

## The Use of FISHER's Discriminant Function in the Taxonomy of Small Algae (*Oocystis* A. BRAUN)

Použití FISHEROVY diskriminační funkce v taxonomii drobnohledných řas (*Oocystis* A. BRAUN)

Pavel J a v o r n i c k ý and Helena Ř e h á k o v á

Hydrobiological Laboratory of Czechoslovak Academy of Sciences, Praha 5, Vltavská 17  
Botanical Institute of Charles University, Praha 2, Benátská 2

**Abstract** — A rich material of four species collected in Czechoslovakia was studied by ŘEHÁKOVÁ (in press). In *O. solitaria* WITTR., it is possible to distinguish two forms based only on the dimensions of the cells. In the three planktonic species, *O. parva* W. et G. S. WEST, *O. lacustris* CHOD. and *O. marssonii* LEMM., there are some morphological characteristics, besides differences in cell size, but they are not distinguishable throughout the entire ontogeny. FISHER's discriminant function enables us to unite several measurements (in our case only two: length and width of the cells) in one numerical value, the so-called discriminant  $X$ , and to minimize the frequency of misclassification. This can be used not only in taxonomic studies, but also in keys for determining algae.

In the taxonomy of small algae there exist two kinds of distinctive features, qualitative and quantitative ones. As qualitative features we consider for example a different sculpture of the membrane, the presence or absence of some cell organells such as chromatophore, pyrenoid and stigma, or the mode of reproduction. The quantitative features are suggested, when the cells or their parts (autospores, chromatophores, spines) are counted or measured. The qualitative characteristics are naturally separated from one another, even though there exist transgressions according to their variability. These transgressions, however, are typical for quantitative features. The characteristics of this kind are much used in the taxonomy of microscopic algae, as it is apparent from a number of names of varieties and forms, e. g. *parva*, *minor*, *minima*, *major*, *maxima*, frequently based only on the small differences in size. Some of them may be considered as ontogenetic stages, but there are also constant differences in the means, both within and among the species. Such differences in the means of the transgressive quantitative features may be used successfully in taxonomy of the small algae. Modern taxonomy, however, cannot be satisfied with the limiting values or means. It searches for methods of how to verify objectively the differences and to divide the material into groups (taxa) with the smallest possible frequency of misclassification. For this reason attempts have been made to use FISHER's discriminant function in the taxonomy of the genus *Oocystis* A. BRAUN (*Chlorococcales*).

### Method

FISHER's discriminant function combines the measurements of two or more taxonomical features, giving one numerical value, the so-called discriminant  $X$ . This value gives in most cases more information about taxonomical differences than any of the features used apart. This statistical method may be variously used, but it was calculated for the use in botanical taxonomy:

FISHER (1936, 1940) demonstrated it on some species of the genus *Iris* L. We have no evidence whether it was used practically, but up to the present it has not been applied to taxonomical problems in phycology. Perhaps there exists a more acceptable or simple method; any suggestions about this are appreciated. Since besides the works named above, FISHER's discriminant analysis is explained very clearly in the handbook by MATHER (1943), the details of the method are here presented only in Czech.

The value of the discriminant  $X$ , when two measurements  $x_a$  and  $x_b$  are available, may be obtained from the expression

$$X = L_a x_a + L_b x_b \quad (1)$$

The linear equations of estimation of the coefficients  $L_a$  and  $L_b$  are

$$\begin{aligned} S_{aa}L_a + S_{ab}L_b &= d_a \\ S_{ba}L_a + S_{bb}L_b &= d_b \end{aligned} \quad (2)$$

where  $d$  is the mean difference between the species,  $S$  is the sum of the squared ( $aa$ ,  $bb$ ) or multiplied ( $ab$ ,  $ba$ ) deviations from the means within the species.

