GENUS PLASMODIUM : STUDIES IN NUMERICAL TAXONOMY

by

A. BOUCKAERT and M. WERY

Summary — Three methods were used to produce a numerical hierarchical classification of malaria parasites. The results are generally in agreement with established use for the subgenera. Disagreements are observed for the definition of some subspecies.

KEYWORDS : Malaria; Plasmodium, Taxonomy; Cluster Analysis; Biometry.

Introduction

The participants to a recent symposium devoted to the nomenclature and systematics of malaria parasites (Garnham, 1973) have suggested that numerical taxonomy could help by providing some objective criteria in this field.

In the present study, which is a preliminary contribution, three methods of cluster analysis were used to built a hierarchical classification of thirty three members of the genus Plasmodium.

The classification of parasitic Protozoa is a special problem in zoological systematics. Some particularly disturbing features will be met when dealing with taxonomy of these organisms:

1) No fossil data are available.

2) The classification of parasites is related to the classification of their hosts. However, this relationship is ambiguous, since parasitism can modify the evolution of the parasites in several respects. For example, an obligatory parasite is subjected to a selective pressure directing its evolution along the host's particular evolution: parasite and host are likely to co-evolve (Bruce-Chwatt, 1965). On the other hand, host exchange is certainly another important process in the general evolution of parasitic species. Since such host changes are likely to follow a direction determined by ecological availability rather than by phylogenetic ancerstorship, no strict parallelism can be expected between parasite taxonomy and host phylogeny.
3) Morphology of protozoa is known almost exclusively by use of the optical microscope together with various staining techniques. Few comparative studies of ultrastructure are available at this time but their number is increasing (Aikawa, 1971). They might become more important in future systematics, but it seems unlikely that, even with the increased resolving power of the electron microscope, the anatomical descriptions of the malaria parasites will ever reach the complexity found in metazoan species. Attention should also be drawn on the artifacts that may appear in preparations during fixation and staining of material.

4) Paucity of morphological criteria is still more evident in bacterial systematics. As a consequence, biochemical characters have been more frequently used as taxonomic criteria. Such a situation is far from unsatisfactory. Biochemical criteria, at least at the level of enzyme activities are more closely related to the genetic structure. On this basis, the rules of bacterial systematics can be considered as continuous with the principles used in zoology and botany where the « gene pool » is considered as the material background of the species concept. Unfortunately, the biochemical description of the malaria parasites is still in its infancy, although stimulating developments are appearing in this field (Gutteridge et al., 1971; Carter et al., 1973).

5) Morphological changes during the life cycle of malaria parasites, as well as the striking changes of the host cells exposed to their influence, are distinctive advantages in establishing comparisons and taxonomic relationships. The problem of separating genetic from environmental factors is however largely unsolved. Answers could be drawn from comparative observations on the evolution of the parasite in natural and experimental hosts. However, examination of malaria parasites in their natural hosts is often recommended for type material collection and description (Bafort et al., 1968).

Material and methods

1. Species of parasites

Thirty three species of the genus Plasmodium were selected for the analysis (table 1). The selection includes six parasites of birds and twenty seven parasites of mammals. The latter group includes all four human species, four parasites of rodents, one parasite of chevrotain and nineteen parasites of apes and monkeys. The descriptive data for most species were borrowed from the textbook of Garnham (1966) where all data published at the time as well as unpublished observations made in the author’s laboratory are reviewed and summarized. Individual references to original material for most species will be found in Garnham’s impressive bibliography. The monography on Primate malarias (Coatney et al., 1971) was also very useful while compiling characters.
TABLE 1
Species and subspecies of the genus *Plasmodium* choosen for this study

