \{(p \mathbf{I} q) \wedge [(x = p) \wedge (x \mathbf{A} p)]\} \qquad \text{(D7)}$$

$\mathbf{SD}x$ reads as 'organism x possesses at least one sexually produced descendant'.

$$\mathbb{DYN}(y) = \{x \mid x(\mathbf{PC}_{\geq y})y \wedge \mathbf{SD}x\} \qquad \text{(D8)}$$

$\mathbb{DYN}(y)$ reads as 'the (proper) dynasty set of y'. The (proper) dynasty of y is the set of those organisms x of gross dynasty set $\mathbb{GDYN}(y)$ which are $\mathbf{SD}$ organisms.

A useful **lemma**: If $r \in \mathbb{DYN}(y)$ and $r \equiv y$, then $\mathbb{DYN}(y) = \mathbb{DYN}(r)$. (L1)

Proof. Let $x \in \mathbb{DYN}(y)$. Then $x(\mathbf{PC}_{\geq y})y$ and $\mathbf{SD}x$, and as $r \equiv y$ also $x(\mathbf{PC}_{\geq r})y$. Since $r \in \mathbb{DYN}(y)$ we have $r(\mathbf{PC}_{\geq y})y$, so $y(\mathbf{PC}_{\geq y})r$ by symmetry, and $y(\mathbf{PC}_{\geq r})r$ because $r \equiv y$. We conclude $x(\mathbf{PC}_{\geq r})r$ by transitivity, and, since $\mathbf{SD}x$, we find $x \in \mathbb{DYN}(r)$.

2.3 Definition of **INTSD**

Write E for the earlier born and L for the later born of organisms $\{x, y\}$, with the convention that, whenever $x \equiv y$, we take $E = x$ and $L = y$. (Observe that consequently E and L are always defined.) Further, let $\geq(x)$ denote $\{y \mid y \geq x\}$.

Suppose **SD**x and **SD**y, then

$$x\textbf{INTSD}y :\Leftrightarrow L \in \mathbb{DYN}(E) \wedge$$

$$\forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(E) \cap \geqq (r) = \mathbb{DYN}(r))\} \qquad \text{(D9)}$$

$x\textbf{INTSD}y$ reads as 'x and y are **SD** organisms that are co-SDinternodal'. Note that the relation **INTSD** is defined for **SD** organisms only; the **INT** relation is an extension of the **INTSD** relation for all **GN** organisms (organisms in genealogical networks), including non-**SD** organisms. First we will prove that **INTSD** is an equivalence relation.

3 Proof that INTSD is an equivalence relation

Theorem. INTSD *is an equivalence relation on* **SD**.

For **INTSD** to be an equivalence relation on the subset of **SD** organisms of U it must be proved that on that set **INTSD** has the properties of symmetry, reflexivity, and transitivity.

3.1 Symmetry

The symmetry of **INTSD** follows immediately from its definition (D9).

3.2 Reflexivity

Let $E = L = x$. We have to show that

$$x \in \mathbb{DYN}(x) \wedge \forall r\{(r \in \mathbb{DYN}(x) \wedge r \leqq x) \Rightarrow (\mathbb{DYN}(x) \cap \geqq (r) = \mathbb{DYN}(r))\}$$
$$\text{(r1)}$$

$x \in \mathbb{DYN}(x)$ holds since **SD**x and $x \in \mathbb{GDYN}(x)$ ((D5) and (D8)). For the second part of conjunction (r1), note that if $r < x$, it is impossible that $r \in \mathbb{DYN}(x)$; if $r \equiv x$ and $r \in \mathbb{DYN}(x)$ then by lemma (L1), $\mathbb{DYN}(x) = \mathbb{DYN}(r)$. By definition of $\mathbb{DYN}(x)$ we have $\mathbb{DYN}(x) \cap \geqq (x) = \mathbb{DYN}(x)$, so we find

$$\mathbb{DYN}(x) \cap \geqq (r) = \mathbb{DYN}(x) \cap \geqq (x) = \mathbb{DYN}(x) = \mathbb{DYN}(r) .$$