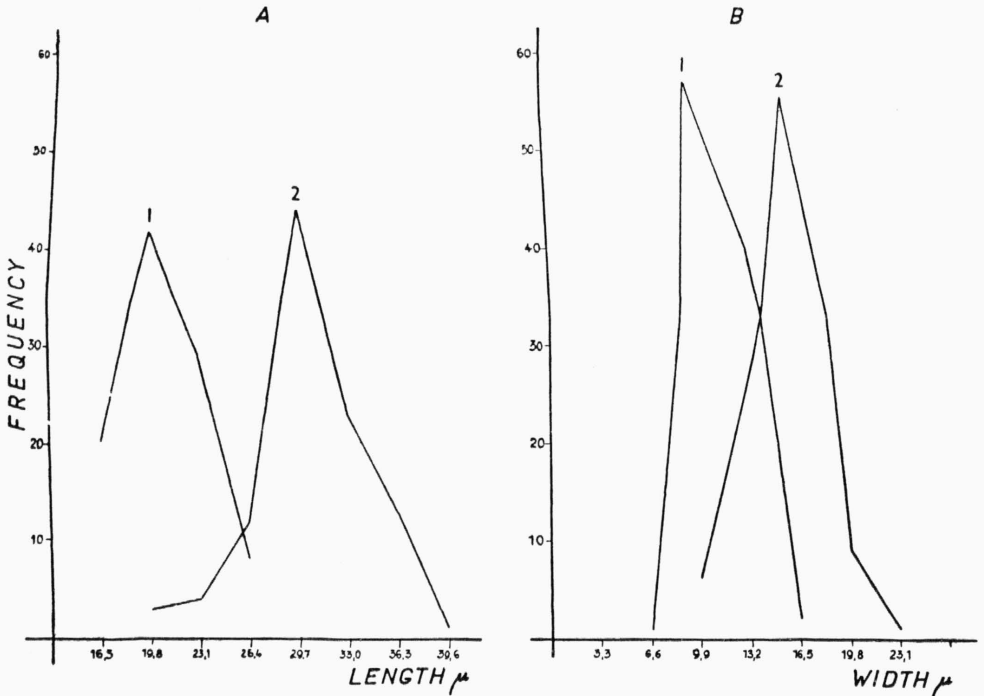


Fig. 1. — *Oocystis solitaria* WITTR. f. *solitaria* (1) and *O. solitaria* WITTR. f. *major* WILLE (2). A — distribution of cell lengths, B — distribution of cell widths.

Fig. 1. — *Oocystis solitaria* WITTR. f. *solitaria* (1) a *O. solitaria* WITTR. f. *major* WILLE (2). A — rozdělení buněčných délek, B — rozdělení buněčných šířek.

The analysis of variance of  $X$  enables us to test the significance of the specific difference. The critical value  $C$  for the classification into groups (taxa) I an II is

$$C = \bar{X}_I + \frac{D}{2} = \bar{X}_{II} - \frac{D}{2} \quad (3)$$

where  $\bar{X}_I$  and  $\bar{X}_{II}$  are the mean values of the discriminant calculated for the both groups, and  $D = \bar{X}_{II} - \bar{X}_I$ .

The frequency of misclassification can be found by the use of the *t*-test. This allows us also to verify, whether the function *X* is really more discriminative than any of the involved measurements used apart.

## Material

In the present study we worked only with natural populations of algae collected in Czechoslovakia. Algae were measured alive immediately after collection.

To prove the homogeneity of the littoral species *Oocystis solitaria* WITTR. (Tab. V.: a, b), 200 individuals from 20 various localities were measured. Two groups of different size were found, which, however, have never grown together in the same peat bog or pool. They occurred at small distances without special

