THE POWER OF MULTIVARIATE STATISTICAL METHODS IN THE TAXONOMY OF PTEROMALIDAE (HYMENOPTERA: CHALCIDOIDEA)

Hannes BAUR

Department of Invertebrates, Natural History Museum, Bernastrasse 15, Bern, CH-3005 Switzerland (e-mail: hannes.baur@nmbe.unibe.ch)

Abstract – The present paper illustrates the use of multivariate statistical methods for separating species in Pteromalidae (Hymenoptera). Linear discriminant (DA) and principal components analysis (PCA) are applied to polythetic species of the *Pteromalus albipennis* group, i.e. to *P. albipennis, P. decipiens*, and *P. solidaginis*. A DA including 22 characters reveals clear gaps among those species and eliminates redundant characters. With PCA *P. solidaginis* differs from the other species by its shape or shape and size, whereas *P. decipiens* is separated from *P. albipennis* mainly by its smaller size. Possible implications of the results are discussed with regard to the status of the taxa.

Key words: principal components, linear discriminant functions, multivariate statistics, taxonomy, *Pteromalus*

Introduction

In Pteromalidae complexes of polythetic species are frequently encountered. This situation poses difficulties for the taxonomist, since such species cannot be separated by a single diagnostic character. Nevertheless, there is often apparent variation in proportions of many characters. Multivariate statistical methods could therefore be used to examine the variation in these structures and to uncover hidden differences between taxa. Studies that make explicit use of such statistical methods are rather rare in the taxonomic literature on Pteromalidae. Janzon (1986) investigating a subgroup of the *Pteromalus albipennis* species-group pointed out the effect of allometry on ratios used in taxonomy. However, the data were explored in a bivariate context, multivariate statistical methods were not applied.

The *Pteromalus albipennis* species-group was erected by Graham (1969) who included 23 Western European species. The group is certainly more widely distributed, since numerous species are known to the author from the Nearctic region (Baur, *unpubl.*). There is strong evidence that most species are parasitoids of fruit flies (Diptera: Tephritidae) on Asteraceae (Graham 1969; Janzon 1984; Baur, *unpubl.*). Despite investigations by Graham (1969) and Janzon (1984, 1986) on the *Pteromalus albipennis* species-group, there are still many unresolved problems. One of them concerns *P. solidaginis* Graham & Gijswijt 1991 (*sol*) and the very similar *P. albipennis* Walker, 1835 (*alb*) and *P. decipiens* Graham, 1969 (*dec*). Graham (1969) and Graham & Gijswijt (1991) mentioned qualitative characters, such as the colour of the flagellum, the veins and the body, but also certain ratios for the separation of these species. A re-examination of the type series and further material revealed that these characters were not reliable for a significant part of the material. Colour characters are difficult to interpret and some of the ratios showed considerable overlap between species. In an attempt to find additional evidence for separating these species, they were included in a multivariate morphometric analysis.

Materials and Methods

The study is based on an analysis of 66 dry mounted females (Appendix). These had to be assigned to a species before the analysis. Specimens of *alb* and *dec* were identified according to Graham (1969: 502) and compared with the lectotype (*alb*) or with paratypes (*dec*). For *sol* only paratypes were available. Individual specimens were provided with an identification label and a number to assure unambiguous recognition. Terminology and morphology follow Gibson (1997). 22 characters were selected for the analysis (Table 1). Measurements were made under a Leica MZ12 stereo-microscope with different magnifications using a calibrated eye-piece micrometer (12 mm subdivided into 120 units) (Table 1). To avoid additional variability resulting from possible fluctuating asymmetry, only the left hand side of a specimen was considered.

The morphometric analysis of the data followed a two-step procedure. A first step served to find the best separation between taxa (e.g. between *sol* and the other two species) and to identify redundant characters. Here, a linear discriminant analysis (DA) was the most appropriate method. In a second step the nature of the differences (e.g. size or shape differences) found with the reduced data set was investigated using a principal components analysis (PCA). The DA, as used here, is based on multiple regression and its application is discussed in detail by Flury & Riedwyl (1988). The procedure requires an additional code variable for the taxa which is treated as the ependent variable of the analysis. Thus an arbitrary number ("0" and "1") was assigned to each group. The characters, on the other hand, are considered the independent variables of the regression analysis. In an iterative process, redundant characters are identified and eliminated: First, a regression analysis including all independent variables (characters) is computed. Second, the variable showing the highest probability of F, i.e. contributing the least to the separation of taxa, is removed and a new regression analysis is run on the reduced data set. These steps are repeated until the probability of F is < 0.1 for the remaining variables. This so-called backward elimination method is fully implemented in the SPSS statistical software package. Thus, the best characters for separating the taxa can be identified very effectively. Finally, a PCA computed with this set of characters further reduces the dimensionality in the data and illuminates the nature of the differences between taxa. The PCA presented below was carried out on a correlation matrix. For further details concerning the computation and interpretation of this fundamental ordination technique I refer to Flury & Riedwyl (1988), Pimentel (1992), Manly (1994), and Podani (2000). Calculation of DA, PCA and the statistics presented below was done with SPSS for Windows (Release 11.0, SPSS Inc., 2001).