| Species | \( P. (G) \) circumflexum | \( P. (H) \) gallinaeum | \( P. (H) \) matutunum | \( P. (H) \) subpraecoix | \( P. (H) \) cathemerium | \( P. (H) \) relictum | \( P. (V) \) tregutili | \( P. (V) \) berghei yoelii | \( P. (V) \) berghei borghei | \( P. (V) \) vinckei chabaudi | \( P. (V) \) vinckei vinckel | \( P. (L) \) falciparum | \( P. (L) \) reichenovi | \( P. (P) \) cynomolgi bastianellii | \( P. (P) \) cynomolgi ceylonensis | \( P. (P) \) cynomolgi cynomolgi | \( P. (P) \) malariae | \( P. (P) \) schwetzii | \( P. (P) \) eylesi | \( P. (P) \) coatneyi | \( P. (P) \) knowlesi | \( P. (P) \) brasiliense | \( P. (P) \) shortii | \( P. (P) \) inui | \( P. (P) \) fieldi | \( P. (P) \) gonderi | \( P. (P) \) simiovale | \( P. (P) \) ovale | \( P. (P) \) vivax | \( P. (P) \) youngi | \( P. (P) \) jefferyi | \( P. (P) \) sylvaticum | \( P. (P) \) pithecii |

2. Characters selected for the description

The sixty seven characters listed in table 2 were used as descriptive items for each of the thirty three species. Forty three characters considered as qualitative are expressed as binary numerals. The states of the dichotomy are usually evident from the way the character has been expressed. E.g. « large nucleus » (char. 1) can assume the values « yes » or « no ». « Vacuole » (char. 2) assumes the values « present » or « absent ». Characters 8, 9 and 10 deserve further comment however. From a strictly geometrical point of view, if the shape of the trophozoite is globular, then its main diameters must be equal. In the present context, if characters 8 and 9 are both assumed to be present, the overall shape is ovoid.

In the geographical characters, we have assumed that the localisation could be described by longitudes (char. 59 to 64) and latitudes (char. 65 to 67). The former were selected rather arbitrarily and the sizes of the covered areas are unequal. E.g. the Pacific zone has the same status as the Middle East. The reason is that we tried to obtain some ecological cohesion in the areas. Moreover, sizes become more equivalent if we consider marine areas as irrelevant to the distribution of the genus.

Actually, all information for a given species was not always available. Therefore, our characters were ternary rather than binary valued, with a neutral state used for unavailable information. As will be seen later, another use of the neutral state was the coding of irrelevant characters. On punching the numerical descriptions of the species, a value of 1 was selected for the « positive » state, 0 for the « negative » state and −1 for the « neutral » state. Any triplet of integers would of course have been satisfactory for computational purposes.

Of the twenty four non dichotomic characters, twenty one were actually polychotomic. They were transformed into binary numerals by suitable coding process. Coloration of the infected red cell (char. 5, 14 and 25) could assume the following values :

1. No change.
2. Lightening.
3. Darkening.
TABLE 2
Descriptive items used in this study

I. Trophozoite (Stage I)
1. Large nucleus
2. Vacuole present
3. Abundant cytoplasm
4. Frequent multiple invasion
5. Coloration of red cell (+)
6. Enlarged size of red cell
7. Granulation of red cell (+)

II. Trophozoite (Stage II)
8. Globular shape
9. Elongated shape
10. Amoeboid shape
11. Vacuoles present
12. Large size
13. Thin pigment granules
14. Red cell coloration (+)
15. Red cell granulations (+)
16. Enlarged red cell

III. Trophozoite (Stage III)
17. Globular shape
18. Size (+)
19. Vacuoles present
20. Nuclear pattern (+)
21. Large nuclear size
22. Heavy pigmentation
23. Dark pigmentation
24. Thin pigmentation
25. Red cell coloration (+)
26. Red cell granulations (+)
27. Enlarged red cell
28. Number of nuclei in mature schizont (+)
29. Length of schizogonic cycle (+)
30. Presence of schizonts in peripheral blood

IV. Gametocytes
31. Shape (+)
32. Size (+)
33. Dispersed pigment in male
34. Dispersed pigment in female
35. Rare occurrence of gametocytes

V. Red Cells
36. Infection of reticuloocytes
37. Invasion of nucleated red cells
38. Shape distortion other than enlargement

