Discrimination of the notifiable pathogen *Gyrodactylus salaris* from *G. thymalli* (Monogenea) using statistical classifiers applied to morphometric data

E. S. McHUGH[1], A. P. SHINN[2]* *and* J. W. KAY[1]

[1] *Department of Statistics, University of Glasgow, Glasgow G12 8QQ, Scotland, UK*
[2] *Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK*

SUMMARY

The identification and discrimination of 2 closely related and morphologically similar species of *Gyrodactylus*, *G. salaris* and *G. thymalli*, were assessed using the statistical classification methodologies Linear Discriminant Analysis (LDA) and *k*-Nearest Neighbours (KNN). These statistical methods were applied to morphometric measurements made on the gyrodactylid attachment hooks. The mean estimated classification percentages of correctly identifying each species were 98·1 % (LDA) and 97·9 % (KNN) for *G. salaris* and 99·9 % (LDA) and 73·2 % (KNN) for *G. thymalli*. The analysis was expanded to include another 2 closely related species and the new classification efficiencies were 94·6 % (LDA) and 98·0 % (KNN) for *G. salaris*; 98·2 % (LDA) and 72·6 % (KNN) for *G. thymalli*; 86·7 % (LDA) and 91·8 % (KNN) for *G. derjavini*; and 76·5 % (LDA) and 77·7 % (KNN) for *G. truttae*. The higher correct classification scores of *G. salaris* and *G. thymalli* by the LDA classifier in the 2-species analysis over the 4-species analysis suggested the development of a 2-stage classifier. The mean estimated correct classification scores were 99·97 % (LDA) and 99·99 % (KNN) for the *G. salaris–G. thymalli* pairing and 99·4 % (LDA) and 99·92 % (KNN) for the *G. derjavini–G. truttae* pairing. Assessment of the 2-stage classifier using only marginal hook data was very good with classification efficiencies of 100 % (LDA) and 99·6 % (KNN) for the *G. salaris–G. thymalli* pairing and 97·2 % (LDA) and 99·2 % (KNN) for the *G. derjavini–G. truttae* pairing. Paired species were then discriminated individually in the second stage of the classifier using data from the full set of hooks. These analyses demonstrate that using the methods of LDA and KNN statistical classification, the discrimination of closely related and pathogenic species of *Gyrodactylus* may be achieved using data derived from light microscope studies.

Key words: *Gyrodactylus*, Monogenea, *Gyrodactylus salaris*, statistical classifiers, linear discriminant analysis, k-nearest neighbours.

INTRODUCTION

The reports and, in certain cases, misidentifications of *Gyrodactylus salaris* Malmberg, 1957 from fish stocks from a number of neighbouring European countries (see Table 1), is justifiably cause for concern to the UK's *G. salaris*-free status. While Shinn, Sommerville & Gibson's (1995) study of 234 salmonid sites throughout the UK and Platten, McLoughlin & Shinn's (1994) survey of 17 fish farms and 7 rivers in Northern Ireland demonstrated the British salmonid stocks to be *G. salaris*-free, the gyrodactylid fauna of British thymallids was at that time unknown. A survey of grayling, *Thymallus thymallus* L., in 9 English rivers during the period 1995–1996 by Denham & Long (1999) found that grayling from 6 of these rivers were infected with *G. thymalli* Zitnan, 1960. The impact *G. salaris* has had on fish stocks throughout Norway has warranted the development of techniques for its identification from

* Corresponding author: Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland. Tel: +44 1786 473171. Fax: +44 1786 472133. E-mail: aps1@stir.ac.uk

other and relatively benign close relatives. Using the nucleotide sequence of the srRNA V4 region, Cunningham *et al*. (1995) demonstrated the discrimination of *G. salaris* from 2 species of *Gyrodactylus*, *G. derjavini* Mikailov, 1975 and *G. truttae* Gläser, 1974, commonly occurring on European salmonids. There was, however, no variation in the 150 base nucleotide sequence between *G. salaris* and *G. thymalli*. Cunningham *et al*. (1995) concluded that while the V4 region may not provide the information necessary to discriminate these 2 morphologically similar species, other variable regions of the genome may. *G. thymalli* and its discrimination from *G. salaris* represents a problem within gyrodactylid taxonomy. Malmberg (1987*b*) suggested that forms of the 2 species found on certain salmonid hosts are morphologically similar as shown by Fig. 1 (see also Mo & Appleby, 1990; Shinn *et al*. 1995; Denham & Long, 1999); and work by Bakke (1991) and Bakke & Jansen (1991) on the experimental cross-transfer of *G. salaris* and *G. thymalli* between their recognized fish hosts found that both species of *Gyrodactylus* were able to survive and reproduce. Later, in 1994, Mo's (1994) exam-

Table 1. Reports of *Gyrodactylus salaris* in Northern Europe

Report of *G. salaris*	Reference
Bosnia-Herzegovina	Imamovic (1987)
Czechoslovakia	Ergens (1961)*; Rehulka (1973)†
Denmark	Malmberg (1973)
Finland	Rimaila-Pärnänen & Wiklund (1987)
France	Johnston, MacKenzie, Cunningham, Eiras & Bruno (1996)
Germany	Schmahl (1988); Lux (1990)
Karelia	Ieshko, Berland, Shulman, Bristow & Shurov (1995)
Norway	Malmberg (1973)
Portugal	Johnston, MacKenzie, Cunningham, Eiras & Bruno (1996)
Russia	Ergens (1983)‡; Rumyantsev (1989)
Scotland	Campbell (1974)§
Slovakia	Zitnan (1974)
Spain	Santamarina, Tojo, Ubeira, Quinteiro & Sanmartin (1991)
Sweden	Malmberg (1957)
Ukraine	Tesarcik & Ivasik (1974)
Yugoslavia	Kulakovskaya (1967)