Results

Median and range of each character and species are given in table 1. From these Figures eight ratios, commonly used in pteromalid taxonomy (Graham 1969; Bouèek 1988), were calculated (Fig. 1). These ratios show that *sol* is well separated from the two other species using POL/OOL and gaster L/B. Clearly, the ranges of several other ratios overlap, e.g. of eye H/B and propodeum L/B. Generally, *sol* seems to be closer to *dec*. On the other hand, *dec* and *alb* cannot really be diagnosed by any of those ratios, despite the fact, that sometimes statistically significant differences may occur (e.g. in head B/H). But the medians indicate, that *dec* is distinctly smaller than *alb*. With regard to these figures, some differences in size *and* form are identified between the species.

Table 1 Median, minimum, and maximum (in mm) of 22 characters, listed by species. Delimitation of characters mostly follows Janzon (1986), points of references were equidistant from the objective of the microscope. "Mag." is the magnification used for measurements

Character	Mag.	Spec	Med.	Min.	Max.	Character	Mag.	Spec	Med.	Min.	Max.
head B	80x	*alb*	0.903	0.763	1.094	pronotal collar	160x	*alb*	0.097	0.063	0.134
(Janzon 1986)	80x	*dec*	0.650	0.581	0.719	(Janzon 1986)	160x	*dec*	0.059	0.050	0.075
	80x	*sol*	0.856	0.794	0.888		160x	*sol*	0.088	0.075	0.100
head H (clypeus-	80x	*alb*	0.675	0.600	0.800	marginal vein	80x	*alb*	0.425	0.350	0.513
lower edge	80x	*dec*	0.500	0.438	0.569	(Janzon 1986)	80x	*dec*	0.325	0.281	0.363
median ocellus)	80x	*sol*	0.663	0.600	0.700		80x	*sol*	0.438	0.375	0.463
upper face (lower	80x	*alb*	0.388	0.344	0.463	postmarginal vein	80x	*alb*	0.413	0.363	0.488
edge toruli-lower	80x	*dec*	0.294	0.256	0.331	(Janzon 1986)	80x	*dec*	0.319	0.275	0.344
edge median	80x	*sol*	0.388	0.363	0.413		80x	*sol*	0.432	0.350	0.488
ocellus)											
POL	160x	*alb*	0.236	0.194	0.266	stigmal vein	80x	*alb*	0.313	0.263	0.350
(Janzon 1986)	160x	*dec*	0.181	0.156	0.191	(Janzon 1986)	80x	*dec*	0.231	0.206	0.250
	160x	*sol*	0.219	0.197	0.238		80x	*sol*	0.300	0.275	0.325
OOL	160x	*alb*	0.133	0.119	0.169	metatibia	80x	*alb*	0.813	0.694	1.013
(Janzon 1986)	160x	*dec*	0.100	0.091	0.109	(length of tibia)	80x	*dec*	0.588	0.513	0.638
	160x	*sol*	0.150	0.141	0.163		80x	*sol*	0.763	0.688	0.819
eye H	80x	*alb*	0.428	0.375	0.500	propodeum L	160x	*alb*	0.166	0.125	0.213
(Janzon 1986)	80x	*dec*	0.313	0.281	0.350	(median area	160x	*dec*	0.125	0.100	0.144
	80x	*sol*	0.400	0.369	0.425	length, Janzon	160x	*sol*	0.181	0.159	0.200
						1986)					
eye B	80x	*alb*	0.294	0.263	0.338	propodeum B	80x	*alb*	0.394	0.319	0.513
(Janzon 1986)	80x	*dec*	0.225	0.200	0.244	(median area	80x	*dec*	0.263	0.244	0.313
	80x	*sol*	0.263	0.250	0.275	breadth, Janzon	80x	*sol*	0.375	0.338	0.425
						1986)					
pedicel L	160x	*alb*	0.094	0.081	0.106	gaster L	32x	*alb*	1.789	1.547	2.266
(length of pedicel	160x	*dec*	0.075	0.069	0.081	(Janzon 1986)	32x	*dec*	1.266	1.094	1.406
in profile)	160x	*sol*	0.091	0.084	0.097		32x	*sol*	1.469	1.203	1.625
funicle1 L	160x	*alb*	0.106	0.088	0.125	gaster B	80x	*alb*	0.713	0.563	0.975
(length of first	160x	*dec*	0.066	0.050	0.075	(Janzon 1986)	80x	*dec*	0.500	0.413	0.563
funicular segment	160x	*sol*	0.091	0.075	0.100		80x	*sol*	0.800	0.675	0.925
in profile)											
scape	80x	*alb*	0.338	0.300	0.413	tergum 7 L (length	80x	*alb*	0.288	0.231	0.388
(Janzon 1986)	80x	*dec*	0.250	0.219	0.288	of seventh gastral	80x	*dec*	0.200	0.150	0.263
	80x	*sol*	0.313	0.281	0.338	tergum, Janzon	80x	*sol*	0.225	0.172	0.238
						1986)					
malar space	160x	*alb*	0.225	0.188	0.275	tergum 7 B	80x	*alb*	0.278	0.231	0.325
(Janzon 1986)	160x	*dec*	0.156	0.138	0.175	(breadth of	80x	*dec*	0.206	0.169	0.231
	160x	*sol*	0.200	0.175	0.222	seventh gastral	80x	*sol*	0.300	0.213	0.344
						tergum, Janzon					
						1986)					