3.3 Transitivity

For **INTSD** to be transitive it must hold that

$$x\textbf{INTSD}y \wedge y\textbf{INTSD}z \Rightarrow x\textbf{INTSD}z \qquad \text{(t1)}$$

There are six possibilities for the relative times of birth of x, y and z with respect to $\leqq$, which by symmetry reduce to three:

$$x \leqq y \leqq z, \quad z \leqq y \leqq x, \quad \text{or} \quad \text{(I)} \ E \leqq y \leqq L \ ,$$

$$x \leqq z \leqq y, \quad z \leqq x \leqq y, \quad \text{or} \quad \text{(II)} \ E \leqq L \leqq y \ ,$$

$$y \leqq x \leqq z, \quad y \leqq z \leqq x, \quad \text{or} \quad \text{(III)} \ y \leqq E \leqq L \ .$$

We have to prove in each case that $E\mathbb{INTSD}L$ is implied.

I. We have to show that for $E \leqq y \leqq L$:

$$y \in \mathbb{DYN}(E) \wedge \forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq y) \Rightarrow (\mathbb{DYN}(E) \cap \geqq (r) = \mathbb{DYN}(r))\}$$
(t2a)

and

$$L \in \mathbb{DYN}(y) \wedge \forall r\{(r \in \mathbb{DYN}(y) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(y) \cap \geqq (r) = \mathbb{DYN}(r))\}$$
(t2b)

implies

$$L \in \mathbb{DYN}(E) \wedge \forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(E) \cap \geqq (r) = \mathbb{DYN}(r))\}$$
(t2c)

Ia. We start by proving $L \in \mathbb{DYN}(E)$.

The first part of the conjunction in (t2a) informs us that

$$y \in \mathbb{DYN}(E)$$

Obviously $y \leqq y$, so according to the second part of (t2a) it is implied that

$$\mathbb{DYN}(E) \cap \geqq (y) = \mathbb{DYN}(y) \tag{t3}$$

From (t2b) we know $L \in \mathbb{DYN}(y)$. So by (t3) it follows that $L \in \mathbb{DYN}(E)$.

Ib. This leaves us to prove the second part of the conjunction (t2c):

Let $r \in \mathbb{DYN}(E)$ and $r \leqq L$. We have to show

$$\mathbb{DYN}(E) \cap \geqq (r) = \mathbb{DYN}(r)$$

If $r \leqq y$, this is immediate by (t2a).
    Otherwise $y < r \leqslant L$.
    By (t3)

$$\mathbb{DYN}(E) \cap \geqq (y) = \mathbb{DYN}(y)$$

Since $r \in \mathbb{DYN}(E)$ and $r > y$, it follows that $r \in \mathbb{DYN}(y)$.
    Obviously, since $r > y$, $\geqq (y) \cap \geqq (r) = \geqq (r)$
    Therefore

$$\mathbb{DYN}(E) \cap \geqq (r) = \mathbb{DYN}(E) \cap \geqq (y) \cap \geqq (r) = \mathbb{DYN}(y) \cap \geqq (r) \tag{t4}$$

Since $r \in \mathbb{DYN}(y)$ and $r \leq L$, by (t2b)

$$\mathbb{DYN}(y) \cap \geqq(r) = \mathbb{DYN}(r)$$

So by (t4)

$$\mathbb{DYN}(E) \cap \geqq(r) = \mathbb{DYN}(r)$$

II. We have to show that for $E \leq L \leq y$:

$$y \in \mathbb{DYN}(E) \wedge \forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq y) \Rightarrow (\mathbb{DYN}(E) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t5a)}$$

and

$$y \in \mathbb{DYN}(L) \wedge \forall r\{(r \in \mathbb{DYN}(L) \wedge r \leqq y) \Rightarrow (\mathbb{DYN}(L) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t5b)}$$

implies

$$L \in \mathbb{DYN}(E) \wedge \forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(E) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t5c)}$$

IIa. We start by proving $L \in \mathbb{DYN}(E)$.