VI. Pre-erythrocytic schizont
39. Only one localisation
40. Localisation in liver parenchymal cells
41. Localisation in reticulo-endothelial cells
42. Localisation in capillary endothelium cells
43. Secondary exo-erythrocytic cycle
44. Prepatent period (+)

VII. Sporogony
45. Dispersed pigment in young oocyst

VIII. Vertebrate hosts (+)
46. Primates
   a. Cebidae
   b. Callithricidae
   c. Cercopithecidae
   d. Pongidae
   e. Hominidae
47. Rodentia
   a. Sciuridae
   b. Cricetidae
   c. Muridae
48. Artiodactyla
49. Galliformes
50. Charadriiformes
51. Coraciiformes
52. Passeriformes
53. Large specific distribution

IX. Insect hosts
54. Anopheles sp.
55. Culex sp.
56. Aedes sp.
57. Mansonia sp.
58. Large specific distribution

X. Zoogeography

The (+) quotation means that more than one binary digit is used for coding purposes.

This trichotomic character was coded as two binary characters by replacing «1» by «00», «2» by «01» and «3» by «10». A similar problem was encountered for red cell granulation (char 7, 15 and 26) where the following states were considered:

1. No granules.
2. Rare thin granules.
3. Numerous thin granules.

This tetrachotomic character was reduced to two binary characters by the following natural binary counting sequence: 00, 01, 10 and 11.

The size of the stage III trophozoite (char. 18) and of the gametocytes (char. 32) could assume the following three values:

1. Smaller than the red cell.
2. Equal to red cell's size.
3. Larger than the red cell (unaltered).
The binary characters in this case were obtained by transforming «1» into «00», «2» into «01» and «3» into «11». By comparison with the code used for characters 7, 15, 26, it can be seen that we have tried to conserve the rank of differences between the qualitative states. With the transformation used here, the binary distances (expressed as the number of digits with unequal values) between states 1 and 2, 2 and 3, 1 and 3, are respectively 1, 1 and 2 which is exactly what would have been achieved with decimal integers. Obviously a different and unsatisfactory result would have been obtained if state 3 had been expressed as «10» as the natural number of the binary system corresponding to the decimal number 3.

The nuclear pattern of stage III trophozoite (character 20) was expressed by the three states:
1. Random.
2. Rosette-like.
3. Chessboard-like.

They were coded respectively as 00, 01 and 10.

The shape of the gametocyte (char. 31) could assume the following values:
1. Crescent.
2. Round.
3. Ovoid.
4. Round or ovoid.

These four states were coded by a two-digit binary selected with the corresponding ordinal rank, i.e. 00, 01, 10 and 11.

Parasitism of nucleated red cells (char. 37) deserves a special mention. This character has no meaning so far as mammalian hosts are considered. Consequently, the «neutral value» (−1) was assigned to this character for all species parasitic of mammals.

Host characters (char. 46 to 52 and 54 to 57) could always assume the three following values:
1. (00) Refractory host.
2. (01) Moderately susceptible host.
3. (11) Very susceptible host.

The binary transformation used here conserves the mutual distances (cfr. char. 18 and 32).

Characters 28, 29 and 44 are quantitative rather than qualitative-polychotomic. The number of nuclei of the mature schizont was expressed by five binary integers:
1. Less than 8.
2. From 8 to 12.
3. From 13 to 16.
4. From 17 to 20.
5. Over 20.

E.g. *Plasmodium knowlesi*, with ten nuclei, is described by the following binary quintuplet: «01000». *Plasmodium ovale*, with 4 to 16 nuclei, is described by «11100». For *P. sylvaticum*, we only know that the number of nuclei does not exceed 20. It has been described by «11110».

For the other quantitative characters, we preferred to select a transformation which conserved the rank of mutual differences. The prepatent
period (char. 44) was expressed by a sequence of six binary integers as follows:

1. Over 280 H.  
2. Over 240 H.  
3. Over 200 H.  
4. Over 160 H.  
5. Over 120 H.  
6. Over 80 H.

E.g. *Plasmodium knowlesi* has a prepatent period of 132 H. The corresponding binary sequence becomes: «000011». Similarly, we find «111111» for *P. malariae* (360 H), «000000» for *P. cathemerium* (72 H) and «001111» for *P. circumflexum* (less than 192 H).