Moreover, there may also be much redundancy in the data, as all characters are strongly intercorrelated (Spearman's rho, $p < 0.01$ in all instances). Here, multivariate statistics offer convenient methods for a straightforward exploration of the data.

Table 2 lists the discriminant functions found in several different DAs of comparisons of species or groups of species. In a first analysis *sol* was compared with the other two species together but subsequently only single species were included for pairwise comparisons. In each of the four analyses a large number of variables were eliminated and only 9–11 characters were retained for the final model. The power of the discriminant function was always higher in the reduced than in the full model. Generally, the groups are widely separated for values of those discriminant functions (Fig. 2).

Table 2 Coefficients of different linear discriminant functions found with DA

	rest-*sol*	***alb-sol***	***dec-sol***	***alb-dec***		**rest-*sol***	***alb-sol***	***dec-sol***	***alb-dec***
head B	–	–	3.959	–	malar space	–	–	–	7.597
head H	7.078	–	8.462	–	marginal vein	–	–	1.633	–
upper face	–5.806	–4.365	–9.148	–	postmarginal vein	1.700	2.317	–	–
POL	–3.259	–	–9.950	–	stigmal vein	–	–2.474	–	–4.598
OOL	22.354	21.926	14.777	–	metatibia	–	–	–	3.873
eye H	–	3.859	–	–	propodeum B	–	–	–5.001	4.438
eye B	–	-7.366	–	–	gaster L	–0.946	-0.493	–	-1.334
pedicel L	–11.569	–11.530	22.605	–	gaster B	0.485	0.711	–	–
funicle1 L	–8.562	–	–13.512	–17.884	tergum7 L	–	–2.269	–1.996	4.212
scape	–5.466	–	–4.852	–11.115	tergum7 B	–	–	–	–6.094

With the DA the species can be very well separated using a subset of the characters only. However, it may be useful to know, whether the differences of the species are related to size or rather to form. To address this question, a PCA was computed using the ten characters retained in the first DA (rest-*sol*, Table 2). This PCA was very effective in further reducing the dimensionality of the data, as the first and second principal components comprise about 93% of the total variance. Hence, the remaining components can safely be neglected. The scatterplot of all 66 specimens against the values of the first two components (Fig. 3) shows that the species cluster tightly together. There is no overlap between the different species at all. The first component is strongly correlated with each of the original variables, that is its correlation coefficients ('loadings') are uniform (all positive) and very high (Table 3). It may therefore be considered as a 'size' vector (Manly 1994) rather than an 'allometry' vector (see Shea 1985). In other words, variation related to size is summarised in this axis. The second component, on the other hand, shows some high – positive or negative – correlation with only some of the original variables, e.g. OOL, gaster L, and gaster B (Table 3). Thus, variation related to shape differences is summarised in the second

component. With this rather rough but nevertheless important interpretation of the components in mind, it is evident that *sol* is clearly separated from *alb* by differences in shape and from *dec* by shape and size (Fig. 3). Contrary to this, *alb* differs from *dec* mainly in size.