As we know from (t5b), $y \in \mathbb{DYN}(L)$, so by definition (see D8) we have $y(\mathbf{PC}_{\geqq L})L$ and by symmetry $L(\mathbf{PC}_{\geqq L})y$.

Since $E \leq L$ we also have

$$L(\mathbf{PC}_{\geqq E})y \tag{t6}$$

But by (t5a) also $y \in \mathbb{DYN}(E)$, so by definition

$$y(\mathbf{PC}_{\geqq E})E \tag{t7}$$

and by transitivity from (t6) and (t7) we have

$$L(\mathbf{PC}_{\geqq E})E$$

Since $\mathbf{SD}L$ it holds indeed that $L \in \mathbb{DYN}(E)$.

IIb. To prove the second part of the conjunction of (t5c) let $r \in \mathbb{DYN}(E)$ and $r \leq L$. We have to show

$$\mathbb{DYN}(E) \cap \geqq(r) = \mathbb{DYN}(r)$$

which we find immediately by (t5a), as $L \leq y$ implies $r \leqq y$.

III. We have to show that for $y \leq E \leq L$:

$$E \in \mathbb{DYN}(y) \wedge \forall r\{(r \in \mathbb{DYN}(y) \wedge r \leqq E) \Rightarrow (\mathbb{DYN}(y) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t8a)}$$

and

$$L \in \mathbb{DYN}(y) \wedge \forall r\{(r \in \mathbb{DYN}(y) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(y) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t8b)}$$

implies

$$L \in \mathbb{DYN}(E) \wedge \forall r\{(r \in \mathbb{DYN}(E) \wedge r \leqq L) \Rightarrow (\mathbb{DYN}(E) \cap \geqq(r) = \mathbb{DYN}(r))\}$$
$$\text{(t8c)}$$

IIIa. We start by proving $L \in \mathbb{DYN}(E)$.

As we know from (t8a)

$$E \in \mathbb{DYN}(y)$$

Obviously $E \leq E$, so according to the second part of (t8a)

$$\mathbb{DYN}(y) \cap \geq (E) = \mathbb{DYN}(E) \tag{t9}$$

From (t8b) we know $L \in \mathbb{DYN}(y)$. As $L \in \geq (E)$ we find by (t9) $L \in \mathbb{DYN}(E)$.

IIIb. To prove the second part of the conjunction of (t8c), let $r \in \mathbb{DYN}(E)$ and $r \leq L$. We have to show $\mathbb{DYN}(E) \cap \geq (r) = \mathbb{DYN}(r)$.

We know by (t9) $\mathbb{DYN}(y) \cap \geq (E) = \mathbb{DYN}(E)$. Since $r \in \mathbb{DYN}(E)$ it follows that

$$r \in \mathbb{DYN}(y) \quad \text{and} \quad r \in \geq (E) \tag{t10}$$

Moreover, if $r \in \geq (E)$ then $\geq (r) \cap \geq (E) = \geq (r)$. Therefore

$$\mathbb{DYN}(E) \cap \geq (r) = \mathbb{DYN}(y) \cap \geq (E) \cap \geq (r) = \mathbb{DYN}(y) \cap \geq (r) \tag{t11}$$

Since $r \in \mathbb{DYN}(y)$ (by (t10)) and $r \leq L$ it follows from (t8b) that

$$\mathbb{DYN}(y) \cap \geq (r) = \mathbb{DYN}(r)$$

So by (t11)

$$\mathbb{DYN}(E) \cap \geq (r) = \mathbb{DYN}(r)$$

We have proved for all conditions of relative order of birth (**I–III**) that

$$x \mathbf{INTSD} y \wedge y \mathbf{INTSD} z \Rightarrow x \mathbf{INTSD} z$$

holds and therefore that relation **INTSD** is transitive.

