Contributions to Zoology, 67 (3) 197-220 (1998)
SPB Academic Publishing bv, The Hague

Analysis of a species/instars/characters table: a theoretical survey on the use of chaetotaxy in ontophylogenetic studies

Pierre Nayrolles
Laboratoire de Zoologie, Ecobiologie des Arthropodes édaphiques, Université Paul Sabatier, 118 route de Narbonne F-31062 Toulouse Cedex, France. E-mail: pn20@calva.net

Keywords: Ontogeny, phylogeny, chaetotaxy, probabilistic organs, ontogenetic trajectories, arthropods, Collembola, Symphypleona

Abstract

Phylogenetic studies on several groups of arthropods, such as Acari (mites) or Collembola (springtails), make wide use of chaetotaxy. Chaetotaxic characters, besides being morphological features (setal shape), often have a binary nature, that is presence vs. absence. A seta can be variable in a population, and one may attribute a presence probability to this seta. In fact, the presence probability should be defined for each instar. One might think that a variable seta corresponds to a polymorphism (e.g., two or more alleles in a population); in fact, setal variability should be regarded as the result of a propensity intrinsic to individuals, i.e., a potentiality being expressed at random among specimens. The phenomenon known as lateral inhibition explains how an early random fluctuation is at the root of the cell fate. Probabilistic organs are likely to originate from a similar phenomenon. A tridimensional species/instars/characters table (SIC table), containing setal presence probabilities, can be built up. Two kinds of analyses may be applied to such a table. In these analyses, the problem of homeotypic setae is emphasized. Related issues are the accuracy of using setal probabilities and the problem of reducing redundancy of information. A first process leads to a bidimensional table with characters in columns, and the species divided into their instars in rows. By using multidimensional statistics we can access ontogenetic trajectories, and in this way, ontogenetic comparisons can be achieved. The aim of the second process is to produce a species/characters table which can be used in a cladistic analysis.

Résumé

Dans l'étude phylogénétique de plusieurs groupes d'Arthropodes, tels que les Acariens ou les Collemboles, il est largement fait appel à la chétotaxie. Outre le fait qu'ils puissent être des traits purement morphologiques (forme des soies), les caractères chétotaxiques sont souvent de nature binaire, opposant présences et absences. Dans une population, une soie donnée peut se révéler variable, une probabilité de présence pouvant alors lui être attachée. Une soie variable pourrait être perçue comme la marque d'un polymorphisme (plusieurs allèles dans une population). En fait, cette variabilité doit être comprise comme la sommation de propensions individuelles, ou en d'autres termes, comme la résultante d'une potentialité propre à chaque individu exprimée au hasard. Vraisemblablement, un phénomène similaire à celui de l'inhibition latérale, phénomène qui explique comment une fluctuation au hasard intervenue précocement peut déterminer le destin de cellules, est à l'origine des organes probabilistes. Un tableau tridimensionnel espèces/stades/caractères contenant les probabilités de présence des soies peut être constitué. Deux sortes de traitements, dans lesquels le problème des soies homéotypes prend une place essentielle, sont alors applicables. Les autres questions relatives à ces analyses sont la pertinence de l'utilisation des probabilités de soies et la réduction de l'information redondante. Un premier traitement conduit à un tableau bidimensionnel dont les caractères sont en colonnes et, en lignes, les espèces divisées en stades. En utilisant des statistiques multidimensionnelles, il est possible de dessiner des trajectoires ontogénétiques, et donc de comparer les ontogenèses. Le but du deuxième traitement est l'obtention d'un tableau espèces/caractères qui pourra être utilisé dans une analyse cladistique.

Introduction

André (1988) emphasized that several authors were surprised that so little is known about insect immature stages. He explained this fact by the lack of a method for the treatment of ontogenetic data. The aim of this paper, as well as those by André, is to provide some theoretical ideas and several

techniques for studying arthropod ontophylogenetics (i.e., ontogenetics and phylogenetics considered together) by means of chaetotaxy, i.e., the study of the arrangement of organs – setae or, strictly speaking, sensilla – produced by the integument of arthropods. Once chaetotaxic patterns have been established, presence vs. absence or number of setae can be used to compare species.

A general scientific procedure by which characters would be generated cannot be defined, because of the great diversity of features used (e.g., morphological, behavioral, biochemical features, etc.). Nevertheless, in some fields, a standard method can be set up (e.g., morphometry).

This article focuses on the erection of characters from brute chaetotaxic data with a special attention paid to the problem of variable setae. Ontogeny, i.e., instar of appearance of setae, is used as a way to distinguish species, so that characters should embrace the ontogenetic dimension. Polarization of characters and phylogenetic inference are next steps and will not be discussed here. Even if phylogenetics is not directly addressed, this theme is underlying since building up characters is a preliminary phase before any phylogenetic – cladistic – work.

An example of data processing is taken from the group of Symphypleona, one of the major divisions of the Collembola (Hexapoda). The ontogenetic patterns of these arthropods are quite simple and constitute a good support for the method presented.

The postembryonic development of arthropods consists of a succession of forms separated by molts. The period of life between two molts is the "intermolt", also called "instar". This paper describes a method for analyzing a table with three dimensions, these dimensions being species, their instars, and characters, in other words a species/instars/characters (SIC) table. In principle, my method can be applied to all arthropod species exhibiting "stases" and associated discontinuous characters. The word "stase" was coined by Grandjean (1938) for designating the morphological stages succeeding one another during the development of mites and distinct from each other by "all or none" characters.

The discrete characters used for differentiating stases are generally based on the presence of setae. The characters considered in this paper concern only idionymic setae. Idionymy (Grandjean, 1949) is the "quality" of a particular organ as distinct from other organs of the same nature; such an organ occupies a defined position so that it can be designated by a name or a symbol, e.g., our vertebrae are idionymic while our hairs are not. Idionymy of a set of setae is established by comparisons of specimens within a species. Developmental stage or sex should also be considered because setae may be idionymic in larvae and not in adults. When setae are not idionymic, their positions are not constant but depend on their number, in other words the inter-setal distance is a function of the setal number. Conversely, when setae are idionymic, each seta has a fixed position, and if a seta is absent, the position of the neighboring setae is not modified.

Setae that appear during development are named "secondary setae", while "primary setae" are present from the first instar on. Sometimes, at a given instar, some setae are variable. Setal presence probabilities must be defined at the level of a population. In this paper, the concept of probabilistic organs will be emphasized.

Analysis of a SIC table permits ontophylogenetic studies. For instance, it is possible to carry out multidimensional statistics to shed light on similarities or dissimilarities among instars of species. A recent paper dealt with application of these statistical methods (Nayrolles, 1996), but missing from it was a discussion of the nature and possible uses of chaetotaxic characters. From my study of Symphypleona, it appears that data processing of chaetotaxic characters is rather burdensome because of the numerous setae that have to be taken into account. Computer software would make studies of SIC tables easier, but before undertaking such a development it is important to outline the purpose and methods of these analyses as well as to specify properties of a SIC table and its subsets.

Three basic concepts: species, instar, and character

André (1986, 1988) discussed the concept of stase and proposed to extend it to other arthropods, particularly Collembola. According to André (1989a), each intermolt of a mite or juvenile Collembola is a stase, and the adult of Collembola, which molts but does not change by discontinuous characters, also corresponds to one stase. Although the concepts of stase and instar are different, I will use the word "instar" instead of "stase", because it has the advantage of being better known among entomologists. My argument is based on commonness of a term, but conceptually André's argument remains worth while.

The term "species" as used in a SIC table refers to a "terminal taxon". In fact, such a table can include any supraspecific taxon, with one condition: character states are constant within every instar of taxa. In other words, attributes should not vary among "elements" taken in one terminal taxon and in one instar. "Elements" are defined as individuals when the terminal taxon is a species, as a species when the terminal taxon is a genus, and so forth. This condition should also be applied at the specific level, e.g., if a seta is present in adults in one population and absent in another population, the seta should not be included in the set of characters, or, as an alternative, the state of character for the species should be treated as unknown. Such a consideration amounts exactly to a principle stated by Nixon & Wheeler (1990a: 218): "Strictly speaking, terminal lineages at any level of cladistic analysis should not vary internally for the characters used in the analysis".

The term "character" as used here is similar to variables or vectors in data analysis terminology, and "species" and "instars" correspond to objects. Consider a set of specimens collected in one or several stations; they may be clustered according to several criteria such as species, ontogenetic levels, sex, and even castes for social insects. The species category is the crux of the species-instar-sex spectrum, because instar and sex categories are merely intraspecific distinctions. In order to simplify the situation, sex will be left out to avoid having four dimensions in the SIC table.

Nixon & Wheeler (1990a, 1990b) made a distinction between attributes that vary and attributes that are constantly distributed among specimens of an elementary division of the species-instar-sex spectrum. The first type was named a trait, the second a character, or more precisely, a character state. Three categories were established: 1) the trait as a variable attribute fitting with variably distributed alleles, 2) the character state as constant attribute permitting species diagnoses, and 3) the character as a set of character states derived from one another through a series of transformations. Nevertheless, many authors (e.g., Platnick, 1979; Eldredge & Cracraft, 1980; Wiley, 1981; Patterson, 1988) dispute the distinction between characters and character states, arguing that, since the essence of systematics is hierarchy, characters and character states are just a series of nested increasingly modified attributes, so that a character state can become a character at a lower taxonomic level. According to another viewpoint, it can be accurate to distinguish between characters and character states. De Pinna rightly points out that this problem is "related to recognition of putative independence among sources of evidence. [...] Character states are attributes that can be proposed as transformations one of another (i.e. as a series of transformations); characters, on the other hand, are putatively independent from one another. If the distinction between character and character state is not made, then theoretically any individual attribute (i.e. any character state) could be transformed into any other, through any number of intermediate steps" (De Pinna, 1991: 380). The difference between character states and traits relies on the distinction between hierarchic patterns as phylogeny, and reticulate patterns as birth relationship (called tokogeny by Hennig, 1966). Nixon & Wheeler (1990b: 122) argued "that tokogenetic systems do not meet the assumptions of cladistic analysis and cannot be considered to be fully hierarchic and are therefore inappropriate for cladistic investigations". The concept of trait was then extended to all "attributes that are variably

distributed within any grouping (population, clade or terminal lineage) that is phylogenetically unresolved internally" (Nixon & Wheeler, 1990a: 218).

Probabilistic organs

Definition of the concept

I deem that a variable seta corresponds to a probabilistic organ. The concept of probability of presence for an organ needs comment because this issue is closely connected with the problem of the phylogenetic informativeness of variable organs which is one of the central claims of this paper (see also Nayrolles, 1995a). The problem we must then solve boils down to this question: do variable setae correspond to simple traits? If the answer is positive, and intuitively it seems positive, the implication would be that such setae do not meet the definition of characters, and consequently should not be used to infer phylogenetic relationships. The last point that will have to be dealt with relates to a practical problem, the confidence one may grant to the use of probability.

A paired seta may be variable in a population of symmetrical specimens, some having the seta on both sides, others wholly devoid of it. Likewise, a paired seta may be variable in asymmetrical specimens; in this case, one should study the frequency on both the left and right sides. Grandjean (1939) in Acari, and myself in Collembola Symphypleona, have remarked that when a paired seta is variable, many asymmetrical specimens generally occur. Given a variable paired seta, let P be the modality "presence of the seta" and A the modality "absence of the seta". Grandjean (ibid.) showed that frequencies on the left and right side are similar (statistically not-significantly different) and frequencies of the different types of specimens (namely PP, PA, and AA) approximate to a binomial law. My observations corroborate this fact.

