

Multi-disciplinary fingerprints reveal the harvest location of cod *Gadus morhua* in the northeast Atlantic

Ruth M. Higgins^{1,11,*}, Bret S. Danilowicz^{1,2}, Juan A. Balbuena³,
Anna K. Daniélsdóttir^{4,10}, Audrey J. Geffen⁵, Wim G. Meijer⁶, Johan Modin^{7,†},
Francisco E. Montero^{3,9}, Christophe Pampoulié⁴, Diana Perdiguero-Alonso³,
Arnd Schreiber⁸, Magnús Ö. Stefánsson⁴, Bryan Wilson^{1,6}

¹School of Biological and Environmental Sciences, University College Dublin, Belfield, Dublin 4, Ireland

²College of Science and Technology, Georgia Southern University, Statesboro, Georgia 30458, USA

³Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, PO Box 22085, 46071 Valencia, Spain

⁴Marine Research Institute, Skúlagata 4, 101 Reykjavik, Iceland

⁵Department of Biology, University of Bergen, PO Box 7803, 5020 Bergen, Norway

⁶School of Biomolecular and Biomedical Sciences, University College Dublin, Belfield, Dublin 4, Ireland

⁷Coastal Research Institute, Swedish Board of Fisheries, 74071 Öregrund, Sweden

⁸University of Karlsruhe, Adenauerring 20, Geb. 50.40, 76131 Karlsruhe, Germany

⁹Department of Animal Biology, Plant Biology, and Ecology, Autonomous University of Barcelona, Campus Universitari, 08193 Bellaterra, Barcelona, Spain

¹⁰Matis, Skúlagata, 101 Reykjavik, Iceland

¹¹Instituto do Mar, Universidade dos Açores, Horta, 9901-862, Portugal

ABSTRACT: Using multiple biological markers to establish the fingerprint of a harvest location, individual cod *Gadus morhua* L. can be classified to their population of origin without error. A combined approach to classification using otolith microchemistry, otolith shape analysis, body morphometry, microbacterial assemblages, internal and external parasites, and microsatellite DNA was found to be more powerful than by any single technique. Binomial and multinomial logistic regression analyses were used to distinguish wild from farmed fish and subsequently to determine the precise harvest origin of each individual. Two new approaches were used: one focusing on optimal or key variables from each discipline and the other using probability values derived on a technique-by-technique basis. Cod from widely separated origins were classified with high (up to 100% correct) placement success. Focusing on the placement of individual fish, this study represents a decisive advance toward identifying fish harvested from protected populations.

KEY WORDS: Traceability · Discrimination · Cod · Biological marker · Molecular marker · Otolith · Morphometry

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Many populations of fish are overexploited (Christensen et al. 2003), which has detrimental socio-economic consequences for nations and communities reliant on sustainable fisheries (Jackson et al. 2001, Pauly et al. 2002, Pauly et al. 2003). Atlantic cod (*Gadus morhua* L.) is one of the most important com-

mercial fish species in the north Atlantic region, having strong social and cultural associations (Kurlansky 1999). Reductions in numbers of cod due to changing environmental conditions and overexploitation have been detected in several regions including the east coast of Canada (O'Driscoll et al. 2000, Bundy 2001), the Baltic Sea (Jonzen et al. 2002), and the North Sea (Cook et al. 1997, Hutchinson et al. 2003). As over-

*Email: ruth@uac.pt

†Deceased

fished populations recover slowly (Hutchings 2000), removing fishing pressure from such populations is critical for their recovery.