This ends the proof of our theorem that **INTSD** is an equivalence relation.

It should be noted that we have in fact shown how to partition an (abstract) network, i.e. an ordered set $(\mathbb{U}, \geq)$ together with a notion of connectedness for elements of $\mathbb{U}$ (the network's universe). An element x in $\mathbb{U}$ represented an organism, and with each x was associated a unique 'point in time', say x's 'moment of birth'.

The attentive reader will have observed that in the above proof no specific properties of the parenthood relation **P** are used; hence the construction might be carried out for *any* relation **P** on $\mathbb{U}$ whatsoever. Also, at no point were properties specific to the subset **SD** used; so the construction goes through for any subset $\mathbb{V}$ of $\mathbb{U}$ whatsoever (including $\mathbb{U}$ itself), giving rise to an equivalence relation **INTV** on $\mathbb{V} \subset \mathbb{U}$. We leave it to the reader to verify that the construction is *stable* in the following sense: if $\mathbb{V} \subset \mathbb{W}$, and x, y are $\mathbb{V}$-organisms such that $x \mathbf{INTW} y$, then $x \mathbf{INTV} y$. In other words, the partition of $\mathbb{V}$ induced by **INTW** is a *refinement* of that induced by **INTV**. Moreover, in general it will be a *proper* refinement, as the converse of the above is easily shown to be false.

4 Definition of the INT relation

The domain of the relation **INT** is the subset of U consisting of **GN** organisms, organisms that are members of a genealogical network. If the **PATH** relationship is the reflexive, symmetric, and transitive closure of the parenthood relation **P**, connecting all organisms of a network, then

$$\mathbf{GN}x :\Leftrightarrow \exists y \{x\,\mathbf{PATH}\,y \wedge \mathbf{SD}y\} \tag{D10}$$

To define the relation **INT**, the definition of **INTSD** has to be accommodated for **GN** organisms which are not **SD** organisms. For this we need the following relation:

$$x\,\mathbf{LBA}\,y :\Leftrightarrow (x\mathbf{A}y \wedge \mathbf{SD}x) \wedge \neg \exists r(\mathbf{SD}r \wedge x < r \wedge r\mathbf{A}y) \tag{D11}$$

$x\,\mathbf{LBA}\,y$ reads as 'x is the latest-born ancestor of y with at least one sexually produced descendant'.

The difference between the domains of **INT** and **INTSD** is now bridged by determining for organisms x and y test-organisms o_1 and o_2. If x and y are **SD** organisms, they serve as their own test-organisms. If, on the other hand, x and y are non-**SD** organisms, the test-organism of each will be its latest-born **SD** ancestor. (Note that it is possible that more than one organism, born at the same time, may satisfy this description, and therefore serve as test-organism.)

$$o\mathbf{T}x :\Leftrightarrow \{[\mathbf{SD}x \wedge (o = x)] \vee [\neg\,\mathbf{SD}x \wedge o\,\mathbf{LBA}x]\} \tag{D12}$$

$o\mathbf{T}x$ reads as 'o is test-organism for x'.

This enables us to give the definition of **INT**:

$$x\mathbf{INT}y :\Leftrightarrow \exists o_1 \exists o_2(o_1\mathbf{T}x \wedge o_2\mathbf{T}y \wedge o_1\mathbf{INTSD}o_2) \tag{D13}$$

$x\mathbf{INT}y$ reads as 'organism x belongs to the same internodon as organism y'.