Grandjean (1948, 1971) reared a thelytokous parthenogenetic mite. A female gave birth to other females that did not display exactly the same chaetotaxy as their mother or each other. There is no genetic polymorphism in this case, and thus all specimens should be alike. Considering that a variable paired seta occurs on the left or right side at random and such a seta is not *per se* hereditary, Grandjean concluded that the variability is due to a propensity intrinsic to each individual, that is, the inherited feature lies in the presence probability at a stage of ontogeny. In this case, the "probability of presence" is a genetic character, even if I cannot explain this phenomenon. What is partly resolved is the cell lineage fate.

Several authors (e.g., Doe & Goodman, 1985; Campos-Ortega, 1988; Held, 1990a; Simpson, 1990; Heitzler & Simpson, 1991) showed that the segregation of neural and epidermal lineages in *Drosophila melanogaster* relies on cellular interactions that are embodied in the concept of "lateral inhibition". A cybernetic model was put forward in which chance played a part. Heitzler & Simpson (1991: 1089) suggested that: "a feedback mechanism whereby cells producing less receptor relative to their neighbors will be more efficient at signaling. In normal development, therefore, if one cell produced slightly less *N* product [protein involved in the lateral inhibition], perhaps through random physiological fluctuations, it would gain an early advantage, generate a greater signal, and inhibit adjacent cells."

A related problem is the geometrical pattern in which setae are often arranged. In his survey of basitarsal setae in *D. melanogaster*, Held (1990a: 61) stated that "the development of the basitarsus appears to be governed by a hybrid mechanism involving *both* a global coordinate system and local cell interactions". The coordinate system controls the arrangement of setae in rows and/or in whorls, and is likely to govern the positions of idionymic setae (i.e., setae with fixed positions). The basitarsus of *D. melanogaster* bears three types of sensory structures: bracted bristles, bractless bristles, and sensilla campaniforma. The bracted bristles are aligned in rows, evenly spaced within the rows, and their positions are not constant but a function of their number. Conversely, bractless bristles and sensilla campaniforma form

two subpatterns in which the elements are few, aperiodic, and idionymic. Using several perturbing factors (e.g., gamma irradiation), Held (1990b) showed that the arrangement of bracted bristles does not develop in a linear direction, and concluded that "the failure to find development waves of sensibility implies that neither bristle nor row positions are established by directional patterning mechanism." Regular spacing of non-idionymic setae was thus interpreted by Held as a pattern which does not develop in an iterative manner along its major axes, but is probably generated by short-range cell signaling and cell rearrangement. For idionymic setae, it may be assumed that setal variability relates to cell interactions and some underlying random fluctuations.

Probabilistic organs are not a novelty. In 1954, Stern showed that the arrangement of thoracic macrochaetae in *D. melanogaster* follows a predetermined pattern in which absence of a macrochaeta's precursor can be replaced by a neighboring cell that switches its fate to become the new precursor. Comparisons between normal and mutant *achaete* flies show that the normal gene leads to differentiation of 11 macrochaetae at specific places, whereas the *achaete* gene does not regularly cause such differentiation: zero to three macrochaetae are lacking in mutated flies. Therefore, in a pure lineage, all specimens are not identical.

The effect of genes on the presence or number of setae has been studied in detail by Held (1990a) on the second-leg basitarsus of *D. melanogaster*. Several mutations entail a decrease in the number of bractless bristles whereas other mutations cause extra bractless bristles to appear. If we compare the effects of the missing-bristle mutations, it appears that all bractless bristles do not have the same "sensibility", e.g., a seta variable in the wild genotype becomes absent, setae constantly present in the wild genotype become variable. Results for the sensilla campaniforma are similar. Therefore, it seems that each seta has an intrinsic "propensity" to develop. Grandjean (1941) asserted a similar view when he claimed that a hierarchy can be developed for a series of probabilistic organs insofar as some elements are "sturdier" than others.

Lateral inhibition has been found in organisms other than insects, e.g., in the alga *Anabaena* (Wilcox et al., 1973). The "microphtalmie aléatoire" (random microphthalmy mutation) described by Signoret & Lefresne (1969) corresponds to a similar phenomenon. This mutation entails anomaly for blood irrigation of the cephalic area during development of the amphibian *Ambystoma mexicanum*. The ophthalmic artery is regressed and the eye undergoes limited growth. The important point is that all mutants do not share the same morphology. Specimens with 0, 1, or 2 microphthalmic eyes occur. As in Grandjean's observations, the probability of microphthalmy for the right eye and left eye follows a binomial law. The authors concluded that the mutation corresponds to a gene with a variably expressed recessive allele. Obviously, this case is exceptional, and it would be incorrect to generalize so far as to see every organ as probabilistic. This case simply shows that, at one moment of development, the conditions to realize such a complex structure as the vertebrate eye can be unstable, so that a minute fluctuation switches the development toward one or another direction. Complex regulating systems prevent this instability for vital organs, and specimens in which regulations would not be firmly established would have a very little chance of survival, but this is not true for repeated small organs such as the setae of arthropods.

From theory to practice

If we now turn to the commonness of probabilistic organs within a series of specimens, Grandjean made several accurate observations. He called "écart" (deviation) the case of one seta being absent in a specimen when this seta is normally present in the considered species and instar, or reciprocally, presence of a seta when the normal condition is absence (Grandjean, 1939). The setae concerned by deviations are almost always the secondary ones. Studying a population of the mite *Platynothrus peltifer*, Grandjean (1948) not-

ed that only 2 out of 26 larvae (first mobile form in Oribatida) exhibited a deviation, while all tritonymphs (form preceding the adult) and all adults had at least one deviation. I have observed a similar situation in Collembola Symphypleona: for the last stages of development, it is rare to find a specimen whose chaetotaxy exactly matches the standard of the species. Specimens with one or several deviations are the rule, but that does not imply that many setae are variable. For instance, Grandjean (1948) counted 655 deviations for a global number of 49870 setae examined (122 specimens observed), and more than a half of these deviations originated from a few setae. In practice, only setae for which more than one deviation is observed will be noted as variable, otherwise chaetotaxic descriptions would be cluttered with many exceptions (Nayrolles, 1993a).

Since a variable seta may be viewed as a probabilistic organ, one can calculate a setal probability within a population and consider this number as a character state. Thus, I claim that variable setae do not correspond to simple traits. I concede that my argument is flawed by an underlying assumption, that is the propensity for bearing a certain seta perhaps varies among individuals, so that the real propensity, i.e., that of each specimen, is beyond the reach of our observations, apart from breeding pure lines. In this context, comparisons between populations can be useful (an example was commented on by Van der Hammen, 1981: 11).

Another approach is to consider that frequencies distribution of the specimens with the seta on both sides (PP), on one side (PA), and without the seta (AA) follows a binomial law (P is the modality "presence of the seta" and A the modality "absence of the seta"). Let us consider a population in which the probability calculated for a certain seta is 0.5, and let us suppose that the population is composed of individuals of two types: some with the seta, others without. We observe half of the specimens with the seta on both sides, and half devoid of the seta. The frequencies distribution does not follow a binomial law. Thus, when the frequencies distribution is different from a binomial law, we can say that the calculated probability results from different propensities. When the frequencies distribution fits with a binomial law, there is a real chance that the calculated probability corresponds to a unique propensity.

A final argument is decisive for using setal probabilities as character states. I showed (Nayrolles, 1993a) that presence vs. absence of certain setae are statistically correlated. Several setae of the fourth antennal segment are variable in Collembola Symphypleona. These setae have a special shape and are arranged along two rows called intergeneratrices, because each is situated between two rows of ordinary and constant setae called generatrices. I studied one species, and made a histogram of the variable "number of setae on the intergeneratrices" and compared it with the histogram as it would have been if the setae were independent. The observed histogram is much narrower than the calculated one. Therefore, in this example the set of correlated setae amounts to only one character, and the "probabilistic nature" of the "synthetic character" is far less pronounced than that of original characters (i.e., setae considered one by one). Similar stochastic relations between organs were observed in mites (Matsakis, 1967).

The stochastic development of individual items is thus counterbalanced by a deterministic process which severely restricts the range of variation for the number of items. Hence, it is accurate to define chaetotaxic variables as setal numbers (normal practice among entomologists) or even ratios of setal numbers. The SIC table analysis chiefly aims to cluster setae belonging to the same phylogenetic set, i.e., those which relate to one character.

In practice, chaetotaxic studies on arthropods take a long time, e.g., in Collembola Symphypleona I used to mount some specimens of every instar for observation, with dissection of legs and furcula. The mounting and observation are generally very time-consuming, so that only a few specimens are examined (about ten specimens for an instar) and comparisons of populations cannot be carried out as a matter of course in species descriptions.

An argument against the use of setal probability is that, if numerous specimens are not observed, the confidence interval of probability is too wide and consequently, the values obtained are nothing more than sampling errors. Nevertheless, we can use results obtained from few observations if the values are very different, because the important thing is the comparison and not the values by themselves. For example, consider a seta observed with 4 presences and 16 absences in one species, and 18 presences and 2 absences in another species; these values are statistically speaking very different. Although one may be reluctant to use such data in phylogenetics, a more convincing argument is that variable setae are generally not considered separately. Indeed, variable setae may be clustered in some sets and if we assume that in each set setae have statistical relations (i.e. compensation) we can guess that the setal number of every set (which amounts to the sum of setal probabilities) is a real measure of the "pilosity" of a body area. The variation of this value, if not directly measured, can be assumed to be quite narrow because of the compensation phenomenon (one may go back to data to estimate the variation).

Definition and properties of a SIC table[1]

In the study of a zoological group, one can distinguish characters which are not constant between species taken at the same level of ontogeny. For each character, the observed differences are assumed to be a consequence of phylogeny. For this reason, such characters will be designated as phylogenetically variable.

By definition, a SIC table will have no character constant both in ontogeny and phylogeny. It is defined by the following three sets:

– set C of characters not constant in ontophylogeny. $C = \{c_1, c_2, \ldots c_i, \ldots c_x\}$

– set T of taxa. $T = \{t_1, t_2, \ldots t_j, \ldots t_y\}$. Generally each taxon is a species (hence the expression species/instars/characters table).

– set O of ontogenetic stages (i.e. instars). $O = \{o_1, o_2, \ldots o_k, \ldots o_z\}$. Ontogenetic stages should be the same for all taxa.

Fig. 1 shows a SIC table. We will deal with chaetotaxic characters, and a three-dimensional cell of the SIC table will reveal the presence probability of one seta in one particular instar of one particular taxon.

Three sections can be defined within a SIC table. The first section cuts off characters, and for each character it makes a bidimensional table described by $T \times O$ (or also $O \times T$ since taxa can be equally presented in rows or columns). With such a section we get species/instars cards (one card per character). I use the word "card" because it refers to data processing, and the tridimensional table can be considered as a database. The second section cuts off taxa, and for each taxon we get a bidimensional table described by $C \times O$ (or also $O \times C$). Finally, the third section cuts off instars, and for each instar we get a bidimensional table described by $C \times T$ (or also $T \times C$).

We can divide the set of taxa, T, into several subsets T_1, T_2, T_3, etc. For example, subsets can correspond to genera. Within C, two subsets are distinguished: the set of the ontogenetically constant characters (i.e., constant during development of each species), written C_o, and the set of the phylogenetically constant characters, written C_p. Note that $C_o \cap C_p$ is the set of characters constant in ontophylogeny, and according to the definition of C: $C_o \cap C_p = \varnothing$. Let C_n be the set of characters variable both in ontogeny and phylogeny, thus: $C_o \cup C_p \cup C_n = C$.