Stock separation and population discrimination techniques have been developing rapidly in fisheries research (e.g. Nielsen et al. 2001, Thorrold et al. 2001). The core principle of discrimination is that individuals from disparate areas (or from discrete stocks or populations) display different characteristics for a given individual trait. Useful grouping factors and identifiers for Atlantic cod include: genetic markers (Bentzen et al. 1996, Jonsdóttir et al. 1999, Ruzzante et al. 1999), life history traits (Begg 2005); meristic (Templeman 1981, Pepin & Carr 1993, Swain et al. 2001); otolith elemental composition (Campana et al. 1999, Gibb et al. 2007), otolith shape (Campana & Casselman 1993, Galley et al. 2006, Stransky et al. 2008), parasite assemblages (Larsen et al. 1997, MacKenzie & Abaunza 2005), and morphometric characters (Pepin & Carr 1993, Swain et al. 2001). Whereas several individual traits may be suitable indicators of the geographic origins of individual fish, it is most likely that the ideal fingerprint is a combination of several characters. Genetics, for example, are useful proxies of the long-term distribution patterns in cod (Ruzzante et al. 1999), whereas features such as otolith chemistry and parasites reflect conditions to which the individual fish has been exposed, reflecting events that occurred during the lifetime of the individual (Williams et al. 1992, Gibb et al. 2007). Thus, multi-disciplinary approaches are probably the best option for fish discrimination and traceability (Waldman 1999, Cadrin et al. 2005), but their potential has rarely been empirically explored.

Traceability of fish is an allied concept to stock discrimination, but it goes further, as it exploits known or suspected differences between fish from different areas to determine the region or stock from which any one individual may have been derived. Traceability techniques can therefore be used to determine the origins of individual fish for which the harvest location is in doubt (Primmer et al. 2000), and, thus, the foremost application of traceability lies in the identification and verification of illegal fishing activity (DEFRA 2004). Further, given increased farming activity that may lead to farmed fish being misrepresented as wild-caught fish, traceability can also benefit consumers through the detection of fraudulent claims to the origin of seafood products (Jacquet & Pauly 2008).

Stock discrimination techniques can be employed for traceability purposes (Nielsen et al. 2001, Thorrold et al. 2001). Nielsen et al. (2001) showed that some specific cod stocks could be traced to their origins with 96 to 100% success using genetic microsatellite tags alone. However, recently established populations, indicated by weak gene genealogies (Carr et al. 1995,

Árnason 2004) and/or weak genetic differentiation (Pampoulie et al. 2008), could be difficult to recognise genetically and would be better distinguished using biological methods that detect variability in the phenotype of the organism rather than the genotype (Swain and Foote 1999). In such circumstances, otolith elemental signatures, morphometric differences, and other such potential markers of origin are likely to enhance discrimination.

The present study brings together 6 different biological techniques to establish a method for determining the harvest origins of individual cod. We evaluate the ability of each technique both singly and in conjunction as identifiers of these origins. We also attempt to define the smallest set of variables that could place fish accurately to their geographic origins. By applying 6 different techniques to the same individual fish, the present study represents one of the first multidisciplinary studies of population structure, and the first to apply a combined marker analysis for traceability.

MATERIALS AND METHODS

Collection of fish and processing. Atlantic cod ($n = 1133$) were collected from 7 areas in northern Europe: 5 open water locations (Baltic, Celtic, Irish, and North Seas and Icelandic waters) and 2 fish farms (Iceland, Scotland) (Fig. 1). The Skagerrak region was considered to belong to the North Sea in accordance with current demarcation by the International Council for Exploration of the Sea (ICES 2005). Farmed fish samples were collected from indoor systems at the Experimental Marine Fishfarm of the Marine Research Institute of Reykjavík, Iceland, and Orkney Marine Hatcheries Ltd, Scotland (see supplement at www.int-res.com/articles/suppl/m404p197_app.pdf).

Sampling occurred during spring 2002 (Feb to May), autumn 2002 (Aug to Dec.) and spring 2003 (Jan to May). Spring aggregations were composed of a greater number of spawning fish (Fig. 2) and were therefore more likely to be representative of the stock in each region. Contrarily, the autumn samples were understood to have been composed of mixed-stock feeding aggregations, representing several spawning stocks. This sampling strategy allowed us to determine an approach for individual fish from each region.

A biological fingerprint of each individual fish origination from each of the sampled regions was determined using 6 techniques: body and otolith morphometrics, microsatellite loci, otolith chemical composition, bacteria, and parasite assemblages. In total, 372 variables were extracted from each individual fish (see supplement for details on the methodology for each technique).