The schizogonic period (char. 29) was expressed by two binary integers:

1. Over 40 H.  
2. Over 60 H.

E.g. the schizogonic period of *P. knowlesi*, *P. malariae* and *P. eylesi* (24 H, 72 H, 48 H) are described respectively by «00», «11», and «10». The code preserves the ranks of differences.

Summarizing, we have now 110 binary digits to represent any species of the genus *Plasmodium*. The distribution of the digits is as follows:

a) Dichotomic features: $43 \times 1 = 43$.

b) Polychotomic features: $27 \times 2 = 54$ (including all host characters).

c) Schizogonic cycle (length of -): 2.


e) Number of nuclei in trophozoite stage III: 5.

Total: 110.

This is of course even more than the 75 characters suggested by Garnham (1973). However, the information actually included in each species description is less than 110 binary units. The difference is due to the fact that, from the way they were defined, some characters are not really independent. Let us examine briefly the causes of this loss of information:

a) Some trichotomic characters were expressed by pairs of binary digits, i.e. characters 5, 14, 18, 20, 25, 32, and all the host characters. One of the four possible states of the binary pairs will thus never be used. E.g. no host character is described by «10». Hence, for these forty six binary unit digits, there will be less than one binary unit (bit) of information by digit.

b) The digits used to binarize quantitative characters are not independent. In the description of the schizogonic cycle, 3 (instead of 4) values are possible. In the description of the prepatent period, we have seven possible states, instead of sixty four if the digits were independent. For the number of nuclei, the real information content is still more difficult to guess: it is possible to have descriptions such as «11100» (4 to 16 nuclei) while «10001» (less than 4 or more than 20 nuclei) is meaningless. In this case, the only possible states are those where an uninterrupted sequence of ones of length equal to or less than five is located somewhere over the sequence of the five positions. There are fifteen such states: without this restriction, thirty two different states could be expressed by a binary sequence of five digits.

Taking into account the various particular restrictions, the total information contained in one single species description is now only 95.5 bits.
instead of 110 bits. The loss of information is of course unequally distributed over the digits, the loss being larger for polychotomic and quantitative characters. The information content of digits could of course have been used as weighting coefficient but we did not use such a procedure. It was thought that the quantitative and polychotomic characters were already disadvantaged by the coding process and that no extra penalty should be imposed to them.

Another kind of weighting was considered and applied. It concerned the relative importance of Vertebrate hosts. It should not be correct to give the same importance to families (Cebidae, Sciuridae) and to orders (Passeriformes). Consequently, weights were given to host characters in order that each order should contribute to the final score for the same amount, whether subdivided into families or not. Characters 46 to 52 had thus all identical weights. A similar rule was not applied to insect vectors however, since we thought that it would have led to an unjustified minimization of the latter's contribution to the final score. Obviously, these options are rather subjective, which seems to be an inevitable feature of all taxonomies, whether numerical or conventional.

Summarizing the suggestions for the collection of data necessary to construct a numerical taxonomy of the genus *Plasmodium*, Garnham (1973) quotes 8 kinds of relevant characters. Since we were unaware of these suggestions at the time this study was undertaken, it is interesting to observe the extent of agreement between them and the data used here.

1. Structure of all stages (char. 1 to 45).
2. Hosts: Vertebrate and invertebrate, including primary and experimental (char. 46 to 58).
3. Zoogeography (char. 59 to 67).
4. Life cycles: dynamics (char. 29, 39, 43, 44).
5. Effect on vertebrate hosts:
   i.e. — Changes in erythrocytes (char. 5, 6, 7, 14, 15, 16, 25, 26, 27, 38).
   — Lethality (not used).
   — Course of infection (not used).
   — Cross immunity (not used).
7. Response to drugs (not used).
8. Immunological reactions (not used).