---

[1] The following mathematical symbols of set theory are used:

{ } symbol of a set. Example: $\{c_1, c_2, c_3\}$ is a set of three elements: c_1, c_2, c_3.

× symbol of the cartesian product of two sets. Example: the cartesian product $\{c_1, c_2, c_3\} \times \{o_1, o_2\}$ is the set of ordered pairs: $\{(c_1, o_1), (c_1, o_2), (c_2, o_1), (c_2, o_2), (c_3, o_1), (c_3, o_2)\}$

$\varnothing$ symbol of the empty set.

$\cap$ symbol of intersection. Example: $\{c_1, c_2, c_3\} \cap \{c_1, c_4\} = \{c_1\}$

$\cup$ symbol of union. Example: $\{c_1, c_2, c_3\} \cup \{c_1, c_4\} = \{c_1, c_2, c_3, c_4\}$

$\subset$ symbol of inclusion, means that a set is a subset of another set. Example: $\{c_1\} \subset \{c_1, c_2, c_3\}$

$\forall$ symbol of the universal quantificator, e.g. $\forall T_\alpha \subset T$, is read: "for any set T_α included in the set T"

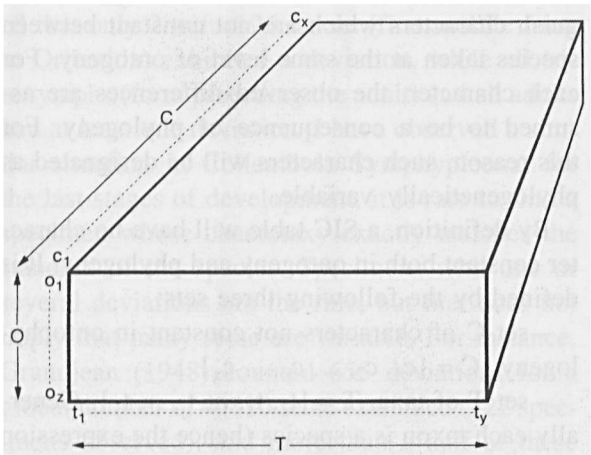


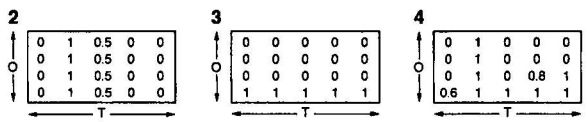
Fig. 1. Graphical representation of a SIC table. C is the set of characters (x characters), T the set of taxa (y taxa), and O the set of instars (z instars).

Figs. 2-4 show species/instars cards for a character belonging to C_o, or C_p, or C_n. These cards contain numbers which are setal presence probabilities.

In phylogenetic studies, the relevant characters are those variable between taxa. By definition, C_q is the set of the phylogenetically variable characters. Thus, it is the complementary set of C_p within C. The characters of C_o are obviously variable in phylogeny, since C does not have any ontophylogenetically constant character; hence: $C_q = C_o \cup C_n$.

Let us consider any subset of T, and let us name it T_α. Defined on T_α, let $C_{o\alpha}$ be the set of the ontogenetically constant characters, $C_{p\alpha}$ the set of the phylogenetically constant characters, $C_{n\alpha}$ the set of the phylogenetically as well as ontogenetically variable characters, and $C_{q\alpha}$ the set of the phylogenetically variable characters. Whether a large table is cut into subsets of taxa, for each subset T_α, the characters which have to be kept are those belonging to $C_{q\alpha}$. Only these characters are useful for differentiating species within T_α.

A character phylogenetically constant in a subset of taxa is not necessarily phylogenetically constant for all taxa. So, the relation $C_{p\alpha} \subset C_p$ is not always true. We can thus establish logical relations as those stated below. The study of such relations will be a prerequisite to develop database application.



Figs. 2-4. Three examples of species/instars cards (5 species, 4 instars). Each card corresponds to a character with one column per species, instars being written in rows and getting older from the top to the bottom; the numbers are setal presence probabilities.

Fig. 2. Species/instars card of a character belonging to the set of ontogenetically constant characters, C_o.

Fig. 3. Species/instars card of a character belonging to the set of phylogenetically constant characters, C_p.

Fig. 4. Species/instars card of a character belonging to the set of characters variable both in ontogeny and phylogeny, C_n.

$\forall\ T_\alpha \subset T,$	$C_{p\alpha} \subset C_p$	false
$\forall\ T_\alpha \subset T,$	$C_{o\alpha} \subset C_o$	false
$\forall\ T_\alpha \subset T,$	$C_{n\alpha} \subset C_n$	true

A good method for comparing subsets T_α, T_β, etc., is to retain their phylogenetically constant characters ($C_{p\alpha}$, $C_{p\beta}$, etc.), and then to compare them. Thereby, we may define the set of characters which differentiate the subsets of taxa. We write it: $C_{d(\alpha, \beta, \ldots \omega)}$. By definition: $\forall\ T_\alpha \subset T$, $\forall\ T_\beta \subset T$, ... $\forall\ T_\omega \subset T$, with T_α, T_β, ... T_ω separated subsets, $C_{d(\alpha, \beta, \ldots \omega)}$ is the set of the characters which are phylogenetically constant within each subset T_α, T_β, ... T_ω and different between at least two of these subsets.

The Appendix provides a concrete example of the use of the sets C_p, C_o, C_q, etc.

The spreading projection

In order to assess distances among species, or instars, or characters, we could use the "three mode-principal component analysis" applied to tridimensional tables by Kroonenberg (1983). In fact, such a method is not necessary. Indeed, we can study the phenotype not of species considered each as a whole, but of species divided into instars. That amounts to opening the tridimensional SIC table and getting a bidimensional table with characters in columns and instars of species in rows (of course it is equivalent to write characters in rows or columns). I name this process a "spreading projection". André (1988, 1989b) proposed to

project into a reduced space the species divided into instars, and then to connect the points representing the successive instars of each species; he coined the phrase "ontogenetic trajectory" for such a pathway. By this means, development of species can be compared.

The bidimensional table achieved by a spreading projection is called instars-species/characters table (e.g. Tables I, II). If x, y, and z are respectively the cardinal numbers of C, T, and O, then the dimensions of the obtained table are x and yz. In fact, only characters phylogenetically variable are of interest, so that a spreading projection performed from the part of the SIC table defined on C_q is sufficient. As far as the distinction among the instars of species is concerned, the bidimensional table must provide the same information content as the original data: we say that the instars-species/characters table must fit the "pertinence principle". For example, if, for every taxon, all the instars have been distinguished among them, that must still be true in the instars-species/characters table. If it is not the case, we add to C_q the smallest set of characters that restores pertinence (Table I).

Some setae have a homeotypic relationship, e.g., a seta in the same place on the three pairs of legs (see Fig. 5). When setae are arranged in a geometrical pattern, homeotypic relationships can be defined with regard to chaetotaxic arrangement. For instance, setae on appendages of Symphypleona are arranged in a "longitudinal structure" composed of eight rows, called generatrices (Nayrolles, 1992). Furthermore, a "transversal structure" is often observed, e.g., whorls of tibiotarsi. Setae of the same whorl or of the same generatrix are liable to have homeotypic relationships (Fig. 5). Moreover, the presence probabilities of these setae can be similar; in this case, we can combine characters in a single one. The new character corresponds to the sum of the presence probabilities. It is preferable to use sum rather than average because that does not alter initial data nor distances between instars of species.

As a rule, homeotypic setae are clustered according to two criteria: 1) their presence probabilities are similar (similarity criterion), and 2) their presence probabilities can be ordered (ordinal criterion). An example is given in Tables II and III, and a table obtained from real data is given in the Appendix (Table XI).

Let u and v be two characters, each corresponding to the presence probability of one seta. Under the first criterion u is similar to v, that we write $u \approx v$, and according to the ordinal criterion either u is less than v ($u < v$) or u is greater than v ($u > v$).

Table I. Determination of the pertinent set of characters. Characters (c_1 to c_8) in columns, and species (t_1, t_2, t_3) divided into their instars (o_1 to o_4) in rows. All instars are distinguished by the characters of C_q, except for o_1 and o_2 of t_2. Character c_6 of C_p permits this distinction: c_6 restores pertinence. Hence, the pertinent set of characters (showed by a double arrow) is $C_q \cup \{c_6\}$.

		C_q					C_p		
		c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
t_1	o_1	0	0	0	0	1	0	0	0
	o_2	1	0	0	0	1	1	0	0
	o_3	1	1	0	1	1	1	1	0
	o_4	1	1	1	1	1	1	1	1
t_2	o_1	0	1	0	0	0	0	0	0
	o_2	0	1	0	0	0	1	0	0
	o_3	0	1	0	1	0	1	1	0
	o_4	0	1	1	1	0	1	1	1
t_3	o_1	0	0	0	0	1	0	0	0
	o_2	1	0	0	0	1	1	0	0
	o_3	1	0	1	0	1	1	1	0
	o_4	1	0	1	1	1	1	1	1

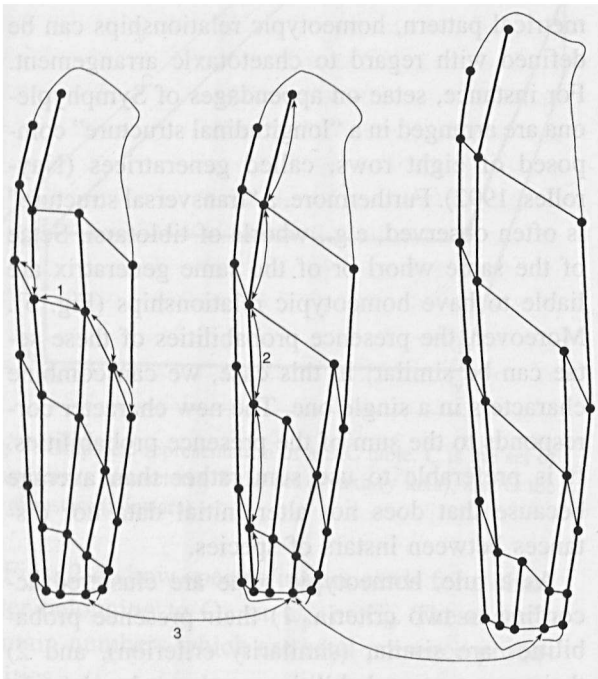


Fig. 5. Homeotypic relationships between setae. From the left to the right are drawn the posterior side of fore, mid, and hind tibiotarsus of the first instar of *Bourletiella hortensis* (Fitch, 1863), sockets of setae being represented by dots. Setae are arranged in generatrices (thick lines) and whorls (thin lines). Certain whorls are incomplete (e.g., posterior generatrix with two setae absent on midleg and three on hindleg). Three homeotypic relationships (represented by double arrows) can be considered: 1) between setae of a same whorl, 2) between setae of a same generatrix, and 3) between setae situated in the same place on the three tibiotarsi.

Similarity is a fuzzy relation that can be mathematically defined with the help of the fuzzy sets theory (to discuss this point is beyond the scope of the present paper, see Zadeh, 1965, for an introduction to the principles).