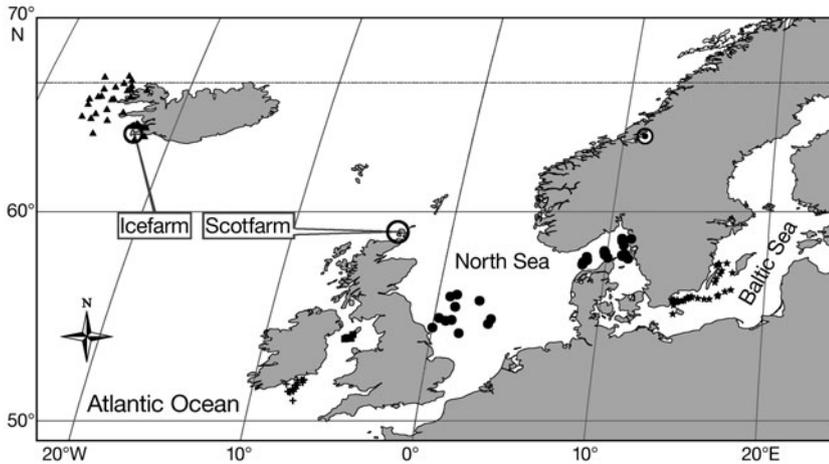


Fig. 1. Locations at which fish were sampled. Wild origins were: Icelandic waters (▲), Celtic Sea (+), Irish Sea (■), North Sea (●) and Baltic Sea (★). Flag labels indicate farmed origins. nd:

Statistical analysis. All fish ($n \geq 564$) sampled from each harvest region were used for analysis by each individual method. The full complement of samples was used in order to optimise the performance of each individual approach, giving individual techniques an advantage over the combined statistical approach that followed (see below) with respect to the number of samples analysed. A selection of biological markers was determined from each individual technique based on expertise and standard statistical approaches (see supplement & Table 1), and these selected biological markers were combined across techniques and used in subsequent analyses.

This initial analysis by each individual technique was used to identify the top 10

variables from that technique most likely to be useful as biological tags (Table 1, and see statistical analysis below), resulting in a total of 60 variables per fish for combined testing (i.e. 10 variables for each of the 6 techniques). For the multi-disciplinary statistical analysis, a prerequisite was the availability of all 60 of these variables from each individual fish, which reduced the number of fish available for multi-disciplinary analysis to 482 (Table 2). First-season Icelandic-farmed fish were not included in this study since samples were collected later than anticipated and as such could no longer be considered alongside other spring samples.

Each cod was classified as being of either wild or farmed origin using cross-validated probabilities derived from logistic regression (binomial classification). Subsequently, fish of wild origin were assigned to 1 of the 5 wild popula-

Table 1. Statistical tool used to determine probabilities for each biological technique. DFA: discriminant function analysis; MLP: multi-layer perceptron. Stepwise procedures were used to select the 10 most useful variables for subsequent analysis. na: not available

Technique	Statistical approach	Stepwise procedure
Microbiology	DFA	Stepwise DFA
Microsatellites	Bayesian assignment (Structure)	na
Body morphometrics	DFA	Stepwise DFA
Otolith elemental composition	DFA	Stepwise DFA
Otolith morphometry	DFA	Stepwise DFA
Parasitology	Neural network (MLP)	Stepwise DFA