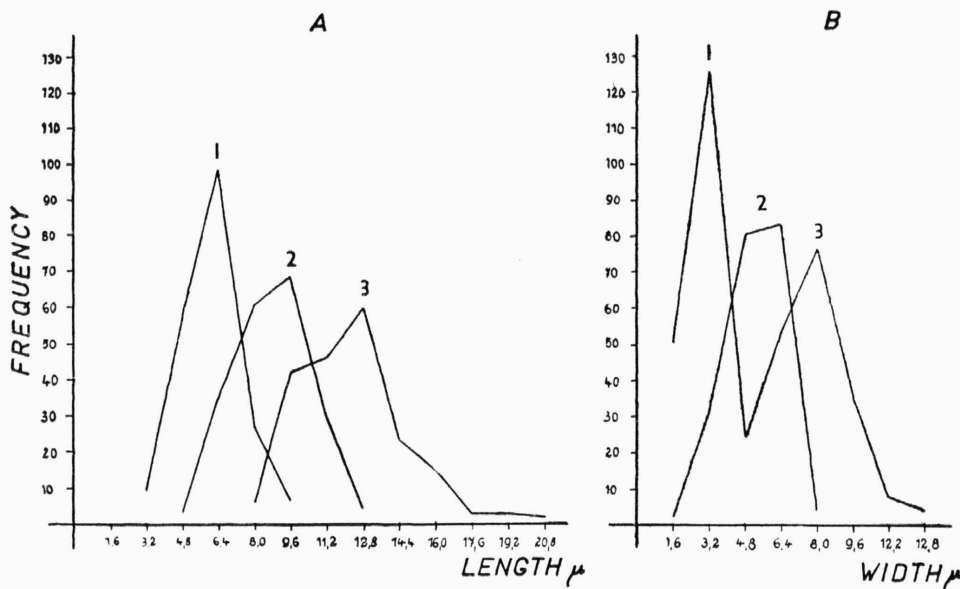


Fig. 2. — *Oocystis parva* W. et G. S. WEST (1), *O. lacustris* CHOD. (2) and *O. marssonii* LEMM. (3). A — distribution of cell lengths, B — distribution of cell widths.

Fig. 2. — *Oocystis parva* W. et G. S. WEST (1), *O. lacustris* CHOD. (2) a *O. marssonii* LEMM. (3). A — rozdělení buněčných délek, B — rozdělení buněčných šířek.

demands on the character of their station or on the *pH* of water. Between them there are differences both in the length and width of the cells (Fig. 1A, B), but none in the cell's morphology.

Three planktonic species were investigated, i.e. *Oocystis parva* W. et G. S. WEST, *O. lacustris* CHOD. and *O. marssonii* LEMM. (Tab. V.: c–f), and 300 individuals (200 of each species) from 25 various localities were measured (Fig. 2A, B). There are morphological differences among these species, besides differences in length and width of the cells; the latter can be used for onto-

genetic stages with indistinguishable main morphological features, i.e. for solitary cells with several chromatophores.

## Results

### *Oocystis solitaria* WITTR.

When the symbol for the values of cell length is  $a$  and of cell width  $b$ , that for the group of minor size  $I$  and of major size  $II$ , the discriminant function was calculated as follows in the table 1.

Table 1

$L_a = 0.003294$	$\bar{X}_I = 0.1008$	$D = 0.0494$
$L_b = 0.002927$	$\bar{X}_{II} = 0.1502$	$C = 0.1255$

If we classify into groups, the values of discriminant  $X$  for each individual are to be calculated from expression (1). The individuals with  $X$  less than the critical value 0.1255 will be attributed to taxon  $I$ , those with a higher  $X$  to taxon  $II$ . The group of larger dimensions was identified with *O. solitaria* WITTR. *f. major* WILLE, the less form being the typical one, i.e. *O. solitaria* WITTR. *f. solitaria*. Forma *major* was isolated in culture and the mean dimensions were proved to be constant (ŘEHÁKOVÁ in press).

When the analysis of variance of  $X$  is calculated and the table of variance ratios consulted, the probability is found:  $F_{2; 197} (0.005) < 5.298 \ll 243$ . There is a good difference in the value of our discriminant  $X$ .