The localization of the suggested characters not used for the present study points out clearly the kind of supplementary information that will be needed in future studies of this kind. Moreover, some species are not only less completely studied, but their life-cycle is often incompletely described, only from a small number of observations. Strains may have been completely unavailable after the original description with the risk of possibly atypical features being considered as highly characteristic, and on the other hand, morphology of sporogonic stages as well as exo-erythrocytic schizont is mostly unknown.

3. Methods of cluster analysis

Using the terminology introduced by Sokal and Sneath (1963), the following methods of cluster analysis were applied to the data:

2. Total linkage (TL).
3. Average linkage (AL).
The result of any of these methods is a hierarchical classification, which can be explained as an agglomerative process starting from a level $L_0$ where all species are considered as separate entities, and proceeding through successive agglomerative fusions up to a level $L_f$ where all species are united into a single entity (in this case, the genus). The intermediate level $L_i$ where species number $i$ is fused with species number $j$ depends on the nature of the particular method of classification used. Without going into too many details about these methods, let us briefly mention that:

1. **The levels are defined as levels of dissimilarity.** The dissimilarity $d_{ij}$ between species $i$ and $j$ is defined as:

$$d_{ij} = 1 - \frac{W_{ijoo} + W_{iill}}{W_{ijoo} + W_{iill} + W_{iiol} + W_{iiilo}}$$

where $W_{ijkm} = \sum_{n=1}^{110} W_n S_{ijkm}$

$s_{ijkm}$ is a variable restricted to the following values:

- 1 if character $n$ assumes value $k$ for species $i$ and value $m$ for species $j$
- 0 in all other cases.

$w_n$ is the weight of character $n$. Its value is 1 except for family host characters as explained before. The clustering process will use a descending scale of $d_{ij}$ for admitting or rejecting the fusions. $L_{ij}$ will be said to be the junction level between species number $i$ and number $j$.

2. **$L_{ij}$ is defined as follows:**

a) $L_{ij} = d_{ij}$ if both species did not fuse with any other at an earlier stage of the clustering process.

b) $L_{ij} = d_{ij}$ if species number $i$ is isolated while species number $j$ is already member of a larger cluster.

c) $L_{ij} = d_{ij}$ if species number $j$ is isolated while species number $i$ is already member of a larger cluster.

d) $L_{ij} = d_{ij}$ if both species are members of separate clusters.

Hence, $d_{mk}$ is the dissimilarity coefficient between species $k$ and the cluster containing species $m$. In the single linkage method, $d_{mk}$ is the minimum value of $d_{jk}$ where species $j$ belongs to the same cluster as species $m$. In the total linkage method, $d_{mk}$ is the maximum value of $d_{jk}$ when species $j$ belong to the same cluster as species $m$. In the average linkage method, $d_{mk}$ is the weighted average of $d_{jk}$ where $j$ takes its value over all elements of the cluster to which species $m$ belongs.

4. **Computing material**

All computations were carried out with the IBM system 1440 using Fortran IV programming. The three linkage criteria were obtained by use of an
algorithmic procedure described e.g. by Wishart (1969), where the linkage
criterion is incorporated into the main routine in a flexible way as a para-
metric function. Owing to the small size of the computer, all intermediate
results had to be stored on external devices.

Results

1. The classifications obtained with methods SL, TL and AL are repre-
sented graphically in figures 1 to 3.

2. The distribution of the 528 dissimilarity coefficients is shown in
figure 4. The mean value is 0.329 with a standard deviation of 0.103. The
shape of the distribution curve is clearly not unimodal. Aside a major
mode centered on 0.280 lesser modes are recognisable around 0.520. The
major mode is also probably heterogenous.

Figure 1
Taxonomic tree according to the single linkage classification method. The values indicated on the
vertical axis are similarity coefficients i.e. $10 \times (1 - \text{dissimilarity})$. 
Discussion

By considering the hierarchical classifications shown on figures 1, 2 and 3, it appears that some patterns are consistently found. These patterns are in some way method-independent. They will be discussed first.