Table 3 Correlation coefficients of principal components with original variables

Character	Component	
	1	2
head H	0.9937	–0.0186
upper face	0.9853	0.0418
POL	0.9379	–0.1683
OOL	0.8790	0.4150
pedicel L	0.9593	0.0075
funicle1 L	0.9443	–0.2145
scape	0.9662	–0.1718
postmarginal vein	0.8990	0.1985
gaster L	0.8418	–0.4746
gaster B	0.8407	0.4196

Discussion

In the taxonomy of Pteromalidae, where qualitative characters for the separation of closely related species are frequently lacking, one has to rely on quantitative characters. Graham (1969), who was one of the first using morphometric measurements, calculated single ratios to discriminate among species. The above results demonstrate that such ratios are often insufficient for an unambiguous identification of species. This is also the case for *Pteromalus solidaginis* (*sol*), which is generally quite well distinguished from the other species. The best ratios (e.g. POL/OOL, gaster L/B) still show some amount of overlap with *P. albipennis* (*alb*). Eye H/B mentioned by Graham & Gijswijt (1991) as one of the key characters for the separation of *sol* from *dec* allows the classification of only 70% of the specimens (Fig. 1). This overlap is also likely to increase when more material also from other localities is available for study. Specimens of *sol* and *dec* originate from a few localities in Southern France and England respectively (Appendix) and their overall variation is thus distinctly smaller than in *alb* (Fig. 3). Therefore, the discriminant functions which were found in the DA, are much more valuable. They are based on the information of several characters and are thus more reliable and have a much better separating power (Fig. 2). A discriminant function is also a very useful tool for the identification of new specimens. In this respect, Flury & Riedwyl (1988) invented a practical method for which they coined the term identification analysis.