* Believed by Tanum (1983) to be another species not *G. salaris*.
† Believed by Mo (1983) to be *G. truttae*, however, Malmberg (1987*c*) believes that *G. truttae* is also erroneous but Rehulka's specimens resemble another *Gyrodactylus* sp. closely related to *G. salaris*.
‡ Recorded as *Gyrodactylus* sp. but suggested by Malmberg (1987*b*) to most probably represent *G. salaris*.
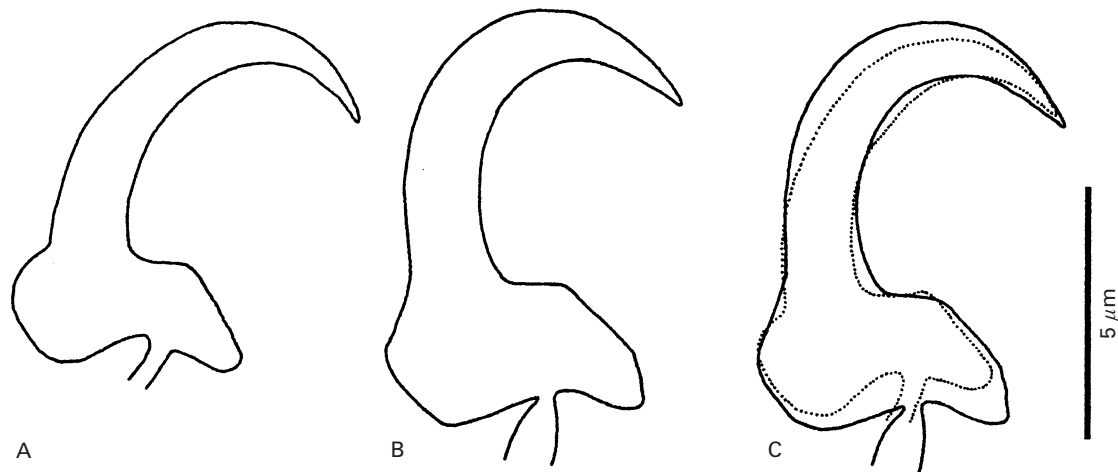§ Mis-identification, more likely to be *G. derjavini* or *G. truttae*.

Fig. 1(A) Marginal hook sickle of *Gyrodactylus thymalli* Zitnan, 1960 collected from *Thymallus thymallus* in the River Test, UK in April 1997. (B) Marginal hook sickle of *G. salaris* Malmberg, 1957 collected from *Salmo salar* in the River Ätran (Nydale), Sweden in April 1992. (C) Overlay of the *G. salaris* marginal hook sickle (solid line) with that of *G. thymalli* (dotted line) to show differences in hook morphology.

ination of variation in the morphology of the opisthaptoral sclerites of *G. salaris* between localities sampled led him to consider that the form of *G. salaris* on rainbow trout was indistinguishable from the sclerites of *G. thymalli*.

An investigation by Kay, Shinn & Sommerville (1999) demonstrated the potential for the use of modern statistical classifiers in differentiating species of the genus *Gyrodactylus*. The study applied 2 statistical methods, nearest neighbours and linear discriminant analysis, to data derived from light microscope images of the opisthaptoral hooks. This approach resulted in the perfect discrimination of *G. salaris* from other closely related species. The current study investigates the ability of these 2 statistical classification methodologies to discriminate *G. salaris* from *G. thymalli* and the latter species from other gyrodactylids parasitizing salmonids. The notion of a suggested synonymy of *G. thymalli* with *G. salaris* that could potentially invalidate the *G. salaris*-free status of the UK or that present methodologies may give rise to false positives with regard to

the identification of *G. salaris* is worrying. Serious consequences could stem from inaccurate species determinations either as *G. salaris* being mis-identified as *G. thymalli* and thus allowing *G. salaris* to go unchecked in the environment or, *G. thymalli* being identified as *G. salaris* and costly treatments being implemented to contain and prevent its spread. Both errors could have serious impacts on fish communities and their environment.

MATERIALS AND METHODS

Specimen collection and preparation

Specimens of *Gyrodactylus* von Nordmann, 1832 parasitizing salmonids collected during the 1990–1992 survey conducted by Shinn *et al.* (1995) were re-analysed in conjunction with specimens of *G. salaris* Malmberg, 1957 and *G. thymalli* Zitnan, 1960. Details pertaining to the sites of collection for the 134 specimens of *G. truttae*, 1974 and 254 *G. derjavini* Mikailov, 1975 taken during the survey have been given by Shinn, Kay & Sommerville (2000). Ten specimens of *G. thymalli* parasitizing *Thymallus thymallus* in the River Test, Hampshire, England were collected in April 1997 and provided for this study by Dr K. L. Denham (CEFAS). The British gyrodactylid material was compared to 85 specimens of *G. salaris* collected from *Salmo salar* from various watercourses across Scandinavia by Dr G. Malmberg (University of Stockholm, Sweden) and Professor T. A. Bakke (Zoological Museum, University of Oslo, Norway) throughout the 1990–1992 time-period. All the gyrodactylid material was prepared for light microscopy, mounted in am-monium picrate glycerine, according to the meth-odology given by Malmberg (1957).