The frequency of misclassification was tested according to the expression

$$t_{(f)} = \frac{\frac{D}{2}}{\sqrt{\frac{S}{f}}}$$

(degrees of freedom  $f$  and the mean square  $\frac{S}{f}$  within groups). For  $t_{(197)} = 1.5633$  the tabulated probability is  $0.25 > P > 0.10$ . Should misclassification occur, the deviation must be in one given direction; it may be, therefore, only found in 5–12,5% of cases. It actually occurred in 7,5% of the cases.

A similar analysis may be made on the data of length and width separately according to the expression

$$t_{(n_1+n_2-2)} = \frac{d}{s_x} ;$$

for  $d$  see in equations (2). It represents for length  $t_{(198)} = 1.4543$  and for width  $t_{(198)} = 0.1249$ . The discriminant enjoys an advantage over any of the measurements used apart:  $t_{(197)} = 1.5633$ .

When the above symbols are used, the values in Tab. 2 for the discrimination of three planktonic species were calculated.

Table 2

	Combination		
	$\alpha$ <i>O. parva</i> <i>O. lacustris</i>	$\beta$ <i>O. lacustris</i> <i>O. marssonii</i>	$\gamma$ <i>O. parva</i> <i>O. marssonii</i>
$L_a$	0.000463	0.000758	0.001732
$L_b$	0.004045	0.002093	0.004055
$\bar{X}_I$	0.014928	0.017636	0.022742
$\bar{X}_{II}$	0.025297	0.026649	0.054862
$D$	0.010369	0.009013	0.032120
$C$	0.020112	0.022142	0.038802

The analysis of variance proves the specific differences to be significant with great probability:

$$F_{2; 397}(0.005) < 5.298 \ll 207 (\alpha); \quad 179 (\beta); \quad 638 (\gamma).$$

The probable and actual frequencies of misclassification are compared in the table 3.

Table 3

Combination	$t_{(397)}$	Frequency of misclassification in %	
		expected	found
$\alpha$	1.0167	12.5—25	15.5
$\beta$	0.9395	12.5—25	17.7
$\gamma$	1.7854	2.5—5	6.0

$t$  — values for the two measurements calculated apart are in table 4.

Table 4

Combination	$t_{(398)}$	
	length	width
$\alpha$	0.8683	1.0107
$\beta$	0.8449	0.9272
$\gamma$	1.5743	1.7254

In this case, the discriminant  $X$  (compare  $t$  — values in table 3) is also somewhat more discriminative than any of the measurements treated apart, but enjoys only little advantage over the width used alone.

The size groups are in good coincidence with the morphological differences among the three planktonic species, i.e. with the morphology of mother membrane and with the number of chromatophores in young daughter cells (Tab. V: c-f).

## Key to the species<sup>1)</sup>

(Giving the arbitrary values of the discriminant function,  $L$  stands for length,  $W$  for the width of cells, both expressed in millimetres.)

- 1a Chromatophores in autospores 1-2, in adult vegetative cells 1-4; planktonic species.
  - 2a Autospores and released daughter cells with one chromatophore.
    - 3a Daughter cells released by gelatinizing of the mother membrane; dimensions of cells: length 4.8-12.8  $\mu$ , width 1.6-8.0  $\mu$ .  
Respecting *O. marssonii*  $L \cdot 0.76 + W \cdot 2.09 < 0.0221$ ; respecting *O. parva*  $L \cdot 0.46 + W \cdot 4.04 > 0.0201$  . . . . . *O. lacustris* CHOD. (Tab. V: f)
    - 3b Daughter cells released by bursting of the mother membrane, which is not gelatinized; dimensions of cells: length 3.2-9.6  $\mu$ , width 1.6-4.8  $\mu$ .  
Respecting *O. marssonii*  $L \cdot 1.73 + W \cdot 4.05 < 0.0388$ ; respecting *O. lacustris*  $L \cdot 0.46 + W \cdot 4.04 > 0.0201$  . . . . . *O. parva* W. et G. S. WEST (Tab. V: d, e)
  - 2b Autospores with 1-2 chromatophores; daughter cells with 2-4 chromatophores released by bursting of the mother membrane; adult vegetative cells with 2-4 chromatophores; dimensions of cells: length 8.0-20.8  $\mu$ , width 4.8-14.4  $\mu$ .  
Respecting *O. parva*  $L \cdot 1.73 + W \cdot 4.05 > 0.0388$ ; respecting *O. lacustris*  $L \cdot 0.76 + W \cdot 2.09 > 0.0221$  . . . . . *O. marssonii* LEMM. (Tab. V: c)
- 1b Chromatophores in autospores at least 4, in adult vegetative cells numerous (to 30 or more); littoral species. . . . . *O. solitaria* WITTR. (Tab. V: a, b)
  - 4a Dimensions of cells: length 13.2-28.0  $\mu$ , width 5.0-18.1  $\mu$ .  
 $L \cdot 3.29 + W \cdot 2.93 < 0.1255$  . . . . . *O. solitaria* WITTR. f. *solitaria*
  - 4b Dimensions of cells: length 16.5-52.8  $\mu$ , width 8.2-29.7  $\mu$ .  
 $L \cdot 3.29 + W \cdot 2.93 > 0.1255$  . . . . . *O. solitaria* WITTR. f. *major* WILLE