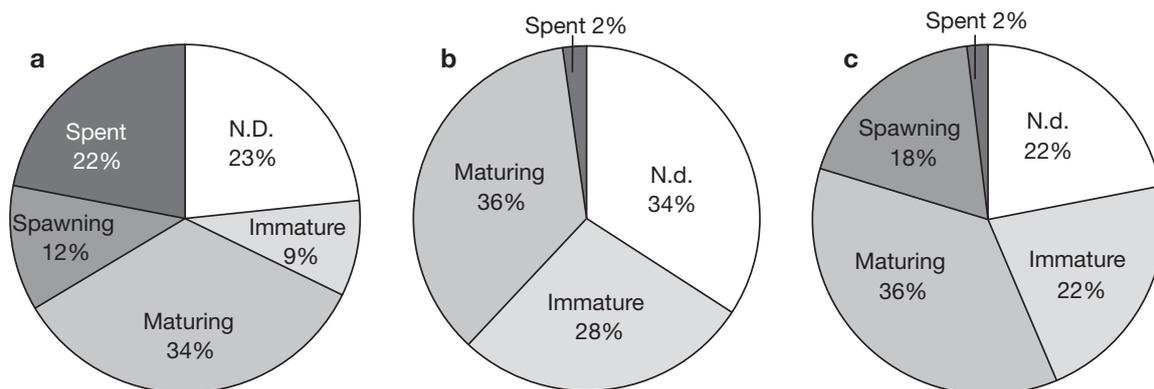


Fig. 2. Maturity status of samples from each season: (a) spring 2002, (b) autumn 2002, (c) spring 2003. N.d.: not determined

Table 2. Sample sizes by area and year for fish included in all analyses. In each region, 60 fish were sampled per season. Fish were only included in the study if viable information was obtained for all techniques. Samples for the first season, spring 2002, were poor for many regions due to logistical difficulties

Region	Spring 2002	Autumn 2002	Spring 2003	Grand total
Baltic Sea	14	16	32	62
Celtic Sea	14	44	23	81
Iceland farm		28	28	56
Icelandic waters	18	22	24	64
Irish Sea	17	29	14	60
North Sea	18	25	26	69
Scottish farm	32	31	27	90
Grand total	113	195	174	482

tions (multinomial classification), while those of farmed origin were assigned to 1 of the 2 potential sources (Icelandic vs. Scottish farm).

Given the number of biological variables collected for this study (despite the removal of noisy variables from each technique), and to avoid over-fitting of the regression model, 2 novel methods were considered for utilising the available data: an optimal variable approach and a probabilities approach (Higgins 2005, Stefánsson et al. 2009).

The optimal variables were selected in a series of steps. Initially, 10 variables from each technique were selected using analyses performed on data from that technique alone (Table 1). Subsequently, 60 variables were used in a combined model, comprising: 10 body morphometric measurements, 10 otolith morphometrics, 10 otolith elemental signatures from the core of the otolith, 10 otolith elemental signatures from the edge of the otolith, 10 microbial signatures, and 10 parasite species. These variables were analysed in concert in a stepwise logistical regression model, and the best overall variables were selected for each step in the classification hierarchy (wild vs. farmed, wild origins only, and farmed origins only).

The probabilities approach integrated data generated by each biological technique, representing the likelihood of every individual fish being harvested from each of the given source populations, using classification by that technique alone. Each biological technique was analysed using a conventional statistical tool for that data type (Table 1), and the resulting probabilities of belonging to each population (as predicted by that technique) were combined in a single classification analysis. For each technique, probabilities were derived for each level of the classification hierarchy. In the first instance, 2 probabilities were derived; the probability of being farmed, and the probability of being wild in origin. In a second analysis, 2 fur-

ther probabilities were derived for the individuals determined to be of farmed origin in the preceding step; the probability of being Icelandic-farmed and the probability of being Scottish-farmed. In the final step, individuals identified as being wild in step 1 were assigned a probability value of being from 1 of the 5 wild origins, resulting in 5 probability values.

Clearly, in terms of the derived probabilities, the binomial choices demonstrated multicollinearity; therefore only one of the provided probabilities was used in the analysis, for example in determining the wild/farmed classification, only the probability (p) of the individual being wild was considered, since the probability (p') of the fish being farmed was $p' = 1 - (p)$. Similarly for farmed fish, only the probability of being Icelandic was included, since the probability of being Scottish was $1 - p(\text{Icelandic})$.

Interpretation of the most useful variables was slightly modified for the multinomial decisions (distinguishing between wild origins) under the probabilities approach. The probabilities for each technique were considered to be a unit, and therefore if one probability was needed to place individuals, all probabilities were considered necessary for that technique. In the case of microsatellites for example, where p (Baltic) and p (Irish) were selected as being the most useful variables, all microsatellite probabilities were retained in the final statistical model.

For the placement of individuals binomial logistic regression was performed using SPSS 10.0, and multinomial logistic regression was carried out using Genstat 6.1 (VSN International) (Payne 2002).