Morphometric data on the attachment hooks were collected from slide preparations of *Gyrodactylus* using an eye-piece graticule at ×100, oil immersion on a BH2 Olympus binocular microscope with phase-contrast. Ten point to point morphometric measurements were made, 3 from the hamulus (total length, shaft length and point length), 2 from the ventral bar (total length and total width) and 5 from the marginal hook (total length, shaft length, sickle length, sickle proximal width and sickle distal width). Using the marginal hook numbering system of Llewellyn (1970), marginal hook number 8 was measured on each individual gyrodactylid. Where this was not possible, then marginal hook number 7 was measured.

Statistical methods

Various statistical classification techniques were used to build classification rules and to assess their likely performance in the classification of new specimens (McLachlan, 1992).

Canonical variate plots were constructed in each case in order to visualize the data with respect to the canonical variate axes. These are determined by the linear functions of the measurements which best separate the different classes. The canonical variates may be used to classify objects into one of the classes. When there are 2 classes, there is just 1 canonical variate and when there are 4 classes there are 3 canonical variates.

The method of Linear Discriminant Analysis (LDA) was used to estimate a classification rule; when it is used just for classification this method is identical to the use of canonical variates. Because the specimens which are available for analysis are not necessarily representative of the naturally-occurring frequencies, and particularly because the number of *G. thymalli* specimens is small relative to the other types, it was assumed *a priori* that the class probabilities are equal. In the absence of precise relevant information of the relative costs of misclassi-fications, it was assumed that these costs are equal; this results in specimens being allocated to the most likely class on the basis of the measurements. The method of k-Nearest Neighbours (KNN) was also used in each case, and this provides a completely non-parametric approach which allows non-linear boundaries to be fitted between the classes. In the applications described herein, 3 nearest neighbours were used.

The method of repeated, stratified, 5-fold cross-validation was used to assess the generalization error likely to be obtained when a classifier is applied to new specimens. In this approach the available specimens are split randomly into 5 groups in proportion to the numbers of each type that are available. One of the 5 groups of data is held back to form a test set, and the remaining data are used to build the classifier. The resulting rule is then applied to the test set and the predictions obtained for each test specimen are compared with the true types; thus the number of misclassifications are calculated. This procedure is repeated taking, in turn, each of the other 4 groups to be the test set, and the numbers are combined to form an overall estimate of the rates of correct classification and also each type of misclassi-fication. This method makes efficient use of the available data. This entire procedure was then repeated 100 times and the mean and range of the resulting estimates of the classification percentages were computed.

The analyses were conducted using scripts using public-domain software provided (Venables & Ripley, 1997) for the S-PLUS statistical package (S-PLUS 4, 1997).

RESULTS

Discrimination of G. salaris *and* G. thymalli

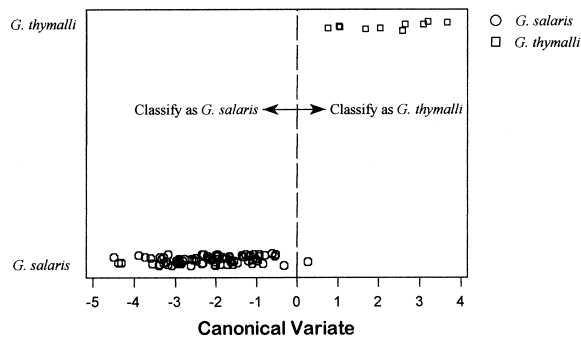A canonical variate plot was constructed using all 10 measurements taken from the 10 *G. thymalli*

Fig. 2. Canonical variate plot of morphometric data from the species *Gyrodactylus salaris* and *G. thymalli*.
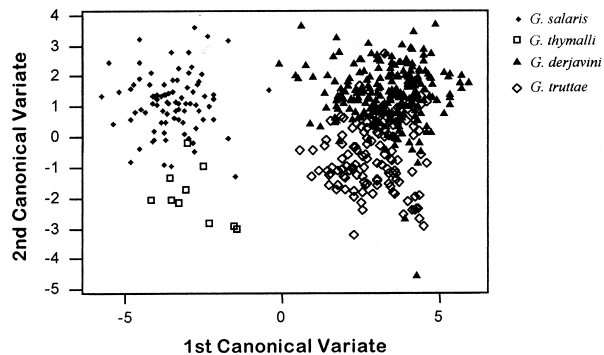

Fig. 3. Canonical variate plot of morphometric data from the 4 species of *Gyrodactylus* (*G. derjavini*, *G. salaris*, *G. thymalli* and *G. truttae*).

specimens and the 85 *G. salaris* specimens and this is displayed in Fig. 2. It is clear that there is very good separation between the classes and that this method produces only 1 misclassification (*G. salaris* as *G. thymalli*) when applied to these 'training' specimens. Of course, as the aim would be to use the classification rule on new specimens we need to assess the predictive ability of this classification rule. The method of repeated, stratified, 5-fold cross-validation was used to estimate the misclassification rates likely to be found with new specimens. The results are given in Table 2.

Using the LDA classifier we may estimate that 98·1 % of *G. salaris* specimens would be classified correctly as being *G. salaris* and 99·9 % of *G. thymalli* specimens would be classified correctly as being *G. thymalli* (on average); also 1·9 % of *G. salaris* specimens would be misclassified as being *G. thymalli* and 0·1 % of *G. thymalli* specimens would be misclassified as being *G. salaris*. The results obtained using the KNN method are similar to the LDA results for *G. salaris*, but not as good for *G. thymalli*.

Discrimination of the 4 species

The canonical variate plot using the first 2 canonical axes and based on all 10 measurements taken from the specimens of *G. salaris*, *G. thymalli*, *G. derjavini* and *G. truttae* is shown in Fig. 3. The most striking feature of the plot is that *G. thymalli* is most similar to *G. salaris* and that these 2 types are clearly separated from *G. derjavini* and *G. truttae*; these

latter 2 types show a fair degree of overlap. The predictive ability of the classifiers for all 4 types was assessed using repeated, stratified, 5-fold cross-validation and the results are given in Table 3.