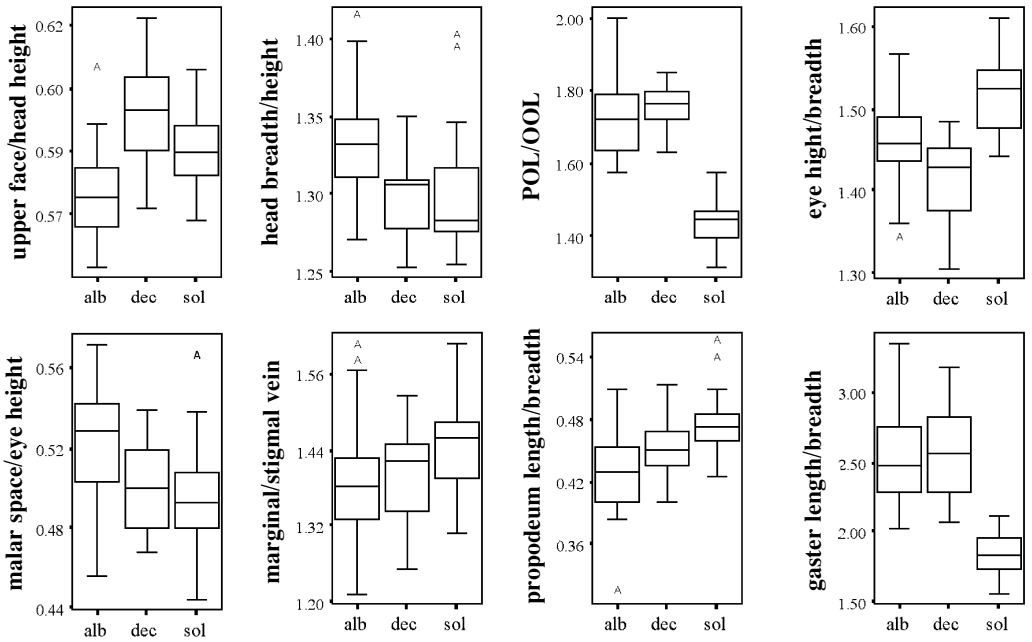

Figure 1 Box-and-whisker plots of eight ratios commonly used in Pteromalidae

According to the results of the PCA, *P. decipiens* (*dec*) is separated from *alb* mainly by its smaller size, despite the fact that their means may also be significantly different for values of the second principal component, the 'shape' vector (Fig. 3). The variance of this axis, however, is only 7.2% of the total variance, hence the differences related to shape may actually be negligible. Differences in size alone are not necessarily a good indication for separate species. For instance, it is well known from examples in parasitic Hymenoptera that specimens of the same species may vary significantly in size when reared from a different host species (Quicke 1997). Unfavourable conditions during development as caused by superparasitism are probably another source of such variation (resulting e. g. in 'dwarf' specimens). So, it might be questioned whether *dec* really deserves a specific status and is not just a form of *alb*. To address this problem, it is important to consider all available data. In this study it was found that some qualitative characters such as the sculpture of the propodeum, were also important. Further characters were mentioned by Graham (1969: 548) in the original description, for instance the colour and shape of the flagellum. These have actually been used by the author for identification of specimens prior to the morphometric analyses. Therefore, multivariate statistics simply represent an alternative in those cases where 'traditional' approaches fail to reveal the difference between taxa.

In conclusion, the application of DA and PCA reveals corroborated evidence of a separate species in the case of *sol*. It is distinguished from both *alb* and *dec* by clear differences in shape or shape and size respectively. On the other hand, it was not possible to further substantiate the status of *dec*, as it is separated from *alb* mainly by its smaller size. However, the results do not contradict earlier findings regarding qualitative morphology, hence the specific status of *dec* is retained.

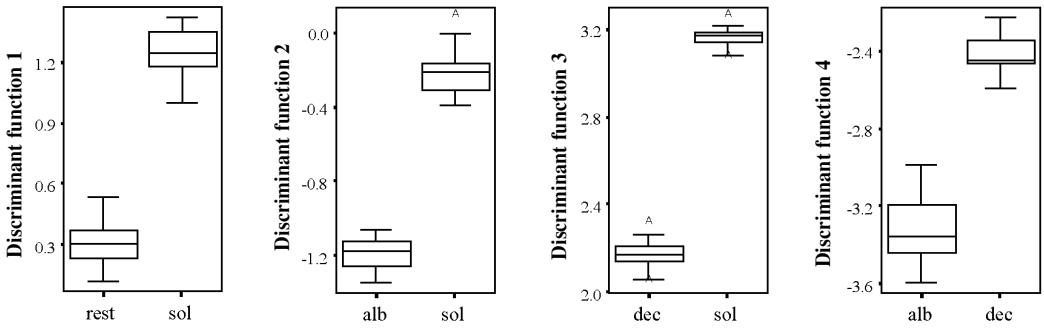

Figure 2 Box-and-whisker plots for values of the discriminant functions found in four different DAs (see Table 2). The separation of groups has strongly increased in comparison with the values of simple ratios (Fig. 1)

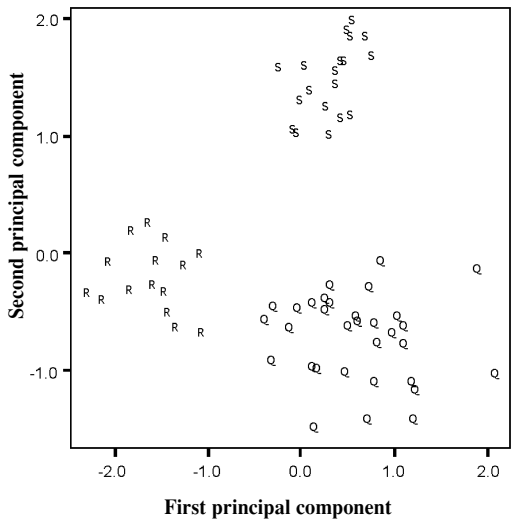

Figure 3 Plot of 66 specimens (+ alb, dec, sol) against values for the first and second principal component. The components comprise about 93% of the total variance

Acknowledgements

I thank Hans Riedwyl, Institute for Mathematical Statistics, Berne, for critical reading of a manuscript draft. I am also grateful to Elsa Obrecht and Charles Huber, both Natural History Museum, Berne (NMBE), and Stefan Ungricht, CEFE-CNRS, Montpellier, for discussion and useful suggestions. For the loan of specimens I thank John S. Noyes, The Natural History Museum, London (BMNH), and Stefan Vidal, Institute for Plant Pathology and Plant Protection, Göttingen (VID).

References

Bouček, Z. (1988) *Australasian Chalcidoidea (Hymenoptera): A biosystematic revision of genera of fourteen families, with a reclassification of species*. C.A.B. International, Wallingford.

Flury, B. & Riedwyl, H. (1988) *Multivariate statistics: A practical approa*ch. Chapman & Hall, London.