Characters u and v belong to the instars-species/characters table in which characters are written in columns and instars of species in rows. There are n rows. We write u_i and v_i the i-th observations of u and v, i varying from 1 to n. The similarity or ordinal relation between u and v can be stated as:

$$u \approx v \Leftrightarrow \begin{cases} \forall\, i \in [1, n],\ u_i \approx v_i \\ \sum_{i=0}^{n} u_i \approx \sum_{i=0}^{n} v_i \end{cases}$$

$$u < v \Leftrightarrow \begin{cases} \forall\, i \in [1, n],\ u_i \leq v_i \\ \sum_{i=0}^{n} u_i < \sum_{i=0}^{n} v_i \end{cases}$$

$$u > v \Leftrightarrow \begin{cases} \forall\, i \in [1, n],\ u_i \geq v_i \\ \sum_{i=0}^{n} u_i > \sum_{i=0}^{n} v_i \end{cases}$$

With regard to the ordinal criterion, it would be relevant to look for anterior-posterior gradients of setal presence probabilities on the legs (by com-

Table II. Instars-species/characters table. Characters in columns, species divided in their instars in rows.

		c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
t_1	o_1	0	0	1	0	0	0	0	0	1
	o_2	0	1	1	0	0	0.4	0	0	1
	o_3	1	1	1	1	0.8	1	1	1	1
	o_4	1	1	1	1	0.8	1	1	1	1
t_2	o_1	0	0	0.5	0	0	0	0	0	0
	o_2	0	1	0.5	0	0	0.8	0	0	0
	o_3	1	1	0.5	0.8	1	0.8	0	1	1
	o_4	1	1	0.5	1	1	0.8	0	1	1
t_3	o_1	0	0	0	0	0	0	0	0	0
	o_2	0	0	0	0	0	0	0	0	0
	o_3	0.2	1	0	0	0	0	0	0	0
	o_4	1	1	0	0.2	0	0	1	1	0.5
t_4	o_1	0	0	1	0	0	0	0	0	0
	o_2	0	0	1	0	0	0	0	0	0
	o_3	0	1	1	0	0	0	0	0	1
	o_4	1	1	1	0	0	0	0	0	1

parisons between fore, mid and hindlegs); such gradients could also be looked for along the rows of setae on appendages.

Clustering setae is justified for two reasons. First, from a statistical viewpoint, it takes account of the initial table and does not greatly alter χ^2 distances between instars of species. This latter point is of interest, since the best distance in the method of ontogenetic trajectories is the χ^2 distance (Nayrolles, 1996). Second, from a biological viewpoint, it is not surprising that several setae with homeotypic relationships have presence probabilities either similar or in ordinal relation. Finally, these setae convey the same information. Thus, the setal clustering reduces the redundancy of information.

The idea of setal clustering lies also in the concept of relationship between setae. I stated above that for certain setae on the fourth antennal segment in Symphypleona, the observed number of setae displays a much narrower range of variation than the number calculated from a model of statistical independence between setae. I called this phenomenon "compensation" (Nayrolles, 1993a, and also Matsakis, 1967) because absences counterbalance presences. Indeed, the successive variable setae in a row (intergeneratrix) along the antennal segment display negative statistical relations for presence vs. absence. That means that when a seta is present on the whorl n, the probabilities of setae on the whorls n - 1 and n + 1 are less than if the seta of whorl n was absent. Setae involved in compensation are thus clustered in one set described by a variable "number of setae".

To cluster homeotypic setae on the basis of their relationship or on the similarity or ordinal criteria relates to the same principle. Setal clustering does not only decrease the information of data but it also improves this information because of the reduction of redundancy and background noise (i.e., the individual variability of each seta in a setal set with compensation). For instance, I used the setal clustering to enhance distinctions between species (Nayrolles, 1995b).

The alphabetic and numeric projections

General points

I stated above that the setal presence probability has phyletic information content. Therefore, the state of a variable character is not to be consid-

Table III. Cluster of several characters of Table II: $c_{1,2}$ corresponds to the combination of c_1 and c_2, $c_{4,5}$ corresponds to the combination of c_4 and c_5, and $c_{7,8}$ corresponds to the combination of c_7 and c_8. The relationships of initial characters are based on homeotypy and the following profile comparisons: $c_1 < c_2$; $c_4 \approx c_5$; $c_7 < c_8$.

		$c_{1,2}$	c_3	$c_{4,5}$	c_6	$c_{7,8}$	c_9
t_1	o_1	0	1	0	0	0	1
	o_2	1	1	0	0.4	0	1
	o_3	2	1	1.8	1	2	1
	o_4	2	1	1.8	1	2	1
t_2	o_1	0	0.5	0	0	0	0
	o_2	1	0.5	0	0.8	0	0
	o_3	2	0.5	1.8	0.8	1	1
	o_4	2	0.5	2	0.8	1	1
t_3	o_1	0	0	0	0	0	0
	o_2	0	0	0	0	0	0
	o_3	1.2	0	0	0	0	0
	o_4	2	0	0.2	0	2	0.5
t_4	o_1	0	1	0	0	0	0
	o_2	0	1	0	0	0	0
	o_3	1	1	0	0	0	1
	o_4	2	1	0	0	0	1

ered as polymorphic, instead it corresponds to a step in a transformation series. For example, if the plesiomorphic state is "presence of a certain seta which appears in adults" and absence the apomorphic state, then species in which the seta in question is variable in adults have an intermediate evolutionary state for this character. In this case, values of presence probabilities are gathered in few ranges, each range fitting with a state, e.g., probability = 1 as plesiomorphic state, [0.4 - 0.6] as intermediate, and 0 as apomorphic.

Analyses of tridimensional tables aim to study kinship of species. For that, it is necessary to build up a table with two dimensions: one for characters, the other for taxa. If we try to give a graphical representation of the process in question, this looks like a projection on the plane defined by the characters and taxa. Nevertheless, such a table should take into account the ontogenetic dimension when species do not get secondary setae at the same instar. Into this type of projection, dimensions of the table obtained are x and y, the cardinal numbers of C and T. Table squares will hold either letters (alphabetic projection) or numbers (numerical projection).

I will support my explanations with a series of tables (IV to VII). These tables fit a didactic purpose, they are not real tables (real tables are much larger). An example of treatment for real data is given in the Appendix.

The alphabetic projection

If we consider the presence vs. absence of any seta, two basic ontogenetic changes can be observed, appearance vs. disappearance. So, during ontogeny, we can consider the following changes:

- (1) a single change that is the appearance of a secondary seta;
- (2) a single change that is the disappearance of a primary seta;
- (3) a combination of both single changes, in this way: appearance of a secondary seta and then its disappearance;
- (4) a combination of both single changes, in this way: disappearance of a primary seta and then its reappearance.

Case (1) is very frequent, while (2), (3) and (4) are very rare in Collembola. For example, (3) is unknown, (2) and (4) are known for only one seta (trichobothrium D of Symphypleona, see Betsch & Waller, 1994). These features are different in Acari: (2) is not rare and presence of calyptostases poses several problems. Here, we will limit the discussion to Collembola, and since cases (2), (3), and (4) are very unusual we will only consider case (1), i.e., ontogenetic change corresponding to setal appearance. In other words, setal presence probability does not decrease during ontogeny.

We give the following rules: character "—" is used for the lack of a seta; when a seta is variable at the instar in which it appears, this instar is written in parentheses, and if in a later instar it becomes constant, this instar is given as well. For example, (o_3) means that a seta appears with variability at the third instar and remains variable, while $(o_3)o_4$ that a seta appears with variability at the third instar and becomes constant at the fourth. Table IV provides an example.

The alphabetic projection yields a bidimensional species/characters table. In such a table, it is unnecessary to consider the phylogenetically constant setae, and it is thus relevant to carry out an alphabetic projection from the part of the SIC table limited by C_q. In practice, an alphabetic table is constructed in the first step, because it permits us to identify the phylogenetically constant characters and to ignore them in other analyses (see Appendix).

The numeric projection

The aim of alphabetic projection was to get a standard presentation of the chaetotaxic ontogeny and to provide a cursory knowledge of data. The purpose of a numeric projection is to achieve a cladistic matrix.

Values produced by this data processing should be regarded as real measures. For each final character, these measures provide a comparison of the "intensity" with which a set of homeotypic setae is present throughout one or several stages of development. An ultimate data processing, which

Table IV. Alphabetic species/characters table corresponding to Table II. See text for further explanations.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
t_1	o_3	o_2	o_1	o_3	(o_3)	$(o_2)o_3$	o_3	o_3	o_1
t_2	o_3	o_2	(o_1)	$(o_3)o_4$	o_3	(o_2)	—	o_3	o_3
t_3	$(o_3)o_4$	o_3	—	(o_4)	—	—	o_4	o_4	(o_4)
t_4	o_4	o_3	o_1	—	—	—	—	—	o_3

Table V. Numeric species/characters table corresponding to Table II. See text for further explanations.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
selected instars	o_3	o_2	o_1	o_3+o_4	o_3+o_4	$o_2+o_3+o_4$	o_3+o_4	o_3+o_4	$o_1+o_2+o_3+o_4$
t_1	1	1	1	2	1.6	2.4	2	2	4
t_2	1	1	0.5	1.8	2	2.4	0	2	2
t_3	0.2	0	0	0.2	0	0	1	1	0.5
t_4	0	0	1	0	0	0	0	0	2

will not be dealt with here, corresponds to delimitation of character states and polarizing and coding them before to achieve a cladistic analysis.

In the first step, for each character, we note instars concerned by the variation among species, then we sum the presence probabilities of the selected instars. Table V, calculated from Table II, corresponds to this step. For instance, consider character c_4, two instars are selected, o_3 and o_4, the value of c_4 for the taxon t_2 (1.8 in Table V) is equal to the setal presence probability of c_4 at the instar o_3 of t_2 (0.8 in Table II) plus the setal presence probability of c_4 at the instar o_4 of t_2 (1 in Table II). In case of ontogenetically constant characters (included in C_o), we only keep the first instar, o_1 (e.g., character c_3 in Tables II and V). The table obtained has to mention the selected instars of each character.

This first step can be regarded as quite objective, since it makes no assumption on the polarization of characters. Nevertheless, adding up presence probabilities of one seta observed in several instars implies that this sum holds as much information as the original data. In other words, that amounts to consider c_6 as having the same information content for t_1 and t_2 in Table V, in spite of the fact that t_1 and t_2 in Table II, each divided into instars, have different values for c_6.

I have found this situation very rarely in my studies on Collembola Symphypleona. Apart from this exceptional case, which may deserve special treatment, the table obtained can be considered to be pertinent beside the original SIC table. Obviously, the process of numeric projection would be more involved if there were other types of ontogenetic changes than the setal appearance.

Characters will be combined in order to decrease their number and redundant information. We will combine setae with homeotypic relationship according to the clustering criteria and pertinence principle.

Characters to be clustered have the same evolutionary direction. One could think that this assertion entails a superfluous assumption. How would one be able to assess *a priori* that several characters have the same evolutionary direction? If characters present very similar values, evolutionary states of these characters (e.g., seta present at the fourth instar vs. seta present from the third instar on) will have the same polarity after a cladistic analysis. So, characters to be clustered are correlated and *de facto* have the same evolution. That does not mean that any of the correlated characters can be clustered. Indeed, setae must share homeotypic relationships to be clustered. Entomologists do not cluster clypeal setae with tarsal setae because such setae are on distinct areas and have no homeotypic relationship. In cases where

setae are idionymic, a setal number corresponds to a combination of elementary characters which are presence vs. absence of each seta. This case is quite frequent in literature, but authors do not always realize the underlying assumption involved in such characters (Nayrolles, 1996).

Every new character (combined character) must have its instars' range which encompasses all selected instars of its initial component characters. From a practical viewpoint, only characters showing the same instars' range can be directly clustered. Nevertheless, we may add one or several instars to characters for matching up their instars' ranges. Each time one adds an instar to the instars' range, it is necessary to add also its setal presence probability to the values of the character. For example, in Table V, if we want to gather c_1 and c_2 we have to add o_2 to c_1, and o_3 to c_2. In this process, the values of c_1 do not change (because for c_1 the setal presence probability of the added instar, o_2, is equal to 0), on the other hand we must add 1 to the values of c_2 (because for c_2 the setal presence probability of the added instar, o_3, is equal to 1). Table VI shows the preparation for the size reduction of Table V. Table VII corresponds to the reduced table.