With the LDA method we may estimate that 94·6 % of *G. salaris* specimens, 98·2 % of *G. thymalli* specimens, 86·7 % of *G. derjavini* specimens and 76·5 % of *G. truttae* specimens, respectively, would be classified correctly. With the KNN method, the corresponding percentages are 98·0, 72·6, 91·8, 77·7 %. Hence the KNN method is likely to perform better for the *G. salaris*, *G. derjavini* and *G. truttae* specimens, although overall LDA is better for the discrimination of *G. salaris* and *G. thymalli* and KNN is better for the discrimination of *G. derjavini* and *G. truttae*. Table 3 shows also the percentages of the various types of misclassification; the main problem lies in the confusion between *G. derjavini* and *G. truttae*, although with the KNN method there is confusion between *G. thymalli* and *G. salaris*. It is clear that there is very little confusion between the pairing *G. salaris*–*G. thymalli* and the pairing *G. derjavini*–*G. truttae*.

A 2-stage classifier

Given the results obtained in the previous 2 subsections, namely that (a) the LDA results are better for discrimination of *G. salaris* and *G. thymalli* in the 2-class case than in the 4-class case and (b) the pairings *G. salaris*–*G. thymalli* and *G. derjavini*–*G. truttae* are clearly separated from each other, it seems

Table 2. Estimated classification percentages for the discrimination of *Gyrodactylus salaris* and *G. thymalli* using the LDA and KNN methods of classification

Classification	LDA			KNN		
	Mean	Min	Max	Mean	Min	Max
G. salaris as G. salaris	**98·1**	**95·3**	**98·8**	**97·9**	**95·3**	**100·0**
G. salaris as G. thymalli	1·9	1·2	4·7	2·1	0·0	4·7
G. thymalli as G. salaris	0·1	0·0	10·0	26·8	20·0	30·0
G. thymalli as G. thymalli	**99·9**	**90·0**	**100·0**	**73·2**	**70·0**	**80·0**

Table 3. *Estimated classification percentages for the discrimination of* Gyrodactylus salaris, *G.* thymalli, *G.* derjavini *and* G. truttae *using the LDA and KNN methods of classification*

	LDA			KNN		
Classification	Mean	Min	Max	Mean	Min	Max
G. salaris as G. salaris	**94·6**	**91·8**	**96·5**	**98·0**	**96·5**	**100·0**
G. salaris as G. thymalli	5·2	3·5	8·2	2·0	0·0	3·5
G. salaris as G. derjavini	0·2	0·0	1·2	0·0	0·0	0·0
G. salaris as G. truttae	0·0	0·0	0·0	0·0	0·0	0·0
G. thymalli as G. salaris	1·8	0·0	10·0	27·4	20·0	30·0
G. thymalli as G. thymalli	**98·2**	**90·0**	**100·0**	**72·6**	**70·0**	**80·0**
G. thymalli as G. derjavini	0·0	0·0	0·0	0·0	0·0	0·0
G. thymalli as G. truttae	0·0	0·0	0·0	0·0	0·0	0·0
G. derjavini as G. salaris	0·3	0·0	2·8	0·1	0·0	0·8
G. derjavini as G. thymalli	0·0	0·0	0·0	0·0	0·0	0·0
G. derjavini as G. derjavini	**86·7**	**82·9**	**91·2**	**91·8**	**86·8**	**94·8**
G. derjavini as G. truttae	13·0	8·8	17·1	8·1	5·2	13·1
G. truttae as G. salaris	0·0	0·0	0·0	0·0	0·0	0·0
G. truttae as G. thymalli	0·0	0·0	0·0	0·0	0·0	0·0
G. truttae as G. derjavini	23·5	19·4	26·9	22·3	18·7	26·9
G. truttae as G. truttae	**76·5**	**73·1**	**80·6**	**77·7**	**73·1**	**81·3**

Table 4. *Estimated classification percentages for the discrimination of the 2 hybrid pairings* Gyrodactylus salaris–*G.* thymalli (SalThy) *and* G. derjavini–*G.* truttae (DerjTrut) *using the LDA and KNN methods of classification*

	LDA			KNN		
Classification	Mean	Min	Max	Mean	Min	Max
SalThy as SalThy	**99·97**	**98·9**	**100·0**	**99·99**	**98·9**	**100·0**
SalThy as DerjTrut	0·03	0·0	1·1	0·01	0·0	1·1
DerjTrut as SalThy	0·6	0·3	1·8	0·08	0·0	0·3
DerjTrut as DerjTrut	**99·4**	**98·2**	**99·7**	**99·92**	**99·7**	**100·0**

appropriate to consider a 2-stage classifier. In this classifier, the pairings *G. salaris*–*G. thymalli* and *G. derjavini*–*G. truttae* would be considered as 2 hyper-classes and the LDA and KNN methods applied to classify specimens into 1 of these 2 hyper-classes. If a specimen is classified as a *G. salaris*–*G. thymalli* hyper-class, the methods would be applied to just the data from the *G. salaris* and *G. thymalli* specimens to decide whether it is likely to be *G. salaris* or *G. thymalli* (as in Table 2 and Fig. 2). The method of repeated, stratified, 5-fold cross-validation was applied to all 10 measurements from the 95 *G. salaris*–*G. thymalli* specimens and the 385 *G. derjavini*–*G. truttae* specimens and the results are given in Table 4.