RESULTS

Cod from widely separated origins were classified with high success using both the optimal variable approach and the probabilities approach. Using individual optimal variables, at least 92.8% of the individuals were correctly placed back to their harvest origins (Table 3). The probabilities approach exceeded this level of classification accuracy by correctly identifying the origins of every individual analysed in this study, i.e. 100%. Both of the approaches used on combined biological techniques provided better classification than any one technique used in isolation (Table 3).

The hierarchy used in classification allows for the clear identification of difficult classification decisions in the present study. Wild and farmed individuals, for example, were clearly identified using both the optimal variable and probabilities approaches, with the origins of each fish being correctly attributed (Table 3). The distinction between farms was also highly successful, with the optimal variable technique being

Table 3. Cross-season stable variables selected for classification for each test. WF: Wild or farmed origin; FF: farm of origin; WW: wild location of origin. S1, S2, and S3 are the levels of classification success for each of season 1, season 2, and season 3 for the group of variables listed. For optimised variables, number in parentheses represents the number of markers needed from each technique (see Table 5 for details)

Test	Optimised variables	S1	S2	S3	Derived probabilities	S1	S2	S3
WF	Microsatellites Body morphometrics (1) Otolith elemental concentration (1) Parasite species (4)	100	100	100	Body morphometrics Otolith morphometry Parasites	100	100	100
FF	Microsatellites Microbiology (1)	–	100	98.2	Body morphometrics Otolith morphometry	–	100	100
WW	Microsatellites Body morphometrics (5) Otolith elemental concentrations (7) Microbiology (1) Otolith morphometry (1) Parasite species (6)	100	92.8	99.1	Microbiology Microsatellites Body morphometrics Otolith morphometry Parasites	100	100	100

marginally less useful than the probabilities approach. Considering that in a binomial classification the classification by chance alone equates to 50% correct classification, it becomes clear that both the optimal variable and the probabilities approaches are highly reliable methods of determining the harvest origin for the farmed fish considered here. With the determination of the wild harvest locations, probabilities continued to perform to the highest level correctly placing all of the fish in question. The optimal variable technique was once again slightly less accurate, being 92.8% in season 2 and 99.1% in season 3.

The variables on which the classification analyses were based are outlined in Table 4. Relatively few variables were needed to achieve the high classification accuracy. It was possible to determine the origins of wild and farmed cod using both the derived probabilities and the optimised set of variables from each technique. The list of variables, stable across all seasons and each type of data (optimised variables or derived probabilities), is shown in Table 5.

In terms of the variables required to obtain these high levels of classification, wild and farmed fish (a binomial choice) were best distinguished with logistic regression using optimised variables consisting of microsatellite probabilities, one morphometric measure, one element from the core of the fish otolith, and several parasite species (Table 5). Where derived probabilities were used, wild and farmed individuals were best distinguished with morphometric measurements of the entire body

of the fish, otolith morphometry, and parasites using logistic regression.

The 2 groups of farmed fish (a binomial choice) could be separated with microsatellite probabilities in combination with 1 parasite species and 1 bacterial species (Table 5). With derived probabilities, body morphometrics in combination with otolith morphometry were sufficient.

Wild origins were more difficult to distinguish (a multinomial choice), requiring a greater number of biological markers than either of the binomial choice tests. Several morphometric measurements from the body of the fish, in addition to several otolith elemental concentrations, an otolith morphometric parameter,

Table 4. Percentage of individual cod correctly classified using probability data based on multiple biological techniques (MBT) compared to the separate analysis for bacterial assemblage (BA), body morphometry (BM), microsatellite loci (ML), otolith chemistry (OC), otolith morphometry (OM), and parasite assemblages (PA). Using the same model across seasons and years, each approach was optimised to best distinguish between farmed and wild fish (WF), between farms (FF), and among wild locations (WW). N: minimum number of cod analysed for each test. na: only one farm sampled this season. *Uses Celtic, Irish, and North Seas as a single population, and analysed a minimum of 50 cod per population