The segregation between parasites of Birds and parasites of Mammals is always the first step in subdivision. This dichotomy occurs at level 0.318 in SL, 0.599 in TL and 0.449 in AL. It remains the only subdivision for a large part of the dissimilarity (or similarity) scale. For example, the second subdivision occurs at level 0.230 in SL, at level 0.436 in TL and at level 0.321
in AL. Since the first agglomerative step occurs at level 0.055 in all methods, the part of the hierarchical tree where the only taxonomic structure is the subdivision between avian and mammalian parasites accounts respectively for 33, 30 and 32 percent of the total height of the tree. It was pointed out (Sokal and Sneath, 1963) that such structures deserve special interest because they are stable, i.e. the same kind of subdivision will be found even if some different or supplementary characters are added to the species descriptions.

a) Bird species

Plasmodia of Birds are always subdivided further according to the same pattern: a first cluster is composed of *P. relictum*, *P. cathemerium*, *P. matutinum* and *P. subpraecox*. This cluster remains homogeneous over some length of the tree, *P. circumflexum* and *P. gallinaceum* remaining outsiders for about one third of the total height of the tree. They join the four membered cluster of avian parasites at about the same level at which fusion occurs with all mammalian species.

Garnham (1966) subdivides the avian species of Plasmodia into four subgenera:

1) *Haemamoeba*, whose main distinctive characters is the round shape of the gametocyte. This subgenus is further subdivided into two groups:
   b. Group II : *H. gallinaceum*, *H. durae*, *H. griffithsi*.

2) *Giovannolai*, whose main distinctive characters are the elongated shape of the gametocyte and the large size of the erythrocytic schizonts, e.g. *G. circumflexum*.

3) *Huttia*.

4) *Novyella*.

The results of the present study support this general classification. However, there is no indication of a closer relationship between *H. gallinaceum* and group I *Haemamoeba* species than between the latter and *G. circumflexum*.

b) Mammalian species

There is always, from the present results, a close and exclusive relationship between *P. reichenovi* and *P. falciparum*. The two membered cluster containing these two species remains separated from all other species for about one third of the total length of the tree whatever clustering method is used. Garnham (1966) considers these two species as the only members of the subgenus *Laverania*.

An interesting situation arises in relation with the subspecies. The first case is the relationship between *P. berghei* and its close relative *P. yoelii*. These two parasites of rodents are considered as subspecies and are accordingly named *P. b. yoelii* and *P. b. berghei*. The first was originally
described as a particular strain by Landau and Chabaud (1965). Its subspecific status was recognized later by Landau and Killick-Kendrick (1966) and has never been challenged. The two subspecies are found here with a coefficient of dissimilarity of only 0.055: we did not find such small value for any other pair of parasites in this study.

The parasites of Rodents, *P. vinckei* and *P. chabaudi* are now considered as subspecies and named respectively *P. vinckei vinckei* and *P. vinckei chabaudi*. In contrast with the berghei case, they were first recognised and named as distinct species (Landau, 1965). Complete description of the sporogonic cycle led Bafort (1968) to suggest that *P. chabaudi* should better be considered as a subspecies of *P. vinckei*. A pattern of segregation between berghei-like and vinckei-like parasites is a unique feature appearing in the TL method. With both AL and SL, *P. v. vinckei* is incorporated into the berghei cluster before the latter is allowed to fuse with *P. v. chabaudi*.

Descriptions of recently isolated strains in various parts of Africa confirm the evidence of the extreme variability of rodent malaria parasites. It is possible that the taxonomy of these parasites will be revised when new biochemical criteria will become available (Garnham, 1973; Walliker, 1973).

A third subspecific entity is encountered with *P. shortti*. This parasite has been considered as a subspecies of *P. inui* and named *P. inui shortti* (Eyles, 1963). This nomenclature has not been widely used. The present results support the close relationship existing between these two parasites. The dissimilarity between them is about the same as that found between *P. v. vinckei* and *P. v. chabaudi*. The two-membered cluster when formed, incorporates other members only shortly thereafter however.