With the LDA method, it may be estimated that 99·97 % of specimens from the *G. salaris*–*G. thymalli* hyper-class and 99·4 % of specimens from the *G. derjavini*–*G. truttae* hyper-class would be correctly classified. The corresponding percentages for the KNN method are 99·99 % and 99·92 %. These results combined with those of Table 2 demonstrate that this 2-stage classifier based on the LDA method

is likely to be very successful as long as future *G. salaris* and *G. thymalli* data are similar to the specimens considered here.

Discrimination using subsets

As described in the Materials and Methods section, the available measurements fall naturally into 3 subsets corresponding to the sclerite structures: the hamulus, the ventral bar and the marginal hook, and it is of interest to discover whether classification based on just a single sclerite structure is likely to be sufficiently accurate. Hence the cross-validation experiments which are described in the previous subsections were repeated for each of the 3 subsets. We now report on the experiments corresponding to the use of the 2-stage classifier. The results obtained with LDA and KNN for the hyper-classes are given in Table 5.

While in some cases, particularly with the marginal hook subset, the results are quite good they are not quite as good as those obtained using the full set of measurements. Similarly, the LDA and KNN

Table 5. Estimated classification percentages for the discrimination of the 2 hybrid pairings *Gyrodactylus salaris–G. thymalli* and *G. derjavini–G. thymalli* using the LDA and KNN methods of classification – separate analyses based on the sclerite subsets of the measurements

Classification	LDA			KNN		
	Mean	Min	Max	Mean	Min	Max
Hamulus						
SalThy as SalThy	**96·8**	**96·8**	**96·8**	**93·7**	**91·6**	**96·8**
SalThy as DerjTrut	3·2	3·2	3·2	6·3	3·2	8·4
DerjTrut as SalThy	3·4	3·1	3·6	0·6	0·3	1·3
DerjTrut as DerjTrut	**96·6**	**96·4**	**96·9**	**99·4**	**98·7**	**99·7**
Ventral bar						
SalThy as SalThy	**47·6**	**33·7**	**52·6**	**21·6**	**14·7**	**29·5**
SalThy as DerjTrut	52·4	47·4	66·3	73·4	70·5	85·3
DerjTrut as SalThy	47·9	43·6	54·0	12·8	9·4	16·1
DerjTrut as DerjTrut	**52·1**	**46·0**	**56·4**	**87·2**	**83·9**	**90·6**
Marginal hook						
SalThy as SalThy	**100·0**	**100·0**	**100·0**	**99·6**	**96·8**	**100·0**
SalThy as DerjTrut	0·0	0·0	0·0	0·4	0·0	3·2
DerjTrut as SalThy	2·8	2·6	2·9	0·8	0·5	1·6
DerjTrut as DerjTrut	**97·2**	**97·1**	**97·4**	**99·2**	**98·4**	**99·5**

Table 6. Estimated classification percentages for the discrimination of *Gyrodactylus salaris* and *G. thymalli* using the LDA and KNN methods of classification – separate analyses based on the sclerite subsets of the measurements

Classification	LDA			KNN		
	Mean	Min	Max	Mean	Min	Max
Hamulus						
Salaris as Salaris	**88·2**	**84·7**	**89·4**	**98·8**	**97·6**	**98·8**
Salaris as Thymalli	12·8	10·6	15·3	1·2	1·2	2·4
Thymalli as Salaris	19·9	10·0	20·0	49·9	40·0	50·0
Thymalli as Thymalli	**80·1**	**80·0**	**90·0**	**50·1**	**50·0**	**60·0**
Ventral bar						
Salaris as Salaris	**82·4**	**78·8**	**85·9**	**95·5**	**90·6**	**98·8**
Salaris as Thymalli	17·6	14·1	21·2	4·5	1·2	9·4
Thymalli as Salaris	15·1	10·0	30·0	79·7	70·0	100·0
Thymalli as Thymalli	**84·9**	**70·0**	**90·0**	**20·3**	**0·0**	**30·0**
Marginal hook						
Salaris as Salaris	**90·9**	**87·1**	**94·1**	**99·3**	**97·6**	**100·0**
Salaris as Thymalli	9·1	5·9	12·9	0·7	0·0	2·4
Thymalli as Salaris	15·4	0·0	20·0	21·3	20·0	30·0
Thymalli as Thymalli	**84·6**	**80·0**	**100·0**	**78·7**	**70·0**	**80·0**

methods were applied to the sclerite subsets for the *G. salaris* and *G. thymalli* classes and the results are given in Table 6; again the marginal hook subset shows the best performance but it is not quite as good as that obtained using the full set of measurements.

Again, while the results in individual cases are quite good, the level of likely accuracy is not high enough compared to that obtained with the full dataset.

DISCUSSION

Malmberg (1993) commented that the 'form' of *G. salaris* parasitizing *Oncorhynchus mykiss* (Walbaum) was reminiscent of *G. thymalli* and that the discrimination of such forms could only effectively be resolved by the collection and comparison of both species of *Gyrodactylus* from the full range of hosts and environmental conditions under which both species are found. Because of the scarcity of *G.*

thymalli specimens available for this study, however, it has not been possible to encompass the full range of host and environmental variation for this species. For the comparator species *G. salaris*, material from a number of rivers, taken over a 15-month period, has been available for study and therefore provides a more satisfactory data-set. Whilst the present investigation has indicated clear differences between these species, it is nevertheless unwise to draw definitive conclusions until a larger sample of *G. thymalli* can be obtained.