Test	Factors		N	Classification						
	Season			MBT	BA	BM	ML*	OC	OM	PA
WF	1		113	100	86.3	na	92.1	79.4	92.6	100.0
	2		276	100	92.9	98.5	56.8	68.1	89.1	99.3
	3		175	100	92.5	96.1	71.2	77.8	89.3	100.0
FF	2		59	100	77.4	67.7	95.1	84.0	93.3	54.2
	3		55	100	71.1	70.0	94.1	93.9	98.9	52.7
WW	1		81	100	70.8	55.7	70.1	49.4	50.4	86.2
	2		217	100	63.7	73.6	75.1	60.9	70.5	89.5
	3		120	100	71.7	52.6	48.2	68.8	48.5	85.7
Overall			564	100	71.7	61.3	70.1	66.9	68.0	80.4

Table 5. Cross-season variables selected for classification for each test, using microsatellite probabilities, morphometric measurements (m), bacterial species (gm), elemental compositions of the core (c), and edge (e) of the otolith, otolith morphometry harmonics (Fdreal), and parasite species. Other abbreviations as in Table 4. Information in parentheses is provided to help identify the technique to which the marker belongs. Further information on markers is available in the supplement at www.int-res.com/articles/suppl/m404p197_app.pdf, including a map of morphometric measurements (Fig. S1), an explanation of morphometric harmonics, and otolith elemental analysis. Bacterial species were identified by a T-RFLP peak resulting in numerical nomenclature, e.g. 456 (gm)

	Optimised variables	Derived probabilities
WF	Microsatellites	Body morphometrics
	ab (m)	Otolith morphometry
	Mg24 (c)	Parasites
	<i>Contracaecum osculatum</i>	
	<i>Derogenes varicus</i>	
FF	Microsatellites	Body morphometrics
	456 (gm)	Otolith morphometry
	<i>Contracaecum osculatum</i>	
WW	Microsatellites	Microbiology
	bc (m)	Microsatellites
	di (m)	Body morphometrics
	fg (m)	Otolith morphometry
	gm (m)	Parasites
	ko (m)	
	Ba137 (c)	
	Cu63 (c)	
	Mg25 (c)	
	Mn55 (c)	
	Rb85 (c)	
	Rb85 (e)	
	Sr86 (e)	
	268 (gm)	
	Fdimag2	
	<i>Anisakis simplex</i>	
	<i>Caligus elongatus</i>	
	<i>Contracaecum osculatum</i>	
<i>Derogenes varicus</i>		
<i>Echinorhynchus gadi</i>		
<i>Hemiurus communis</i>		

and one bacterial species, were needed in this classification (Table 5). More probabilities were also needed for discriminating between fish from wild locations. Under the probabilities approach, wild populations were distinguished using body morphometrics, otolith morphometrics, microsatellites, parasites, and bacterial probabilities.

DISCUSSION

Classification studies in fisheries biology have been numerous and diverse. Although the literature is rich, it is difficult to draw any firm conclusions regarding

the ideal set of biological tags for tracing individuals of a given species back to their harvest origins, which has lead biologists to consider the potential of taking a more holistic perspective (Waldman 1999, Cadrin et al. 2005). In fact, most research has focused on only one biological approach, with few studies considering the potential of using several techniques in concert (e.g. Meng & Stocker 1984, Pepin & Carr 1993, Roldan et al. 2000). A set of biological characters and a set of biological techniques that are temporally stable for classification of Atlantic cod have been derived in the present study, and it has been shown that in order to obtain an efficient and accurate classification model for traceability of Atlantic cod, it is essential to combine information from disparate biological disciplines.

In the present study, statistical analysis for traceability was performed in a series of stages. The initial stage was to determine how well individual techniques differentiated fish from each source area. One of the stipulations we made for this approach was that the full complement of samples would be used. For example, where data were available for body morphometry but not for otolith elemental analysis (due to the otolith being aragonite, etc.), these fish were still included in classification using morphometry. In this way we intentionally gave individual techniques an advantage over the combined statistical approach. Despite increased rigour, combined methods consistently outperformed individual techniques.