## Acknowledgement

The authors are much obliged to Ing. Miloš Jílek CSc. from the Institute of Microbiology of the Czechoslovak Academy of Sciences who recommended the statistical method, offered valuable help during its study and kindly revised the manuscript.

## Souhrn

Kvantitativní znaky, kterých se často používá k rozlišování forem, variet i druhů drobnohlédných řas, vytvářejí zpravidla plynulé přechody. Někdy se však jedná o skutečné rozdíly velikosti, což nám ukáží průměry nebo lépe diagramy rozdělení hodnot. Tak ŘEHÁKOVÁ ve své práci o rodu *Oocystis* A. BRAUN (v tisku) shromáždila velký materiál o rozměrech druhu *O. solitaria* WITTR., který se zřetelně rozpadá na dvě velikostní formy (Tab. V: a, b; Fig. 1A, B), a o třech planktonních druzích *O. parva* W. et G. S. WEST, *O. lacustris* CHOD. a *O. marssonii* LEMM. Ty se sice navzájem liší také morfologickými znaky (Tab. V: c-f), ale v určitém ontogenetickém stadiu (solitární buňky s více chromatofory) jsme i u nich odkázáni pouze na rozdíly velikosti (Fig. 2A, B). V této práci jsme se pokusili aplikovat na tento materiál FISHEROVU diskriminační analýzu (FISHER 1936, 1940; MATHER 1943). Je to statistická metoda vypracovaná přímo pro botanickou taxonomii, která umožňuje: 1. shrnout rozdíly měření dvou nebo více kvantitativních znaků do jedné číselné hodnoty, tzv. diskriminantu  $X$ ; 2. objektivně testovat na základě analýsy rozptylu  $X$  rozdílnost skupin (taxonů); 3. omezit počet špatně zařazených jedinců na minimum a předvídat pravděpodobnost chybné klasifikace. Ideální případ použití této metody je tam, kde máme 3 nebo více znaků, které nejsou v těsné korelaci. Protože v žádné české statistické příručce není tato metoda uvedena, považujeme za vhodné předložit zde podrobnější pracovní návod.

Pro měření třech kvantitativních znaků  $a$ ,  $b$ ,  $c$  u dvou skupin (taxonů)  $I$  a  $II$ , vypočteme průměrnou hodnotu diskriminantu  $X$  pro každou skupinu zvlášť ze vztahu

$$\bar{X} = L_a \bar{x}_a + L_b \bar{x}_b + L_c \bar{x}_c \quad (1)$$

<sup>1)</sup> The key according to ŘEHÁKOVÁ (in press) completed by the data of discriminant analysis.