Mo's (1991*b*) report of a species, tentatively identified as *G. salaris*, from *O. mykiss* cited differences from *G. salaris* taken from *S. salar* in terms of the possession of a stouter ventral bar and hamuli. In the same paper, no difference in the shape of the marginal hook was noted between these two morphs. In the present study it has not been possible, however, to obtain specimens of '*G. salaris*' from *O. mykiss* and therefore the affinity of this species appears uncertain. The differences in shape between the marginal hook sickles of *G. thymalli* and *G. salaris*, however, are presented in Fig. 1C of the present study. While both sickles collected from the same month of the year are of a similar size, the marginal hook of *G. salaris* has a deeper foot to the sickle, the toe being noticeably broader. Both sickle points follow a similar curve but the sickle shaft and point of *G. salaris* are more robust. The variation in the size of the opisthaptoral sclerites of *G. salaris* with season was investigated by Mo (1991*a*, *b*, *c*), who found that the hooks were larger in colder water temperatures and smaller in warmer water temperatures. Variation in the shape of the sclerites with season was, however, less pronounced, the shape of the marginal hooks being notably constant. Regardless of season, the attachment hooks are fully developed at birth and undergo no further development or growth, only the hamulus root portion may continue to grow (Ergens, 1965; Malmberg, 1970).

Both *G. salaris* and *G. thymalli* belong to the *G. salaris*-group but are designated within different subgroups, their placement within these being based on the relative robustness of their hamuli and ventral bars (Malmberg, 1993). Whilst subtle differences in the shape of the marginal hook sickle, which are highlighted by the measurements used in the present study, allow experts to discriminate these species, it is questionable whether discrimination of these species may be achieved using univariate statistics on the classical point to point morphometrics gathered from the gyrodactylid attachment hooks.

The discrimination of gyrodactylid species has been achieved to a certain level in earlier studies using multivariate analyses such as principal components analysis (PCA) applied to morphometric data derived from specimens prepared for light microscopy (Shinn, 1994; Shinn *et al.* 1996) and on sonication liberated hooks examined using the scanning electron microscope (SEM) (Shinn, Gibson & Sommerville, 1993; Shinn, 1994). While these studies allow for the discrimination of a species as a collective of individuals, however, the confidence that one might place upon the discrimination of an individual gyrodactylid is dependent upon its relative position in the PCA matrix and thus its position in relation to all other specimens within the analysis. Specimens positioned at the distal periphery of its species range in a PCA matrix might be clearly discriminated from another specimen situated on the distal periphery of its own species range but the confidence in discriminating 2 species diminishes where 2 species overlap. Thus it was not surprising that, where *Gyrodactylus* species were difficult to classify using conventional taxonomic procedures, the progression in gyrodactylid taxonomy and/or discrimination was to move away from methodologies with a numeric or morphometric basis to techniques with a molecular biology basis (Cunningham *et al.* 1995). However, Kay *et al.* (1999) and Shinn *et al.* (2000) established that the discrimination of closely related and morphologically similar *Gyrodactylus* species was possible using a variety of statistical classification approaches applied to morphometric data. The use of a statistical classification system once optimized and validated presents a rapid, reliable and cheap alternative to traditional methods and permits the early detection of pathogens from large numbers of samples.

This paper demonstrates with some success, the high correct percentage classifications of 2 closely related and morphologically similar species of *Gyrodactylus*, *G. salaris* (98·1 % LDA; 97·9 % KNN) and *G. thymalli* (99·9 % LDA; 73·2 % KNN), using 2 statistical classification methodologies, LDA and KNN. Expansion of the analysis to include additional species of *Gyrodactylus* illustrates that *G. thymalli* can also be discriminated from these (*G. salaris* – 94·6 % LDA; 98·0 % KNN); (*G. thymalli* – 98·2 % LDA; 72·6 % KNN); (*G. derjavini* – 86·7 % LDA; 91·8 % KNN); and (*G. truttae* – 76·5 % LDA; 77·7 % KNN). It is important to reiterate at this stage that the classification of these species is based upon relatively crude light-derived morphometric data i.e. 10 morphometric characters (3 hamulus, 2 ventral bar and 5 marginal hook). Thus by the addition of extra feature descriptors and/or by the development of algorithms for the detailed shape analysis of all or individual attachment hooks it would perhaps be possible to obtain better classification performance. The results from these early analyses illustrated that higher correct classification scores were obtained for *G. salaris* and *G. thymalli* by the LDA classifier in the 2-species analysis over the 4-species analysis. Also, the clear discrimination of the morphologically similar *G. salaris–G. thymalli* (99·97 % LDA; 99·99 % KNN)

pairing from the *G. derjavini–G. truttae* (99·4 %
LDA; 99·92 % KNN) pairing proposed the de-
velopment of a 2-stage classifier. When the in-
dividual attachment hooks were considered, the
classification percentages for the marginal hook were
equally good (*G. salaris–G. thymalli* pairing 100 %
LDA, 99·6 % KNN; *G. derjavini–G. truttae* pairing
97·2 % LDA, 99·2 % KNN) but not so for the other
hook structures. While the benefits of using a 2-stage
classifier are suggested here, a combined procedure
working as an entire system is as yet to be evaluated,
but it is clear that it would work well. Mis-
classifications, particularly in the case of *G. salaris*,
would have serious consequences should they allow
this species to slip through undetected. It is possible,
however, to incorporate estimates regarding the
relative seriousness of each type of misclassification
into the classifier and thereby weight the discrimi-
nation to minimize more serious errors. For the
purposes of this study no such estimates were built
into the system, with the costs of the various types of
misclassification assumed to be equal.

This work is promising, but more *G. thymalli*
specimens are required in order to be more confident
about the success of the application of our rules to
new specimens. The high classification efficiencies
are based on the data at hand and our estimates of
likely future performance are contingent on the
future data being similar to that of the present data.
While *G. thymalli* material from a variety of hosts
and environmental conditions was sought from
national helminth collections across northern
Europe, virtually no deposited material was avail-
able. The problem here is that with only 10 *G.
thymalli* specimens collected from a single locality,
we must be cautious in our statement of discrimi-
natory power.