The fourth subspecific distinction encountered is related with the simian malaria parasite *P. cynomolgi*. *P. bastianellii* was first described as a subspecies and named *P. cynomolgi bastianellii* (Garnham, 1959). It was considered later as a distinct species (Bray, 1963). Another subspecies *P. cynomolgi ceylonensis* was first described by Dissanaike et al. (1965). Subsequently, Garnham (1966) expressed some doubt about the close nature of the relationship between *P. c. cynomolgi* and *P. c. ceylonensis*. However, the recent symposium on malaria parasites taxonomy (Garnham, 1973) did not suggest to raise *P. c. ceylonensis* to a specific level. In the three methods, TL, SL and AL, a close relationship was observed between *P. c. cynomolgi* and *P. c. bastianellii*. *P. c. ceylonensis* remains separated from the two other subspecies up to the point where all three are merged in polyspecific clusters in AL and SL. A three-membered cluster including all three subspecies of *P. cynomolgi* appears at some stage of the classification process in TL.

A fifth possible subspecific relation is described for *P. vivax* and *P. schwetzi*. Bray (1958) preferred to consider the differences between them as subspecific and suggested to use the designations *P. vivax vivax* and *P. vivax schwetzi*. This nomenclature did not become generally used. Observations by Collins et al. (1969) revealed supplementary differences between the two parasites and made the subspecies nomenclature appear less appropriate. The present data results support the view that *P. schwetzi* and *P. vivax* have many points of similarity. However, they do never join into one single separate cluster with any of the three methods. The first
fusion step occurs first between *P. schwetzi* and *P. gonderi* or between *P. vivax* and *P. cynomolgi* or both processes occur in parallel before *P. vivax* and *P. schwetzi* are allowed to be fused.

While the close relationship between *P. malariae* and *P. brasilianum* is widely acknowledged and has given rise to intesting evolutionnary hypotheses (Dunn, 1965; Livingstone, 1971), their specific level has never been questionned. The relationship between these two species appears clearly in the present results. They form a two-membered single cluster with all three classification methods, but the junction level occurs lower than for the berghei, vinckei or cynomolgi complexes.

For mammalian malaria parasites, Garnham (1966) has suggested a classification along there subgenera : *Laverania* (quoted above), *Vinckeia* (all non Primate parasites) and *Plasmodium* (all simian and human parasites, excepting *Laverania* species). Only in the AL method can these three subgenera be identified with separate clusters at some stage of the hierarchical agglomerative process (figure 3). The differences between the three methods is clear when the position of *P. (vinckeia) traguli* is considered in relation with other species of the same subgenus. With the SL and TL methods, *P. traguli* remains an outsider up to the final merging of all the mammalian parasites. In contrast, an homogeneous *Vinckeia* cluster persists for some time with the AL method. The relative position of simian parasitic species of the subgenus *Plasmodium* is greatly dependent on the classification method.

c) Influence of the method on the results

The influence of the method of classification on the nature is a disturbing feature of numerical taxonomy. The problem can be stated as follows : the raw data have been used, prior to the classification process, for the computation of dissimilarity coefficients. The classification itself organises these coefficients in a hierarchical taxonomic structure. By the use of a constraint on the nature of the clusters (overlapping is never allowed), some distortion is brought to the original coefficients. E.g. in SL, species with a large dissimilarity coefficient will be placed into one single cluster at a smaller dissimilarity level (at a higher level in the taxonomic tree) because an intermediate species with small dissimilarity coefficients with either species will be used as a connecting bridge (chaining effect). In TL, species with smaller dissimilarity coefficient will be held into separate clusters because other species included in the same cluster will increase the inter-cluster dissimilarity (reserve chaining effect). AL can be expected to combine the advantages (or disadvantages) of SL and TL. The overall distortion effect of the classification scheme is conveniently summarized by use of the classification operator C :