Koeficienty  $L$  vypočteme ze soustavy lineárních rovnic

$$S_{aa}L_a + S_{ab}L_b + S_{ac}L_c = d_a$$

$$S_{ba}L_a + S_{bb}L_b + S_{bc}L_c = d_b$$

$$S_{ca}L_a + S_{cb}L_b + S_{cc}L_c = d_c$$

2

kde  $d$  jsou rozdíly průměrů každého znaku v obou skupinách (např.  $d_a = \bar{x}_{aII} - \bar{x}_{aI}$ ) a  $S$  součty čtverců ( $aa, bb, cc$ ) nebo součinů ( $ab, ac, bc$ ) odchylek od průměrů uvnitř skupin. To znamená, že odečítáme od každé jednotlivé hodnoty aritmetický průměr toho znaku v té skupině ( $I$  nebo  $II$ ). Samozřejmě vyjdou i hodnoty záporné, jejichž čtverce jsou sice kladné, ale jejich součiny mohou být rovněž záporné. Hodnoty, vypočtené pro každý znak ( $a, b, c$ ), pak sečteme z obou skupin ( $I, II$ ) dohromady. V rozepsaném tvaru je tedy první rovnice soustavy (2)

$$S(x_a - \bar{x}_a)^2 L_a + S(x_a - \bar{x}_a)(x_b - \bar{x}_b) L_b + S(x_a - \bar{x}_a)(x_c - \bar{x}_c) L_c = d_a$$

Při velkém počtu naměřených hodnot si usnadníme práci jejich grupováním: rozdělíme materiál na několik tříd s určitým rozmezím znaku. S průměrnými hodnotami znaku v každé třídě pak pracujeme jako s jednotlivými hodnotami. Zásady grupování nalezneme v každé statistické příručce (MACEK a POKORNÝ 1956). Protože však v našem případě musíme při vytváření součinů odchylek hodnoty znaků párovat tak, jak se vyskytovaly spojené v materiálu, použijeme pro grupování dvojsoých tabulek. Tak např. pro délku a šířku buněk *Oocystis lacustris* CHOD. byla sestavena tabulka č. 5, ve které  $x_i$  je střední hodnota znaku v každé třídě,  $f$  četnost v třídách. Uvnitř tabulky jsou vepsány četnosti v kombinovaných třídách šířky a délky.

Tabulka 5

Š í ř k y

	$x_i$	1	2	3	4	5	
$x_i$	$x_i - \bar{x}$	-2,28	-1,28	-0,28	0,72	1,72	$f$
3	-2,48	2	1				3
4	-1,48		27	8			35
5	-0,48		3	55	3		61
6	0,52			16	52		68
7	1,52			1	26	2	29
8	2,52				2	2	4
	$f$	2	31	80	83	4	200

Kritická hodnota diskriminantu mezi oběma skupinami (taxony) je rovna

$$C = \bar{X}_I + \frac{D}{2} = \bar{X}_{II} - \frac{D}{2}, \quad (3)$$

kde  $D = \bar{X}_{II} - \bar{X}_I$ .

Třídění jedinců pak provádíme tak, že pro každého jedince vypočteme individuální hodnotu

diskriminantu  $X$  dosazením jeho rozměrů místo průměrů do vztahu (1). Hodnoty koeficientů  $L$  jsou již známy. Jedinci s hodnotou diskriminantu menší než je kritická hodnota  $C$  patří ke skupině  $I$ , jedinci s diskriminantem větším než  $C$  ke skupině  $II$ .

Hodnoty vypočtené pro uvedené taxony rodu *Oocystis* jsou v anglickém textu na tabulkách č. 1 a 2 a jsou také použity v určovacím klíči.

Rozdílnost obou skupin (taxonů) testujeme na základě analýsy rozptylu  $X$  podle schématu na tabulce č. 6.

Tabulka 6

	Součet čtverců $S$	Stupně volnosti $f$	Průměrný čtverec	Variační poměr $F$
Mezi skupinami	$\frac{n}{2} D^2$	$k$	$\frac{S}{f}$	$\frac{S}{f}$
Uvnitř skupin	$D$	$2n - k - 1$	$\frac{S'}{f'}$	$\frac{S'}{f'}$

(Hodnoty označené čárkou se rozumějí „uvnitř skupin“.)