Finally we show that **INT** inherits from **INTSD** the property of being an equivalence relation. The relation **INTSD** differs from **INT** in that **INTSD** applies only to **SD** organisms (organisms x for which it holds: $x \in \mathbb{SD}$), while **INT** ranges over **SD** and non-**SD** organisms. Whether a non-**SD** organism belongs to one and the same internodon as some other organism z depends entirely on whether the **SD** test-organism of the non-**SD** organism belongs to the same internodon as organism z. Therefore we are entitled to state that **INT** inherits from **INTSD** the property of being an equivalence relation on its domain, the set of **GN** organisms. In the (easy) formal verification one shows, to prove transitivity, that $o\mathbf{T}y$ and $o'\mathbf{T}y$ implies $o\mathbf{INTSD}o'$.

5 Discussion

In the preceding treatment, we built upon the interpretation of Hennigian or internodal species proposed by Kornet (1993). This interpretation offered, through the relation **INT**, a formalization of the notion of groups of organisms contained between permanent splits in the genealogical network. In the

present paper we prove that **INT** is indeed an equivalence relation, and therefore partitions its domain exhaustively into mutually exclusive sets.

To conclude, some broader considerations about this approach to species in biology. We believe that the formalization based on **INT** is both minimal and unique. We invite the reader either to give a counterexample or to prove that the relation **INT** as defined does not derive uniquely from the requirement that it should formalize the definition of Hennigian species as groups of organisms contained between two successive permanent splits, or between a permanent split and an extinction event, in the genealogical network.

Even though internodons as defined by **INT** are the ultimate consequence of adhering strictly to the basic Hennigian idea, they are, as explained in the introduction, very short-lived, contrary to the intuition of Hennig and his followers. Their short life span renders internodons poor candidates for the status of species. However, it has already been shown that, if grouped together into larger units on a suitable criterion, internodons are valuable building-blocks for a species concept (Kornet and McAllister, submitted). These larger units inherit the desirable properties of internodons, and yet have a longevity which conforms more closely to classical intuitions about historical species. Even this entity may not be the last word. We therefore invite the reader to join us in extending and developing the above proposals for a constructive approach to basic taxonomic methodology.

References

1. Cracraft, J.: Species concept and species analysis. Curr. Ornithol. **1**, 159–187 (1983)
2. Ghiselin, M. T.: A radical solution to the species problem. Syst. Zool. **23**, 536–544 (1974)
3. Gregg, J. R.: The Language of Taxonomy: An Application of Symbolic Logic to the Study of Classificatory Systems. New York: Columbia University Press (1954)
4. Hennig, W.: Phylogenetic Systematics. Urbana: University of Illinois Press (1966)
5. Hull, D. L.: Are species really individuals? Syst. Zool. **25**, 174–191 (1976)
6. Kornet, D. J.: Permanent splits as speciation events: a formal reconstruction of the internodal species concept. J. Theor. Biol. **164**, 407–435 (1993)
7. Kornet, D. J. and McAllister, J. W.: The composite species concept (subm.)
8. Mayr, E.: Speciation phenomena in birds. Amer. Nat. **74**, 249–278 (1940)
9. Mayr, E.: Origin and history of some terms in systematic and evolutionary biology. Syst. Zool. **27**, 83–88 (1978)
10. Mayr, E.: The Growth of Biological Thought. Cambridge, Mass.: Harvard University Press (1982)
11. Nelson, G. and Platnick, N. I.: Systematics and Biogeography. Columbia University Press, New York (1981)
12. Nixon, K. C. and Wheeler, Q. D.: An amplification of the phylogenetic species concept. Cladistics **6**, 211–223 (1990)

13. Ridley, M.: The cladistic solution to the species problem. Biol. Phil. **4**, 1–16 (1989)
14. Simpson, G. G.: Principles of Animal Taxonomy. (Reprinted, 1990.) New York: Columbia University Press (1961)
15. Wiley, E. O.: Phylogenetics: The Theory and Practice of Phylogenetic Systematics. New York: John Wiley (1981)
16. Williams, M. B.: Deducing the consequences of evolution: a mathematical model. J. Theor. Biol. **29**, 343–385 (1970)
17. Woodger, J. H.: From biology to mathematics. Brit. J. Phil. Sci. **3**, 1–21 (1952)