The similarity criterion can be used in the same way as in the spreading projection. On the other hand, to apply the ordinal criterion may raise a problem. For instance, in Table II, we have: $c_8 > c_7$, and, after the numeric projection (Table V), if we add c_7 and c_8, we get the same value (2) for t_2 and t_3. It is problematic because this value is built up from two different ways, and generally that is due to two distinct evolutionary pathways. Indeed, suppose that the character polarity of c_7 and c_8 is so that the primitive instar of setal appearance is the third one, the intermediate evolutionary state corresponds to the setal appearance at the fourth instar (ontogenetic delay), and the fully apomorphic state is the setal lacking, then taxa t_2 and t_3 are really separated: t_2 is evolved for c_7 and primitive for c_8, whereas t_3 has an intermediate evolutionary state for both characters. Thus, in this example, we must hold c_7 and c_8 separated. We have defined the possibility of characters clustering in Table II, and we have performed this process from Table VI to Table VII. However, this process can be made directly from Table II (but for very large tables, the intermediate step of Tables V-VI is very useful).

Conclusion

A SIC table based on setal presence vs. absence provides a good tool for ontophylogenetic studies on arthropods. The characters of this tridimensional table correspond to setal presence probabilities. In this paper, I define the main properties of such a table, particularly the problem of its division into taxonomic groups.

Two methods can be used for analyzing a SIC table. The first aims to compare instars of species. The tridimensional table is "opened out" (spreading projection), with, for example, the characters in columns and the instars of species in rows. Then we gather characters with the same information content, this information taking into account statistical as well as biological factors (homeotypy). By this means, we get the table with the fewest informative characters and keeping at best differences between instars of species of the initial table (pertinence principle). It is then possible to apply statistical methods such as Factor analysis. By joining points representing the successive instars into a reduced space, we can draw ontogenetic trajectories (André, 1988, 1989b; Nayrolles, 1996).

A second process seems to erase the ontogenetic dimension, but in fact, the ontogeny, when it differs among taxa, is not forgotten. We construct a table with two dimensions, e.g., characters in columns and taxa in rows. This process is an alphabetic projection when the achieved table is filled with letters symbolizing instars of setal appearance and a numeric projection when the table is filled with numbers. In this last case, for each character, we add the setal presence probabilities of the selected instars. The obtained table can be used to achieve a cladistic analysis.

Klompen & O'Connor (1989) showed that the use of ontogenetic transformations yields better results for cladistic analyses than a stase by stase approach. Klompen & O'Connor worked on Acari.

Table VI. Preparation for the size reduction of Table V; characters to be combined are shown by double arrows. See text for further explanations.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
selected instars	o_2+o_3	o_2+o_3	o_1	o_3+o_4	o_3+o_4	$o_2+o_3+o_4$	o_3+o_4	o_3+o_4	$o_1+o_2+o_3+o_4$
t_1	1	2	1	2	1.6	2.4	2	2	4
t_2	1	2	0.5	1.8	2	2.4	0	2	2
t_3	0.2	1	0	0.2	0	0	1	1	0.5
t_4	0	1	1	0	0	0	0	0	2
	⟷			⟷					

Table VII. Correspondence to Table V reduced: $c_{1,2}$ corresponds to the combination of c_1 and c_2 of Table V; $c_{4,5}$ corresponds to the combination of c_4 and c_5 of Table V. See text for further explanations.

	$c_{1,2}$	c_3	$c_{4,5}$	c_6	c_7	c_8	c_9
selected instars	o_2+o_3	o_1	o_3+o_4	$o_2+o_3+o_4$	o_3+o_4	o_3+o_4	$o_1+o_2+o_3+o_4$
t_1	3	1	3.6	2.4	2	2	4
t_2	3	0.5	3.8	2.4	0	2	2
t_3	1.2	0	0.2	0	1	1	0.5
t_4	1	1	0	0	0	0	2

Their data were more difficult to code than those for Collembola. I here outline a simple method for use in Collembola (this method cannot directly be applied to the data of Klompen & O'Connor). Nevertheless, the two approaches are similar. In fact, when I add setal presence probabilities of several selected instars, I do not make an instar by instar study, but I really carry out a transformation pattern approach. The main difference between my study and that of Klompen & O'Connor lies in the nature of the data. The biological material I have studied – Collembola Symphypleona – shows transformation patterns that are quite homogeneous, because, as a rule, setal presence probability either does not change or increases during development.

In Collembola Symphypleona the number of juvenile instars is fixed within every species, but varies between families (e.g., Bourletiellidae have three juvenile instars whereas Sminthuridae have four). The SIC table has been defined for taxa displaying the same number of instars. If we consider both families Bourletiellidae and Sminthuridae together (likely they are sister groups), a spreading projection can be performed on each group, and the instars-species/characters tables can be joined before the setal clustering. As far as the numeric projection is concerned, analyses will be performed on each table, and two cladistic matrices will be thus achieved. It can be difficult or irrelevant to merge these matrices and the practical solution consists in analyzing each matrix separately (cladistic analysis). Nevertheless, comparisons between groups, i.e., between matrices, may be useful for polarizing characters, but this is another topic.

The SIC table has been defined for species always passing by the same number of instars. Yet, many arthropods present an extensive variability in the number of instars. However, several instars may generally be gathered in "stases" (André, 1989a). Every stase shows a fixed chaetotaxic pattern (direct consequence of the definition of the concept of stase), so that the SIC table analysis can also be applied providing that the number of stases does not vary between taxa. Comparative studies of the first instar constitute a worthwhile introduction to ontophylogenetic tasks. Recently Pomorski (1996) has successfully applied the morphology of the first instar to the generic classification of Onychiurinae (Collembola, Arthropleona).

It would be valuable to develop database software to make analyses of tridimensional tables easier. Such software would spur phylogenetic research based on chaetotaxy in arthropods.

Our thinking about evolution is probably too influenced by studies on vertebrates. The nature of characters, at the specimen level, such as presence vs. absence of many "petits caractères discontinus" (i.e., setae, Grandjean, 1951), is sharply different from the nature of the characters usually used in vertebrate analyses. Indeed, we are used to see organs as fixed characteristics of species. This perspective is not consistent with the study of setae in arthropods; it is much too deterministic and cannot explain the frequency of specimens which display a chaetotaxy with one or several differences from the standard of the species. As probabilistic organs are not simply anomalies in developmental genetics, systematists should pay more attention to this phenomenon.

Acknowledgments

I thank H. M. André, L. Bonnet, D. R. Brooks, K. Christiansen, L. Deharveng, S. Hopkin, P. Tassy, Q. D. Wheeler, and an anonymous reviewer for comments and helpful suggestions on the manuscript.

References

André, H. M., 1986. The stases in Collembola: a case study. In: R. Dallai (ed.), Second international seminar on Apterygota. Università di Siena, Italy: 301-305.

André, H. M., 1988. Age-dependent evolution: from theory to practice. In: C. J. Humphries (ed.), Ontogeny and systematics: 137-187 (Columbia University Press, New York).

André, H. M., 1989a. The concept of stase. In: H. M. André & J.-Cl. Lions (eds.), L'ontogenèse et le concept de stase chez les Arthropodes / Ontogeny and concept of stase in arthropods: 3-14 (AGAR Publishers, Wavre, Belgium).

André, H. M., 1989b. Ontogenetic trajectories in arthropods. In: H. M. André & J.-Cl. Lions (eds.), L'ontogenèse et le concept de stase chez les Arthropodes / Ontogeny and concept of stase in arthropods: 83-90 (AGAR Publishers, Wavre, Belgium).

Betsch, J. M. & A. Waller, 1994. Chaetotaxic nomenclature of the head, thorax and abdomen in Symphypleona (Insecta, Collembola). Acta zool. Fenn., 195: 5-12.

Campos-Ortega, J. A., 1988. Cellular interactions during early neurogenesis of Drosophila melanogaster. Trends Neurosci., 11 (9): 400-405.

De Pinna, M. C. C., 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics, 7: 367-394.

Doe, C. Q. & C. S. Goodman, 1985. Early events in insect neurogenesis. II. Role of cell interactions and cell lineage in the determination of neural precursor cells. Dev. Biol., 111: 206-209.

Eldredge, N. & J. Cracraft, 1980. Phylogenetic patterns and the evolutionary process: 1-349 (Columbia University Press, New York).

Grandjean, F., 1938. Sur l'ontogénie des Acariens. C. r. hebd. Séanc. Acad. Sci., Paris, 206: 146-150.

Grandjean, F., 1939. La répartition asymétrique des organes aléatoires. C. r. hebd. Séanc. Acad. Sci., Paris, 208 : 861-864.

Grandjean, F., 1941. Sur la priorité dans les groupes d'organes homéotypes qui évoluent par tout ou rien. C. r. hebd. Séanc. Acad. Sci., Paris, 213: 417-421.

Grandjean, F., 1948. Sur les écarts dans un clone de Platynothrus peltifer (Acarien). C. r. hebd. Séanc. Acad. Sci., Paris, 227: 638-661.

Grandjean, F., 1949. Au sujet des variations individuelles et des polygones de fréquence. C. r. hebd. Séanc. Acad. Sci., Paris, 229: 801-804.

Grandjean, F., 1951. Les relations chronologiques entre ontogenèses et phylogenèses d'après les petits caractères discontinus des Acariens. Bull. Biol., 3: 269-292.

Grandjean, F., 1971. Caractères anormaux et vertitionnels rencontrés dans les clones de Platynothrus peltifer (Koch). Première partie. Acarologia, 13 (1): 209-237.

Hammen, L. van der, 1981. Numerical changes and evolution in actinotrichid mites (Chelicerata). Zool. Verh. Leiden, 182: 1-47.

Heitzler, P. & P. Simpson, 1991. The choice of cell fate in the epidermis of Drosophila. Cell, 64: 1083-1092.

Held, L. I., Jr., 1990a. Arrangement of bristles as a function of bristle number on a leg segment in Drosophila melanogaster. Roux's Arch. dev. Biol., 199: 48-62.

Held, L. I., Jr., 1990b. Sensitive periods for abnormal patterning on a leg segment in Drosophila melanogaster. Roux's Arch. dev. Biol., 199: 31-47.

Hennig, W., 1966. Phylogenetic systematics: 1-263 (University of Illinois Press, Urbana).

Klompen, J.S.H. & B.M. O'Connor, 1989. Ontogenetic patterns and phylogenetic analysis in Acari. In: H. M. André & J.-Cl. Lions (eds.), L'ontogenèse et le concept de stase chez les Arthropodes / Ontogeny and concept of stase in arthropods: 91-103 (AGAR Publishers, Wavre, Belgium)

Kroonenberg, P. M., 1983. Tree-mode principal component analysis: theory and applications: 1-399 (DSWO Press, Leiden).

Matsakis, J. Th., 1967. Sur certaines variations quantitatives affectant les appendices arthropodiens: cas des griffes de l'Oribate Rhysotritia ardua. Bull. Soc. Hist. nat. Toulouse, 103: 590-631.

Nayrolles, P., 1992. Aspects structuraux de la chétotaxie appendiculaire des Collemboles Symphypléones. Nouv. Revue Ent. (N.S.), 9 (4): 345-356.