V tabulce 6 znamená  $n$  počet proměřených jedinců v jedné skupině (má být stejný i ve druhé),  $k$  počet měřených znaků, pro  $D$  viz vztah (3). S dvojnásobnou stupni volnosti vstupujeme do tabulky kritických hodnot  $F$  (JANKO 1958, tab. 12). Pro zvolenou hladinu významnosti (např. 0,005, tj. pro pravděpodobnost 99,5 %) zde nalezneme tabelovanou hodnotu  $F$ . Jestliže je naše vypočtená hodnota větší, zamítá se na dané hladině významnosti hypotéza o totožnosti obou skupin (taxonů).

Abychom odhadli, v jakém procentu případů se dopustíme užitím diskriminantu  $X$  chybného zařazení, což je při použití transgresivních znaků nevyhnutelné, testujeme  $t$ -testem podle vztahu

$$t_{(2n-k-1)} = \frac{\frac{D}{z}}{\sqrt{\frac{S'}{f'}}$$

V tabulce kritických hodnot rozdělení  $t$  (JANKO 1958, tab. 8) nalezneme s příslušnými stupni volnosti  $(2n - k - 1)$  pravděpodobnost pro hodnotu nejbližší nižší a nejbližší vyšší než je vypočtená hodnota  $t$ . Tak pro klasifikaci *O. solitaria* WITTR. podle  $X$  jsme našli hodnotu  $t_{(197)} = 1,5633$ . Počet stupňů volnosti je větší než 120 a pro něj nejbližším tabelovaným hodnotám  $t$  odpovídá pravděpodobnost  $0,25 > P > 0,10$ . Protože však ke špatnému zařazení může dojít jenom v jednom směru, očekáváme je jenom v polovině, tj. asi v 5–12,5 % případů. Skutečně k němu došlo v 7,5 % případů.

Chceme-li ověřit, zda je použití diskriminantu  $X$  výhodnější než použití jednotlivých znaků, můžeme znaky testovat odděleně podle vztahu (pro znak  $a$ )

$$t_{(2n-2)} = \frac{\frac{d_a}{2}}{\sqrt{\frac{S_{aa}}{2n-2}}}$$

Nejvýhodnější je to kritérium, pro které vychází největší hodnota  $t$  a tedy nejmenší pravděpodobnost chybné klasifikace. Porovnání těchto hodnot viz v anglickém textu v tabulkách č. 3 a 4.

## References

- FISHER R. A. (1936): The use of multiple measurements in taxonomic problems. — Ann. Eugen. 7, 2 : 179–188.  
 — (1940): The precision of discriminant functions. — Ann. Eugen. 10, 4 : 422–429.  
 JANKO J. (1958): Statistické tabulky. — Praha: 251 p.  
 MACEK J. et POKORNÝ V. (1956): Zakládání a hodnocení biologických pokusů. — Skriptum, Praha: 228 p.



MATHER K. (1943): Statistical analysis in biology. — London: 247 p.