\[
C(d_{ij}) = J_{ij}
\]

where \(d_{ij}\) is the original dissimilarity coefficient, while \(J_{ij}\) is the dissimilarity level where species i and j enter into the same cluster. Sokal and Rohlf (1962) used the *cophenetic correlation* as a measure of resemblance between taxonomic structures. The same parameter can be used as a measure of distortion. Originally, the cophenetic correlation was defined as
the correlation coefficient between $d_{ij}$ and $J_{ij}$, where $J_{ij}$ was the rank of the junction level between species $i$ and $j$ on the taxonomic tree. This definition sprung off the particular structure of the classification program which sampled the tree structure at regularly spaced dissimilarity levels. In the present study, the actual junction levels $J_{ij}$ were used: $J_{ij}$ can be considered as the rounded values of $J_{ij}$. $C$ was computed for SL, TL and AL. We obtained respectively the values 0.888, 0.860 and 0.907. It seems thus that AL was the best method so far as distortion is used as the only criterion. However, one must be cautious when dealing with such measures of performance. Single linkage distortion can also be viewed as a profitable feature in an evolutionary perspective, where evolutionary links are of admitted importance.

**Conclusion**

Besides the already mentioned need for biochemical and immunological information, progresses in ultrastructure and finally progresses in knowledge of the pathology caused by each species of parasite, some remarks should be made concerning the choice of characters.

1. It can be seen in table II that the list of descriptive data used in this study is in fact incomplete. In order to increase the number of species studied to a level representative of the genus, very important features had to be rejected from the description. The most striking omissions are the lack of morphological description of any stage of exoerythrocytic schizonts and sporogony, although these are known to be as important as the blood stages for taxonomy.

2. Coding methods for some characters are unable to express very fine variations occurring in a broad scale of values. Such case is clearly happening for differences of less than 40 hours in the length of the prepatent period (char. 44) which reflects the length of exoerythrocytic schizogony. At the lower end of the scale, for rodent malaria parasites, much smaller differences seem nevertheless to be consistent. Logarithmic transformation of the scale cannot solve entirely the problem: such a scale should become too inaccurate in his higher part.

3. More information should also be welcomed for the distribution of most parasites in both vertebrate and invertebrate hosts as well as the susceptibility of a wide range of laboratory hosts for each described parasite.

A full collection of objective morphological and physiological characters should provide a solid basis which, introduced in a computer, would come out as numerical comparisons of dissimilarity between any given pair of parasites and automatically decide whether they belong or not to different subgenus, species or subspecies.

*Résumé — Le genre Plasmodium : Essais de taxonomie numérique.*

Trois méthodes ont été utilisées pour la construction d'une systématique numérique des parasites de la malaria. Les résultats sont en accord avec la nomenclature traditionnelle pour ce qui concerne le niveau subgénérique mais s'en écartent occasionnellement pour la définition des sous-espèces.
Samenvatting — Het genus Plasmodium : Pogingen tot numerische taxonomie.

Door drie verschillende methoden werd getracht een numerische taxonomie der malaria parasieten op te bouwen. Voor wat de subgenera betreft blijkt er geen belangrijk verschil te bestaan tussen de computeruitslagen en de klassieke gegevens. Voor de subspecies kwamen in sommige gevallen de computeruitslagen en de klassieke taxonomie niet overeen.

A. Bouckaert, Service de Physiologie, Faculté de médecine, Kinshasa XI, Zaïre.
Present address : Service d’épidémiologie, Ecole de Santé Publique, Avenue Chapelle-aux-champs 4, B-1200 Bruxelles, Belgium.
M. Wéry, Unité de Parasitologie I.M.T., Faculté de Médecine, Université Nationale du Zaïre, B. P. 747, Kinshasa XI, Zaïre Republic.
Received for publication on December 10, 1973.

REFERENCES

Bray, R. S. (1958) : Studies on malaria in Chimpanzees. II. The sporogonic cycle and mosquito transmission of Plasmodium vivax schwetzi. J. Parasitol. 44, 46.