Gibson, G.A.P. (1997) Morphology and terminology. pp. 16–44. *In* Gibson, G.A.P., Huber, J.T. & Woolley, J.B. (eds), *Annotated keys to the genera of Nearctic Chalcidoidea (Hymenoptera)*. NRC Research Press, Ottawa.

Graham, M.W.R. de V. (1969) The Pteromalidae of North-Western Europe. *Bulletin of the British Museum* (*Natural History*), *Entomology*, *Supplement* **16**: 1–908.

Graham, M.W.R. de V. & Gijswijt, M.J. (1991) A new species of *Pteromalus* (Hymenoptera: Chalcidoidea) from France, associated with *Solidago virgaurea*. *Entomologische Berichten Amsterdam* **51**: 153–155.

Janzon, L.-Å. (1984) Taxonomy of a new subgroup of the *Pteromalus albipennis* group and their host-relationships, with description of four new species (Hymenoptera: Chalcidoidea). Paper VI. pp.1–34 pp. *In* Janzon, L.-Å. (ed), *Taxonomical and biological studies of Tephritis species (Diptera) and their parasitoids (Hymenoptera)*. PhD Thesis, University of Stockholm, Stockholm.

Janzon, L.-Å. (1986) Morphometric studies of some *Pteromalus* Swederus species (Hymenoptera: Chalcidoidea) with emphasis on allometric relationships, or: Are ratios reliable in chalcid taxonomy? *Systematic Entomology* **11**: 75–82.

Manly, B.F.J. (1994) *Multivariate statistical methods: A primer*. Chapman & Hall, London, 2nd edition.

Pimentel, R.A. (1992) An introduction to ordination, principal components analysis and discriminant analysis. pp. 11-28. *In* Sorensen, J.T. & Foottit, R. (eds), *Ordination in the study of morphology, evolution and systematics of insects: Applications and quantitative genetic rationals*. Elsevier, Amsterdam.

Podani, J. (2000) Introduction to the exploration of multivariate biological data. Backhuys, Leiden.

Quicke, D.L.J. (1997) *Parasitic wasps*. Chapman & Hall, London.

Shea, B.T. (1985) Bivariate and multivariate growth allometry: Statistical and biological considerations. *Journal of Zoology* **206**: 367–390.

Appendix

List of female specimens used for the data-matrix (number in square brackets). For abbreviations of depositories see acknowledgements.

Pteromalus albipennis (n=32): – ENGLAND: Cambs, Monks Wood (Straw), ex *Tephritis bardanae* (Schrank) (Diptera: Tephritidae), BMNH: 17.xi.1982 [1], 11.xi.1982 [1], Middlesex, Southgate (Graham), BMNH: 11.viii.1970 [1], 9.ix.1969 [1], 28.viii.1969 [1], 14.viii.1969 [2], 22.viii.1969 [3], 19.v.1970 [1]. – FRANCE: Bretagne, W Rennes, Guitté (Vidal), ex *Urophora quadrifasciata* (Meigen) or *Chaetorellia jacea* (Robineau-Desvoidy) (Diptera: Tephritidae), on *Centaurea nigra* L., VID: 8.ix.1995 [7]; Lozère, 2 km S Aleyrac (Baur), on *Senecio* sp., NMBE: 12.vii.1995 [3]; Lozère, Camprieu-Col de Faubel (Baur), NMBE: 13.vii.1995 [1]. – GERMANY: Niedersachsen, Göttingen, Rainshof (Denys), ex *Tephritis formosa* (Loew) (Diptera: Tephritidae), on *Sonchus oleraceus* L., VID: vii.1995 [6], 7.viii.1995 [2]. – SWITZERLAND: Wallis, Leuk, Brentjong (Baur), on *Achillea millefolium* L., NMBE: 15.vi.1996 [2].

P. decipiens (n=15): – ENGLAND: Birkshire, Newbury, Thatzham (Graham), paratype, BMNH: 29. viii.1964 [1]; Middlesex, Southgate (Graham), BMNH: 25.viii.1970 [4], 28. viii.1969 [5], 9.ix.1969 [1], 1.ix.1970 [4].

P. solidaginis (n=19): – FRANCE: Drőme, Séderon, Col de l'Homme Mort, on *Solidago virgaurea* L., paratypes, BMNH: (Graham) 15.viii.1988 [1], 1.viii.1990 [6], 27.viii.1990 [4], 15.viii.1988 [4], 1.viii.1990 [3]; (Gijswijt) 18.viii.1988 [1].