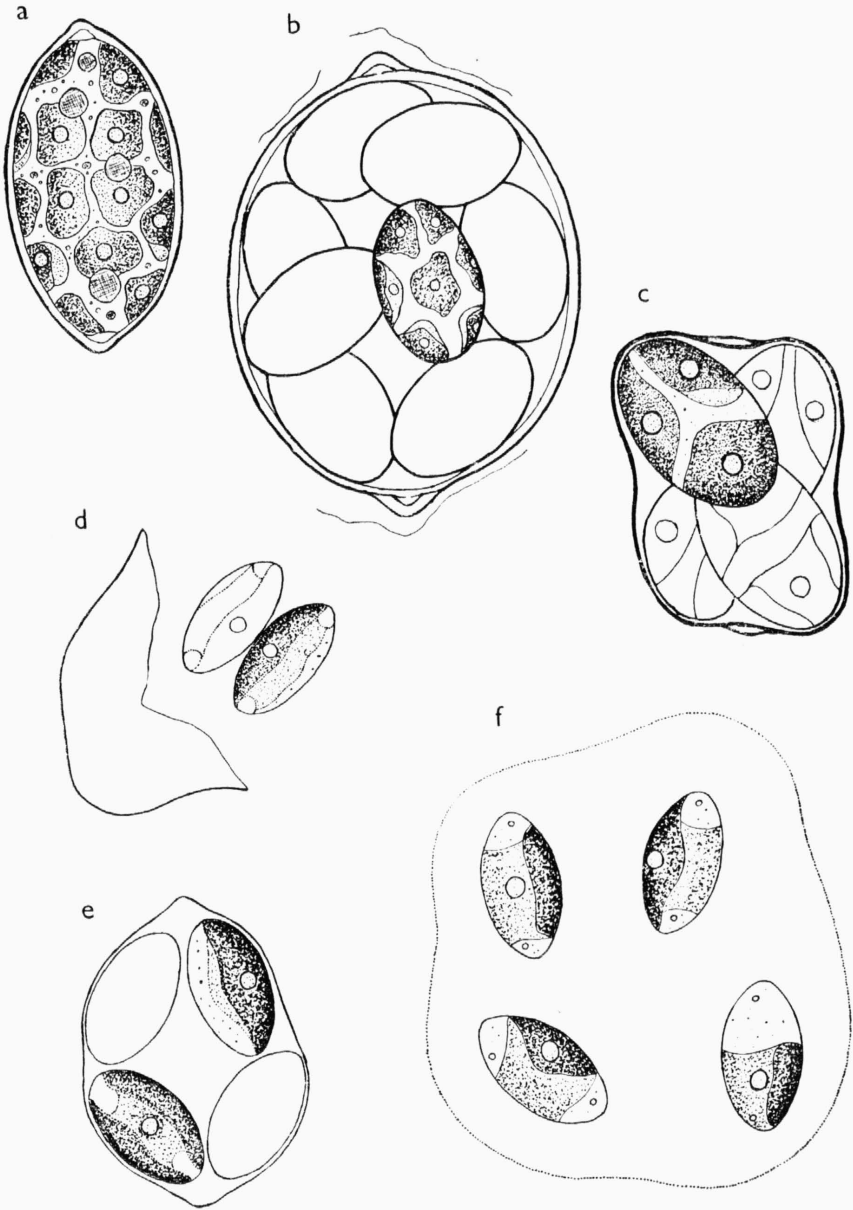
ŘEHÁKOVÁ H.: Die Variabilität der Arten der Gattung *Oocystis* A. BRAUN in der Tschechoslowakei. In press.

Explanations of the plate V — Vysvětlivky k tabuli V

Tab. V: *a* — *Oocystis solitaria* WITTR., vegetative cell. *b* — *O. solitaria*, autospores inside the mother membrane. *c* — *O. marssonii* LEMM. *d* — *O. parva* W. et G. S. WEST, release of daughter cells by bursting of the mother membrane. *e* — *O. parva*, a colony of young cells. *f* — *O. lacustris* CHOD., a colony with mucous mother membrane.

Tab. V: *a* — *Oocystis solitaria* WITTR., vegetativní buňka. *b* — *O. solitaria*, autospory uvnitř mateřské membrány. *c* — *O. marssonii* LEMM. *d* — *O. parva* W. et G. S. WEST, uvolňování dceřinných buněk prasknutím mateřské membrány. *e* — *O. parva*, kolonie mladých buněk. *f* — *O. lacustris* CHOD., kolonie se slizovitou mateřskou membránou.

(Orig. H. ŘEHÁKOVÁ)



P Javornický and H. Řeháková: The Use of Fisher's Discriminant Function in the Taxonomy of Small Algae (*Oocystis* A. Braun)