Nayrolles, P., 1993a. La biométrie des caractères discontinus d'après le revêtement appendiculaire des Collemboles Symphypléones. I. Sur quelques concepts nouveaux dans l'analyse numérique de la chétotaxie. Bull. Mus. natl. Hist. nat., Paris (4e sér.), 15: 79-93.

Nayrolles, P., 1993b. A standardized description of European Sminthuridae (Collembola, Symphypleona), 1: genera Lipothrix, Gisinurus, and Caprainea. Bijdr. Dierk., 63 (1): 43-60.

Nayrolles, P., 1994. A standardized description of European Sminthuridae (Collembola, Symphypleona), 2: review of four species of the genera Allacma and Spatulosminthurus. Bijdr. Dierk., 64 (3): 151-162.

Nayrolles, P., 1995a. Propositions pour un retour à la définition originelle de l'homologie, d'après l'analyse de la chétotaxie des Collemboles. Bull. Soc. zool. France, 120 (1): 3-20.

Nayrolles, P., 1995b. A standardized description of European Sminthuridae (Collembola, Symphypleona), 3: description of seven species of Sminthurus, including four new to science. Bijdr. Dierk., 64 (4): 215-237.

Nayrolles, P., 1996. Ontogenetic trajectories in Symphypleona (Collembola). J. Morphol., 227 (2): 127-143.

Nixon, K. C. & Q. D. Wheeler, 1990a. An amplification of the phylogenetic species concept. Cladistics, 6: 211-223.

Nixon, K. C. & Q. D. Wheeler, 1990b. Extinction and the origin of species. In: M. J. Novacek & Q. D. Wheeler (eds.), Extinction and cladistic analysis: 119-143 (Columbia University Press, New York).

Patterson, C., 1988. Homology in classical and molecular biology. Mol. Biol. Evol., 5 (6): 603-625.

Platnick, N. I., 1979. Philosophy and the transformation of cladistics. Syst. Zool., 28: 537-546.

Pomorski, R. J., 1996. The first instar larvae of Onychiurinae - a systematic study (Collembola: Onychiuridae). Genus (Wroclaw), 7 (1): 1-102.

Signoret, J. & J. Lefresne, 1969. Description et interprétation d'une nouvelle mutation "microphtalmie aléatoire" chez l'Axolotl (Ambystoma mexicanum Shaw). C. r. hebd. Séanc. Acad. Sci. Paris, ser. D, 269 (4): 507-509.

Simpson, P., 1990. Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of Drosophila. Development, 109: 509-519.

Stern, C., 1954. Two or three bristles. Am. Scient., 42: 213-247.

Wilcox, M., G. J. Mitchison & R. J. Smith, 1973. Pattern formation in the blue-green alga, Anabaena. I. Basic mechanisms. J. Cell Sci., 12: 707-723.

Wiley, E. O., 1981. Phylogenetics: the theory and practice of phylogenetic systematics: 1-439 (Wiley, New York).

Zadeh, L. A., 1965. Fuzzy sets. Information and Control, 8: 338-353.

Received: 10 September 1997

Appendix

I provide a concrete example of data treatment based on the tibiotarsal chaetotaxy observed in 14 species of the subfamily Sminthurinae (descriptions in Nayrolles, 1993b, 1994, 1995b). The species and the abbreviations used in the tables are listed below:

Gisinurus malatestai Dallai, 1970	Gi mal
Caprainea bremondi (Delamare & Bassot, 1957)	Ca bre
Caprainea marginata (Schött, 1893)	Ca mar
Allacma fusca (Linnaeus, 1758)	Al fus
Allacma gallica (Carl, 1899)	Al gal
Spatulosminthurus betschi Nayrolles, 1990	Sp bet
Spatulosminthurus lesnei (Carl, 1899)	Sp les
Sminthurus bourgeoisi Nayrolles, 1995	Sm bou
Sminthurus nigromaculatus Tullberg, 1871	Sm nig
Sminthurus viridis (Linnaeus, 1758)	Sm vir
Sminthurus leucomelanus Nayrolles, 1995	Sm leu
Sminthurus bozoulensis Nayrolles, 1995	Sm boz
Sminthurus hispanicus Nayrolles, 1995	Sm his
Sminthurus multipunctatus Schäffer, 1896	Sm mul

The tibiotarsus is the subapical segment of legs in Collembola. It bears eight generatrices and five whorls of primary setae in Symphypleona. Secondary setae appear between whorls during development. Fig. 6 shows the setae liable to be present on tibiotarsi of the species observed and provides a straightforward display of the setal nomenclature. A code for designating the tibiotarsus (or any other segment of legs) is added to the symbol of the seta: TI1 for the fore tibiotarsus, TI2 for the mid tibiotarsus, and TI3 for the hind tibiotarsus. For example, (TI1)Ia is the seta of the fore tibiotarsus situated on the first whorl and on the anterior generatrix. When two setae in the same place on two tibiotarsi are considered together, we write the number of both tibiotarsi separated by a comma, e.g., (TI1,2)IIa. A period replaces numbers "1, 2, 3" to designate a seta on all tibiotarsi, e.g., (TI.)IIIa.

In chaetotaxic descriptions of Symphypleona, the instar of appearance of a seta is given by a letter: P for a primary seta, and respectively D, T, Q and C for a seta emerging at the second, third, fourth or fifth instar. I keep these symbols (they correspond to the first letter of the French adjective "premier", "deuxième", "troisième", etc.) in order to provide a real example based on a current standard of description. The sign – is for a seta absent. When a seta is variable at the instar in which it appears, the letter that symbolizes this instar is written between parentheses; if in a later instar it becomes constant, this instar is given as well. For example, (Q) means that a seta appears with variability at the fourth instar and remains variable; (T)Q means that a seta appears with variability at the third instar and becomes constant at the fourth. These symbols are used in the alphabetic projection of the SIC table (Table VIII). In Sminthurinae, tibiotarsal setae appear in first, third, or fourth instars (letters P, T, and Q in Table VIII); the fifth instar is the adult and shows the same tibiotarsal chaetotaxy as the fourth instar. A number with one digit, named occurrence (Nayrolles, 1993a), provides an estimation of the probability of presence of a variable seta. Below are listed the occurrences of variable setae with the letters of the instar of appearance:

Gi mal	(TI3)2ae	(Q)	0.8
Ca bre	(TI3)O2pe	(T)Q	0.5
Ca bre	(TI3)O3pe	(T)Q	0.6
Sp les	(TI1)FSpe↓	(T)Q	0.3
Sp les	(TI2)FSpe↓	(T)Q	0.4
Sm bou	(TI1)O1ae	(Q)	0.8
Sm bou	(TI3)3a	(T)Q	0.6
Sm bou	(TI3)4a1	(T)Q	0.3
Sm bou	(TI1)4i1	(Q)	0.3
Sm nig	(TI3)3a	(T)Q	0.4
Sm nig	(TI3)4a1	(T)Q	0.2
Sm nig	(TI3)FSpe↓	(T)Q	0.3
Sm vir	(TI2)3a	(T)Q	0.8
Sm vir	(TI1)3a	(T)Q	0.8
Sm vir	(TI1)4a1	(T)Q	0.5
Sm vir	(TI2)4a1	(T)Q	0.6
Sm vir	(TI3)4a1	(T)Q	0.7
Sm vir	(TI3)2p	(T)Q	0.5
Sm vir	(TI1)3p	(T)Q	0.6
Sm vir	(TI2)3p	(T)Q	0.5
Sm vir	(TI3)3p	(T)Q	0.2
Sm vir	(TI2)4p1	(Q)	0.5
Sm leu	(TI1)4a1	(T)Q	0.4
Sm leu	(TI2)4a1	(T)Q	0.6
Sm leu	(TI3)4a1	(T)Q	0.6
Sm leu	(TI3)2p	(T)Q	0.6
Sm leu	(TI1)3p	(T)Q	0.7
Sm leu	(TI2)3p	(T)Q	0.6
Sm leu	(TI3)3p	(T)Q	0.3
Sm boz	(TI2)4a1	(T)Q	0.7
Sm boz	(TI3)4a1	(T)Q	0.7
Sm boz	(TI3)2p	(T)Q	0.7
Sm boz	(TI2)3p	(T)Q	0.7
Sm boz	(TI3)3p	(T)Q	0.6
Sm boz	(TI3)4p1	(Q)	0.6
Sm boz	(TI1)FSpe↓	(T)Q	0.8
Sm boz	(TI2)FSpe↓	(T)Q	0.8
Sm his	(TI2)3p	(T)Q	0.6
Sm his	(TI2)4p1	(Q)	0.6

In practice, we build up the alphabetic table first in order to define the set of phylogenetically constant characters, C_p, and remove it from the rest of the analysis. Generic comparison can then be achieved by defining the subset of characters which differentiate the genera. Table IX presents the results for the genera studied. In this table, the standard of setal notation is used, e.g., (TI.)IIi is the seta IIi considered on the three pairs of tibiotarsi. Another set of characters is of interest for systematists; this is C_o, the set of ontogenetically constant characters (Table VIII), because these characters permit us to distinguish species from the first instar on.

In the part "Definition and properties of a SIC table" it has been stated that the set of characters to be retained for a subset of taxa T_α is $C_{q\alpha}$, the set of characters phylogenetically variable within T_α. An example of this process applied to the genus *Sminthurus* is provided in Table X. The main difference among species relates to the stage of development for which setae appear: either at the third instar or at the fourth. From a study of ontogenetic trajectories applied to tibiotarsi of Symphypleona, it was assessed "that the long ontogenetic trajectory of *S. viridis* fits with an evolution of the tibiotarsal chaetotaxy, this evolution being a neochaetosis (appearance of setae). All adults of the species of subfamily Sminthurinae that I have observed have many secondary setae on tibiotarsi. Therefore, the intensive tibiotarsal neochaetosis would be a derived character of Sminthurinae" (Nayrolles, 1996: 133-134). We can thus assume that the setae appearing at the fourth instar of Sminthurinae get an earlier instar of emergence in certain species of *Sminthurus* (especially *S. multipunctatus*). This hypothesis will be tested "cladistically" in a future paper.

The spreading projection combined with a setal cluster is shown in Table XI. To define setae to be clustered, the similarity and ordinal criteria are applied on sets of homeotypic setae. For the similarity criterion, we will consider that a difference inferior or equal to 0.3 between two values is acceptable, i.e., it does not force us to keep the setae separate. Comparisons between profiles and other features (e.g., setal shape) show that the setae to be clustered are those situated in the same place on the tibiotarsi (relationship 3 in Fig. 5), and/or those situated on the same generatrix (relationship 2 in Fig. 5). As far as the anterior and posterior generatrices are concerned, we observe that the seta in position 3, e.g. (TI1)3a, has an occurrence superior or equal to the seta in position 4, e.g. (TI1)4a1. We can thus cluster the setae in pairs 3a, 4a1 or 3p, 4p1 for every tibiotarsus (e.g., character 6 in Table XI). This process entails a problem for the numeric projection: the same numeric value can be achieved in two different ways. As a consequence, there are more characters in the table of the numeric projection (Table XII) than in the table of the spreading projection (Table XI). For instance, if the setae (TI1)FSp and (TI2)FSp were clustered (characters 16 and 17 in Table XII), the numeric values would have been the same (2) for *Gisinurus* and *Caprainea*, though the ontogeny of both setae is distinct between these genera (the problem raised by the ordinal criterion was discussed for the numeric projection).