REFERENCES

BAKKE, T. A. (1991). A review of the inter- and
intraspecific variability in salmonid hosts to laboratory
infections with *Gyrodactylus salaris* Malmberg.
Aquaculture **98**, 303–310.

BAKKE, T. A. & JANSEN, P. A. (1991). Susceptibility of
grayling (*Thymallus thymallus*) to *Gyrodactylus salaris*
Malmberg (Monogenea). *Bulletin of the Scandinavian
Society of Parasitology* (*Proceedings of the XVth
Symposium of the Scandinavian Society for
Parasitology*) **1**, 61.

CAMPBELL, A. D. (1974). The parasites of fish in Loch
Leven. *Proceedings of the Royal Society of Edinburgh*,
(*1972–1973*) **74**, 347–364.

CUNNINGHAM, C. O., MCGILLIVRAY, D. M., MACKENZIE, K. &
MELVIN, W. T. (1995). Identification of *Gyrodactylus*
(Monogenea) species parasitizing salmonid fish using
DNA probes. *Journal of Fish Diseases* **18**, 539–544.

DENHAM, K. L. & LONG, J. (1999). Occurrence of
Gyrodactylus thymalli Zitnan, 1960 on grayling,
Thymallus thymallus (L.), in England. *Journal of Fish
Diseases* **22**, 247–252.

ERGENS, R. (1961). Zwei weitere Befunde der
Gyrodactylus Art (Monogenoidea) aus der
Tschechoslowakei. *Vestnik Ceskoslovenské Spolecnosti
Zoologické* **25**, 25–27.

ERGENS, R. (1965). Die Morphogenese der chitinoiden
Teile des Haptors bei *Gyrodactylus decorus* Malmberg,
1956 (Monogenoidea) und ihre morphologisch-
metrische Variabilität. *Zeitschrift für Parasitenkunde*
25, 359–370.

ERGENS, R. (1983). *Gyrodactylus* from Eurasian
freshwater Salmonidae and Thymallidae. *Folia
Parasitologica* **30**, 15–26.

IESHKO, E. P., BERLAND, B., SHULMAN, S., BRISTOW, G. A. &
SHUROV, I. L. (1995). Parasites of Atlantic salmon parr
in some Karelian rivers flowing into the White Sea.
*Abstract of the Fourth International Symposium of
Ichthyoparasitology*, *Munich, Germany, October*,
p. 100.

IMAMOVIC, V. (1987). Paraziti i parazitoze riba u nekim
Salmonidnim ribogojilistima u Bosni i Hercegovini.
III. Girodaktiloza. (Parasites and parasitoses of fishes
in some salmonid fish ponds in Bosnia-Herzegovina).
Veterinaria **36**, 159–166. (In Serbo-croat.)

JOHNSTON, C., MACKENZIE, K., CUNNINGHAM, C. O., EIRAS,
J. C. & BRUNO, D. W. (1996). Occurrence of
Gyrodactylus salaris Malmberg, 1957, in Portugal.
*Bulletin of the European Association of Fish
Pathologists* **16**, 89–91.

KAY, J. W., SHINN, A. P. & SOMMERVILLE, C. (1999).
Towards an automated system for the identification of
notifiable pathogens. *Parasitology Today* **15**, 201–206.

KULAKOVSKAYA, O. P. (1967). Fragments on the parasite
fauna of fishes from the upper course of the Rivers
Tisa and Seret. *Helminthologia* **8**, 289–295.

LLEWELLYN, J. (1970). Monogenea. In *Technical Review:
Taxonomy, Genetics and Evolution of Parasites. Second
International Congress of Parasitology. Journal of
Parasitology* **56**, 493–504.

LUX, E. (1990). *Gyrodactylus salaris* – Parasitierung von
Salmoniden, ein diagnostisch – taxomisches
Problem. *Tagung der Fachgruppe 'Fischkrankheiten' in
Verbindung mit der EAFP/Deutsche Sektion European
Association of Fish Pathologists*, *14–16 November, 1990.*
pp. 87–98.

MALMBERG, G. (1957). [On the occurrence of
Gyrodactylus on Swedish fishes.] *Skrifter utgivna av
Södra Sveriges Fiskeriförening Årsskrift* 1956, 19–76
(In Swedish with an English summary).

MALMBERG, G. (1970). The excretory systems and
marginal hooks as a basis for the systematics of
Gyrodactylus (Trematoda, Monogenea). *Arkiv för
Zoologi, Serie* 2, **23**, 1–235.

MALMBERG, G. (1973). *Gyrodactylus* infestations on
species of *Salmo* in Danish and Swedish hatcheries.
Norwegian Journal of Zoology **21**, 325–326.

MALMBERG, G. (1987*a*). Increased intraspecific
divergence in *Gyrodactylus salaris* resulting from

genetic drift in fish farm populations. *Information, Åbo Akademii* **19**, 33.

MALMBERG, G. (1987*b*). *Gyrodactylus salaris* Malmberg, 1957 and *G. truttae* Gläser, 1974 – two problematic species. *Information, Åbo Akademii* **19**, 34.

MALMBERG, G. (1987*c*). *Gyrodactylus* – en haptormask av ekonomiskt interesse for fiskodlare. *Vattenbruk* **2**, 15–20.

MALMBERG, G. (1993). Gyrodactylidae and gyrodactylosis of Salmonidae. *Bulletin Français de la Pêche et de la Pisciculture* **328**, 5–46.

MCLACHLAN, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, Chichester and New York.