Differences between third instars of *Sminthurus* can be portrayed with ontogenetic trajectories (Fig. 7). The initial matrix corresponds to Table XI with the difference that only the instars distinguishable from each other are retained (they correspond to Operational Semaphorontic Units, see André, 1988; Nayrolles, 1996). Thus, for each species, the first instar (identical to the second) and the fifth instar (identical to the fourth) are removed. Correspondence analyses were performed on some matrices which have been produced by different data treatments applied to the initial array. As it has been previously noticed (Nayrolles, 1996), the most accurate analysis corresponds to the splitting of characters in grade and antigrade with the same scale grading for all characters. The inertia of the plan defined by the two first axes is 73% (data are highly structured). Axis F1 corresponds to ontogenetic differences, and F2 to phylogenetic differences. Within the genus *Sminthurus*, it is obvious that the main difference lies

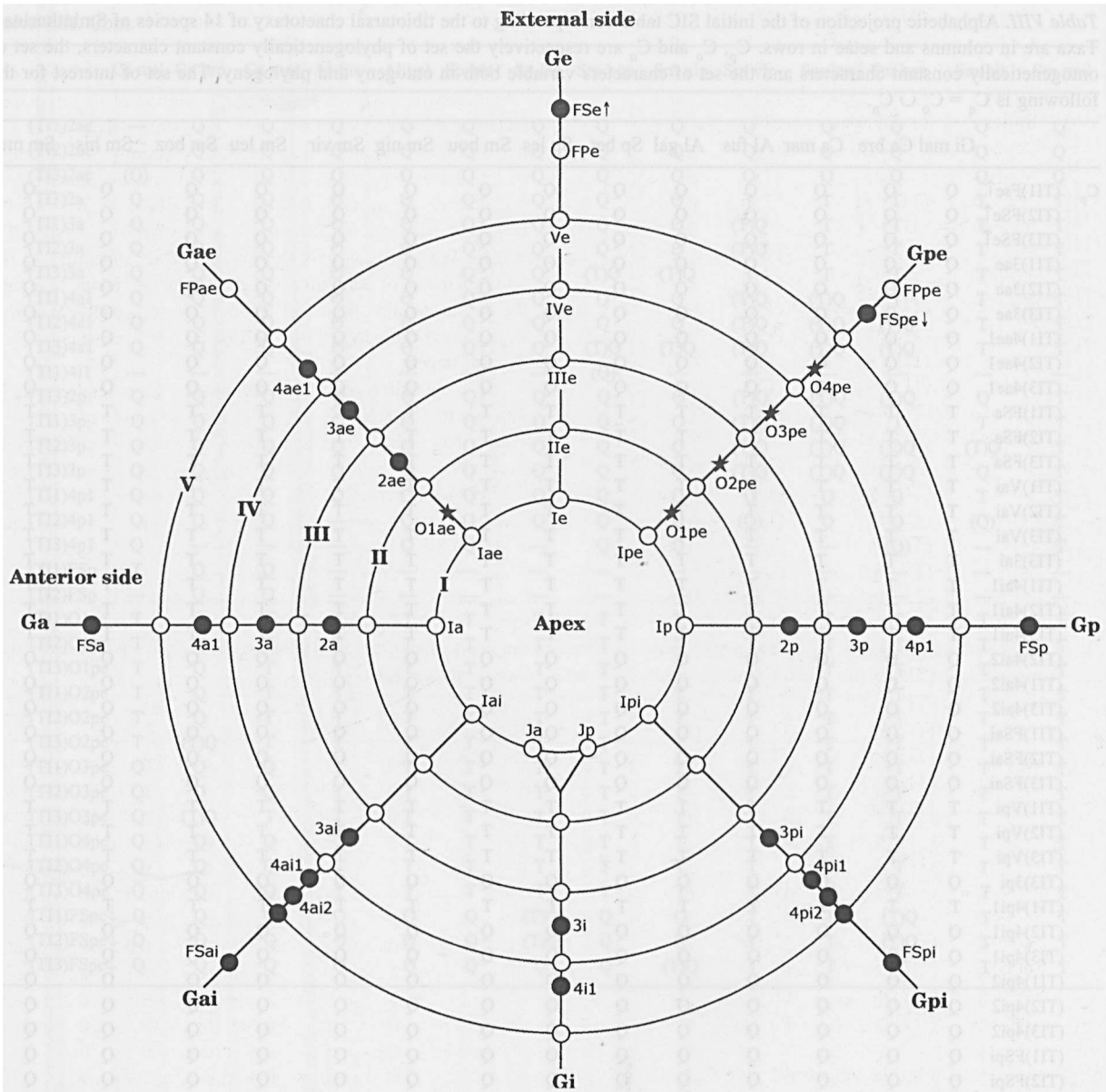


Fig. 6. Schematic representation of the tibiotarsal chaetotaxy. The figure is synthetic and presents all setae which can be observed on tibiotarsi of Sminthurinae. The setae are schematized as follows: an open symbol for a primary seta, a full symbol for a secondary seta, a circle for a normal shaped seta, a star for a special shaped seta (oval organ). Ge, Gae, Ga, Gai, Gi, Gpi, Gp, and Gpe are the eight generatrices in the following positions: external, anterior-external, anterior, anterior-internal, internal, posterior-internal, posterior, and posterior-external. There are five whorls numbered I, II, III, IV and V from apex to basis. The symbol for a seta on a whorl combines the number of the whorl with the letter(s) of the generatrix, e.g., Ie, IIe, ... Ve for the generatrix Ge. The basal part is named F, and the letters FP and FS are used to designate primary and secondary setae. The first whorl bears two setae, Ja and Jp, instead of one at the place of the internal generatrix. The secondary setae which appear in inter-whorls are numbered with Arab numerals.

in the position of third instars along the axis F1, i.e., along the ontogenetic gradient.

Table XII could be used to perform a cladistic analysis, but there are other characters (e.g., antennal chaetotaxy) and data of other species available. A future paper will deal with this issue as well as polarization of characters.

Table VIII. Alphabetic projection of the initial SIC table corresponding to the tibiotarsal chaetotaxy of 14 species of Sminthurinae. Taxa are in columns and setae in rows. C_p, C_o and C_n are respectively the set of phylogenetically constant characters, the set of ontogenetically constant characters and the set of characters variable both in ontogeny and phylogeny. The set of interest for the following is $C_q = C_o \cup C_n$.

		Gi mal	Ca bre	Ca mar	Al fus	Al gal	Sp bet	Sp les	Sm bou	Sm nig	Sm vir	Sm leu	Sm boz	Sm his	Sm mul
C_p	(TI1)Fse↑	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)FSe↑	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)FSe↑	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)3ae	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)3ae	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)3ae	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)4ae1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)4ae1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)4ae1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)FSa	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)FSa	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)FSa	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI1)Vai	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)Vai	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)Vai	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)3ai	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI1)4ai1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)4ai1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)4ai1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)4ai2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)4ai2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)4ai2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)FSai	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)FSai	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)FSai	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)Vpi	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)Vpi	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)Vpi	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)3pi	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)4pi1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)4pi1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)4pi1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)4pi2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)4pi2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)4pi2	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI1)FSpi	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI2)FSpi	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)FSpi	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)3i	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI2)4i1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
	(TI3)4i1	T	T	T	T	T	T	T	T	T	T	T	T	T	T
C_o	(TI1)Ia	P	—	P	P	P	P	P	P	P	P	P	P	P	P
	(TI2)Ia	P	—	P	P	P	P	P	P	P	P	P	P	P	P
	(TI3)Ia	P	—	P	P	P	P	P	P	P	P	P	P	P	P
	(TI1)IIi	—	P	P	P	P	P	P	P	P	P	P	P	P	P
	(TI2)IIi	—	P	P	P	P	P	P	P	P	P	P	P	P	P
	(TI3)IIi	—	P	P	P	P	P	P	P	P	P	P	P	P	P
	(TI2)Vp	—	—	—	—	—	—	—	P	P	P	P	P	P	P
	(TI3)Vp	—	—	—	—	—	—	—	P	P	P	P	P	P	P
C_n	(TI1)O1ae	—	Q	Q	Q	Q	Q	Q	(Q)	Q	Q	Q	Q	Q	Q
	(TI2)O1ae	—	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
	(TI3)O1ae	—	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q

Table VIII. Cont.

	Gi mal	Ca bre	Ca mar	Al fus	Al gal	Sp bet	Sp les	Sm bou	Sm nig	Sm vir	Sm leu	Sm boz	Sm his	Sm mul
(TI1)2ae	—	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
(TI2)2ae	—	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
(TI3)2ae	(Q)	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q
(TI3)2a	Q	Q	Q	Q	Q	Q	Q	Q	Q	T	T	T	T	T
(TI1)3a	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	T	T	T	T
(TI2)3a	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	T	T	T	T
(TI3)3a	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	T	T	T	T	T
(TI1)4a1	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	T	T	T
(TI2)4a1	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	(T)Q	T	T
(TI3)4a1	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	(T)Q	(T)Q	(T)Q	T	T
(TI1)4i1	—	—	—	—	—	—	—	(Q)	—	—	—	—	—	—
(TI3)2p	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	(T)Q	Q	T
(TI1)3p	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	T	T	T
(TI2)3p	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	(T)Q	(T)Q	T
(TI3)3p	Q	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	(T)Q	(T)Q	Q	T
(TI1)4p1	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	Q	T	T
(TI2)4p1	Q	Q	Q	—	Q	Q	Q	Q	Q	(Q)	Q	Q	(Q)	T
(TI3)4p1	Q	—	—	—	Q	Q	—	Q	Q	—	—	(Q)	—	T
(TI1)FSp	T	Q	Q	—	—	—	—	—	—	—	—	—	—	—
(TI2)FSp	—	Q	Q	—	—	—	—	—	—	—	—	—	—	—
(TI1)O1pe	T	Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI2)O1pe	T	Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI3)O1pe	T	Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI1)O2pe	T	Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI2)O2pe	T	Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI3)O2pe	T	(T)Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI1)O3pe	Q	Q	Q	T	T	T	T	T	T	T	T	T	T	T
(TI2)O3pe	Q	Q	Q	T	T	T	T	T	T	T	T	T	T	T
(TI3)O3pe	Q	(T)Q	T	T	T	T	T	T	T	T	T	T	T	T
(TI1)O4pe	Q	Q	Q	T	T	T	T	T	T	T	T	T	T	T
(TI2)O4pe	Q	Q	Q	T	T	T	T	T	T	T	T	T	T	T
(TI3)O4pe	Q	Q	Q	T	T	T	T	T	T	T	T	T	T	T
(TI1)FSpe↓	Q	Q	Q	Q	Q	Q	(T)Q	Q	Q	T	T	(T)Q	T	T
(TI2)FSpe↓	Q	Q	Q	Q	Q	Q	(T)Q	Q	Q	T	T	(T)Q	T	T
(TI3)FSpe↓	Q	Q	Q	Q	Q	Q	Q	Q	(T)Q	T	T	T	T	T

Table IX. Distinction of genera. The characters are constant within each genus and variable between genera.

	Gisinurus	*Caprainea*	*Allacma*	*Spatulosminthurus*	*Sminthurus*
(TI.)IIi	—	P	P	P	P
(TI2,3)Vp	—	—	—	—	P
(TI2,3)O1ae	—	Q	Q	Q	Q
(TI1,2)2ae	—	Q	Q	Q	Q
(TI3)2ae	(Q)	Q	Q	Q	Q
(TI1)FSp	T	Q	—	—	—
(TI2)FSp	—	Q	—	—	—
(TI1,2)O3pe	Q	Q	T	T	T
(TI.)O4pe	Q	Q	T	T	T

Table X. Distinction of the species of *Sminthurus*.

	Sm bou	Sm nig	Sm vir	Sm leu	Sm boz	Sm his	Sm mul
(TI1)O1ae	(Q)	Q	Q	Q	Q	Q	Q
(TI3)2a	Q	Q	T	T	T	T	T
(TI1)3a	Q	Q	(T)Q	T	T	T	T
(TI2)3a	Q	Q	(T)Q	T	T	T	T
(TI3)3a	(T)Q	(T)Q	T	T	T	T	T
(TI1)4a1	Q	Q	(T)Q	(T)Q	T	T	T
(TI2)4a1	Q	Q	(T)Q	(T)Q	(T)Q	T	T
(TI3)4a1	(T)Q	(T)Q	(T)Q	(T)Q	(T)Q	T	T
(TI1)4i1	(Q)	—	—	—	—	—	—
(TI3)2p	Q	Q	(T)Q	(T)Q	(T)Q	Q	T
(TI1)3p	Q	Q	(T)Q	(T)Q	T	T	T
(TI2)3p	Q	Q	(T)Q	(T)Q	(T)Q	(T)Q	T
(TI3)3p	Q	Q	(T)Q	(T)Q	(T)Q	Q	T
(TI1)4p1	Q	Q	Q	Q	Q	T	T
(TI2)4p1	Q	Q	(Q)	Q	Q	(Q)	T
(TI3)4p1	Q	Q	—	—	(Q)	—	T
(TI1)FSpe↓	Q	Q	T	T	(T)Q	T	T
(TI2)FSpe↓	Q	Q	T	T	(T)Q	T	T
(TI3)FSpe↓	Q	(T)Q	T	T	T	T	T

Table XI. Spreading projection with homeotypic setae clustered on the basis of similar or ordered profiles. Each species is divided into five instars, the last is the adult. Characters with their setal components are listed below.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Gi mal	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	1	6	0	0	0
	3	0	0	0.8	1	2	2	2	0	1	2	2	2	1	6	6	2	1
	3	0	0	0.8	1	2	2	2	0	1	2	2	2	1	6	6	2	1
Ca bre	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0.6	0	0
	0	3	0	6	1	2	2	2	0	1	2	2	1	2	6	6	2	1
	0	3	0	6	1	2	2	2	0	1	2	2	1	2	6	6	2	1
Ca mar	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	6	1	0	0
	3	3	0	6	1	2	2	2	0	1	2	2	1	2	6	6	2	1
	3	3	0	6	1	2	2	2	0	1	2	2	1	2	6	6	2	1
Al fus	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	6	6	0	0
	3	3	0	6	1	2	2	2	0	1	2	1	1	0	6	6	2	1
	3	3	0	6	1	2	2	2	0	1	2	1	1	0	6	6	2	1
Al gal	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	6	6	0	0
	3	3	0	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
	3	3	0	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
Sp bet	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	6	6	0	0

Table XI. Cont.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	3	3	0	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
	3	3	0	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
Sp les	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	0	0	0	0	0	0	0	0	0	0	0	0	6	6	0.7	0
	3	3	0	6	1	2	2	2	0	1	2	2	1	0	6	6	2	1
	3	3	0	6	1	2	2	2	0	1	2	2	1	0	6	6	2	1
Sm bou	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0.9	0	0	0	0	0	0	6	6	0	0
	3	3	2	5.8	1	2	2	2	0.3	1	2	2	2	0	6	6	2	1
	3	3	2	5.8	1	2	2	2	0.3	1	2	2	2	0	6	6	2	1
Sm nig	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0.6	0	0	0	0	0	0	6	6	0	0.3
	3	3	2	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
Sm vir	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	1	1.3	1.4	1.7	0	0.5	0.6	0.5	0.2	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	1.5	1	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	1.5	1	0	6	6	2	1
Sm leu	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	1	1.4	1.6	1.6	0	0.6	0.7	0.6	0.3	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	1	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	1	0	6	6	2	1
Sm boz	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	1	2	1.7	1.7	0	0.7	1	0.7	0.6	0	6	6	1.6	1
	3	3	2	6	1	2	2	2	0	1	2	2	1.6	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	1.6	0	6	6	2	1
Sm his	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	1	2	2	2	0	0	2	0.6	0	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	1.6	1	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	1.6	1	0	6	6	2	1
Sm mul	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	2	0	1	2	2	2	0	1	2	2	2	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1
	3	3	2	6	1	2	2	2	0	1	2	2	2	0	6	6	2	1

character number	set of setae	number of setae	character number	set of setae	number of setae	character number	set of setae	number of setae
1	(TI.)Ia	3	7	(TI2)3a & (TI2)4a1	2	13	(TI3)3p & (TI3)4p1	2
2	(TI.)IIi	3	8	(TI3)3a & (TI3)4a1	2	14	(TI1,2)FSp	2
3	(TI2,3)Vp	2	9	(TI1)4i1	1	15	(TI.)O1pe & (TI.)O2pe	6
4	(TI.)O1ae & (TI.)2ae	6	10	(TI3)2p	1	16	(TI.)O3pe & (TI.)O4pe	6
5	(TI3)2a	1	11	(TI1)3p & (TI1)4p1	2	17	(TI1,2)FSpe↓	2
6	(TI1)3a & (TI1)4a1	2	12	(TI2)3p & (TI2)4p1	2	18	(TI3)FSpe↓	1

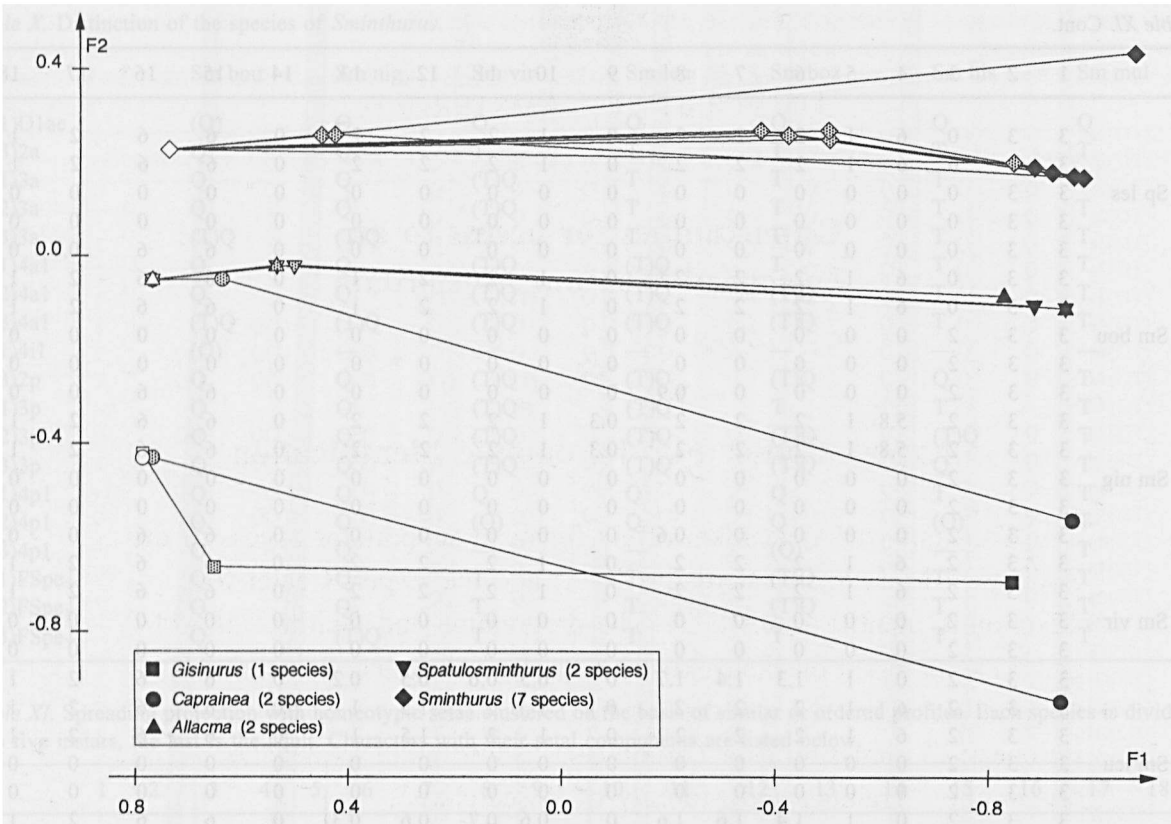


Fig. 7. Ontogenetic trajectories of Sminthurinae. Open symbols: second instars; shaded symbols: third instars; solid symbols: fourth instars. First instars have the same chaetotaxy as second instars, and adults (fifth instars) have the same chaetotaxy as fourth instars.

Table XII. Numeric projection with clustering of homeotypic setae. Characters with their setal components are listed below.

selected instars	1 P	2 P	3 P	4 Q	5 T	6 T	7 T	8 T	9 Q	10 T	11 T	12 T	13 T+Q	14 T	15 T+Q	16 T+Q	17 Q	18 T	19 T	20 T	21 T
Gi mal	3	0	0	0.8	0	0	0	0	0	0	0	0	1	0	1	2	0	6	0	0	0
Ca bre	0	3	0	6	0	0	0	0	0	0	0	0	1	0	0	1	1	0.5	0.6	0	0
Ca mar	3	3	0	6	0	0	0	0	0	0	0	0	1	0	0	1	1	6	1	0	0
Al fus	3	3	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	6	6	0	0
Al gal	3	3	0	6	0	0	0	0	0	0	0	0	1	0	1	0	0	6	6	0	0
Sp bet	3	3	0	6	0	0	0	0	0	0	0	0	1	0	1	0	0	6	6	0	0
Sp les	3	3	0	6	0	0	0	0	0	0	0	0	1	0	0	0	0	6	6	0.7	0
Sm bou	3	3	2	5.8	0	0	0	0.9	0.3	0	0	0	1	0	1	0	0	6	6	0	0
Sm nig	3	3	2	6	0	0	0	0.6	0	0	0	0	1	0	1	0	0	6	6	0	0.3
Sm vir	3	3	2	6	1	1.3	1.4	1.7	0	0.5	0.6	0.5	0.5	0.2	0	0	0	6	6	2	1
Sm leu	3	3	2	6	1	1.4	1.6	1.6	0	0.6	0.7	0.6	1	0.3	0	0	0	6	6	2	1
Sm boz	3	3	2	6	1	2	1.7	1.7	0	0.7	1	0.7	1	0.6	0.6	0	0	6	6	1.6	1
Sm his	3	3	2	6	1	2	2	2	0	0	2	0.6	0.6	0	0	0	0	6	6	2	1
Sm mul	3	3	2	6	1	2	2	2	0	1	2	1	2	1	2	0	0	6	6	2	1

character number	set of setae	number of setae	character number	set of setae	number of setae	character number	set of setae	number of setae
1	(TI.)Ia	3	8	(TI3)3a & (TI3)4a1	2	15	(TI3)4p1	1
2	(TI.)IIi	3	9	(TI1)4i1	1	16	(TI1)FSp	1
3	(TI2,3)Vp	2	10	(TI3)2p	1	17	(TI2)FSp	1
4	(TI.)O1ae & (TI.)2ae	6	11	(TI1)3p & (TI1)4p1	2	18	(TI.)O1pe & (TI.)O2pe	6
5	(TI3)2a	1	12	(TI2)3p	1	19	(TI.)O3pe & (TI.)O4pe	6
6	(TI1)3a & (TI1)4a1	2	13	(TI2)4p1	1	20	(TI1,2)FSpe↓	2
7	(TI2)3a & (TI2)4a1	2	14	(TI3)3p	1	21	(TI3)FSpe↓	1