MO, T. A. (1983). *Gyrodactylus truttae* Gläser, 1975, på ørret, *Salmo trutta* L. og laks *Salmo salar* L. i Sanvikvassdraget. Taksonomi og infeksjonsforhold Hovedfagsoppgave i spesiell zoologi. Thesis, University of Oslo, Zoologisk Museum.

MO, T. A. (1991*a*). Seasonal variations of opisthaptoral hard parts of *Gyrodactylus salaris* Malmberg, 1957 (Monogenea: Gyrodactylidae) on parr of Atlantic salmon *Salmo salar* L. in the River Batnfjordselva, Norway. *Systematic Parasitology* **19**, 231–240.

MO, T A. (1991*b*). Variations of opisthaptoral hard parts of *Gyrodactylus salaris* Malmberg, 1957 (Monogenea: Gyrodactylidae) on rainbow trout *Oncorhynchus mykiss* (Walbaum, 1792) in a fish farm, with comments on the spreading of the parasite in south-eastern Norway. *Systematic Parsitology* **20**, 1–9.

MO, T. A. (1991*c*). Variations of opisthaptoral hard parts of *Gyrodactylus salaris* Malmberg, 1957 (Monogenea: Gyrodactylidae) on parr of Atlantic salmon *Salmo salar* L. in laboratory experiments. *Systematic Parasitology* **20**, 11–19.

MO, T. A. (1994). Status of *Gyrodactylus salaris* problems and research in Norway. In *Parasitic Diseases of Fish*, (ed. Pike, A. W. & Lewis, J. W), pp. 43–56. Samara Publishing Ltd, Dyfed.

MO, T. A. & APPLEBY, C. (1990). A special technique for studying haptoral sclerites of monogeneans. *Systematic Parasitology* **17**, 103–108.

PLATTEN, M., MCLOUGHLIN, M. & SHINN, A. P. (1994). Distribution and identification of gyrodactylid species in fish farms and rivers of Northern Ireland. *Veterinary Record* **135**, 411–412.

REHULKA, J. (1973). Remarks on the occurrence of *Gyrodactylus salaris* Malmberg, 1957 *sensu* Ergens, 1961 (Monogenoidea: Gyrodactylidae). *Vestnik Ceskoslovenské Spolecnosti Zoologické* **37**, 293–295.

RIMAILA-PÄRNÄNEN, E. & WIKLUND, T. (1987). *Gyrodactylus salaris* – loismadon levinneisyydestä makeanveden kalanviljelylaitoksissamme. *Suomen Eläinlääkärilehti* **93**, 506–507.

RUMYANTSEV, E. A. (1989). Some aspects in the studies of fish parasite fauna in the lakes of different type. In Bauer, O. N. (Ed). *Parasites of Freshwater Fishes of N. W. Europe* (ed. Bauer, O. N.) (Materials of the International Symposium of Soviet-Finnish Cooperation, 10–14 January, 1988), pp. 123–130.

SANTAMARINA, M. T., TOJO, J., UBEIRA, F. M., QUINTEIRO, P. & SANMARTIN, M. L. (1991). Anthelminthic treatment against *Gyrodactylus* sp. infecting rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms* **16**, 39–43.

SCHMAHL, G. (1988). Chemotherapy and control (*Gyrodactylus* and *Pseudodactylogyrus*). *First International Symposium on Monogenea, 7–13 August, Ceské Budejovice, Czechoslovakian Academy of Sciences*, p. 49.

SHINN, A. P. (1994). The application of new biosystematic techniques in the discrimination of the genus Gyrodactylus (Monogenea) on salmonoid fish. Ph.D. thesis, University of Stirling.

SHINN, A. P., DES CLERS, S. A., GIBSON, D. I. & SOMMERVILLE, C. (1996). Multivariate analysis of morphometrical features from *Gyrodactylus* spp. (Monogenea) parasitising British salmonids: Light microscope based studies. *Systematic Parasitology* **33**, 115–125.

SHINN, A. P., GIBSON, D. I. & SOMMERVILLE, C. (1993). An SEM study of the haptoral sclerites of the genus *Gyrodactylus* Nordmann, 1832 (Monogenea) following extraction by digestion and sonication techniques. *Systematic Parasitology* **25**, 135–144.

SHINN, A. P., KAY, J. W. & SOMMERVILLE, C. (2000). The use of statistical classifiers for the discrimination of species of the genus *Gyrodactylus* (Monogenea) parasitising salmonids. *Parasitology* **120**, 261–269.

SHINN, A. P., SOMMERVILLE, C. & GIBSON, D. I. (1995). Distribution and characterization of species of *Gyrodactylus* Nordmann, 1832 (Monogenea) parasitizing salmonids in the UK, and their discrimination from *G. salaris* Malmberg, 1957. *Journal of Natural History* **29**, 1383–1402.

S-PLUS 4. (1997). *Seattle: Data Analysis Products Division, Mathsoft*.

TANUM, K. (1983). Studier av taksonomi og vertsforhold hos *Gyrodactylus* – arter på laksefisk av slektene *Salmo* og *Salvelinus* I Norge-Hovedfagsoppgave i spesiell zoologi. Thesis, University of Oslo.

TESARCIK, J. & IVASIK, V. (1974). Comparison of the helminth fauna of trouts in the River Moravice and in carpathian ponds. *Helminthologia* **15**, 773–778.

VENABLES, W. N. & RIPLEY, B. D. (1997). *Modern Applied Statistics with S-Plus*. Springer-Verlag, Berlin.

ZITNAN, R. (1974). Acclimatization of fish in the carpathian region of Czechoslovakia and the role of helminths in this process. *Ichthyologia* **6**, 143–155.