Combined biological identifiers were evaluated using a 2-step hierarchical method. Initially, each cod was classified as being of either wild or farmed origin using cross-validated probabilities derived from logistic regression (binomial classification). Subsequently, if a fish was classified as being of wild origin, it was then assigned to 1 of the 5 wild populations (multinomial classification). Similarly, if a fish was determined to be of farmed origin, it was then placed to 1 of the 2 potential sources (Icelandic vs. Scottish farm). Due to restrictions in the software used, a jack-knife procedure was applied to validate the binomial logistic regression model, whereas the multivariate logistic regression model was validated using a separate holdout validation method as follows: the multivariate predictive model was constructed using 80% of the data, with the remaining 20% used for validation, rotating the training and validation sets. This ensured that each individual was included in a validation set while also taking care not to use that individual when constructing the model in the same instance. The mean classification using these rotated sets was taken as the classification success.

The optimal variable approach used a subset of the variables available in their respective forms (standardised for age and standard length where appropriate),

with the exception of microsatellite data. Due to their co-dominant nature, microsatellite data did not combine legitimately with non-genetic data, and could have been compromised by treating it in such a manner. Therefore, genetic data were analysed with the software Structure (Pritchard et al. 2000), and the raw genetic data were used to produce a series of 7 probabilities, representing the likelihood of inclusion in each source population. Structure uses a clustering approach to classification and consequently, on occasion, individuals from different sampling locations were determined to cluster into a single group. The group North/Celtic for example, composed of individuals from both the North Sea and the Celtic Sea, was identified for the season 1 placement of wild fish, indicating that these populations were not distinguishable using the selected microsatellite markers (see Pampoulie et al. 2008). Structure allows the number of populations to be predefined (there were 5 populations for wild-caught fish), thus the joined category was represented twice in the classification results for microsatellites, once in place of the North Sea, and once in place of the Celtic Sea. In cases such as these, the mean value of the 2 categories was calculated, and this mean value was taken to be the probability of having come from each area in isolation. As clustering with Structure is mainly linked to the genetic divergence of the samples used, determining binomial probabilities of a fish being either farmed or wild in origin proved challenging: genetic divergence did not naturally fall into just 2 groups. Therefore, 7 predefined origins were used for the farm-wild decision, representing each origin separately. The probability of being of farmed or wild origin was then calculated by summing the probabilities of being from each of the individual farms or wild localities, respectively.

Reducing microsatellite data to a set of probabilities also benefited the classification model as the dimensionality of the input was considerably reduced. This method converted qualitative data (genotypes) into quantitative data. Subsequently, genotypes can be handled statistically as if they were quantitative data. A similar approach was taken by Stefánsson et al. (2009) where genotypes (converted to quantitative data with Structure) were correlated with spatial dimensions. In the current application, genetic probabilities were combined with the chosen biological markers from each of the remaining techniques.

Previous studies have reported varying levels discriminatory success among fish populations. In some cases, high levels of classification have been reported: up to 96% for seabream (Palma & Andrade 2002) and bogue (Power et al. 2005), and 96 to 100% for cod (Nielsen et al. 2001), using morphometrics, parasites, and microsatellites respectively. Although relatively

successful, these studies did not test the strength of their technique in a broader context. In the present study, by contrast, it was possible to place wild Atlantic cod with 100% accuracy to origins across a broad geographic area of northern Europe, which included populations that no technique alone could discriminate with more than 89.5% accuracy (Table 3, see WW season 2).

Farmed and wild fish clearly differ in several respects due to sharp contrasts in their environmental conditions, since every aspect of the environment of farmed individuals is controlled by direct human intervention. Water quality, diet, photoperiod, etc., affect the phenotype of farmed individuals. Consequently, during the present study, it was found that farmed fish had distinctly fewer parasites than wild caught individuals, which could be due to regular anti-parasitic treatments at the farms, artificial feeds, etc. Such clear differences should be utilised to distinguish fish from farmed and wild origins, a distinction that has been known to be of interest to consumers (McCarthy 2002). For similar reasons the ability to distinguish between farmed origins can be useful, which can be readily done, as shown in the present study.

Why particular markers are better suited to each level of separation is an interesting subject. In some cases it was clear why certain markers performed best in separating fish from different sources. In separating wild from farmed fish, for example, parasites were virtually absent from the tissues sampled from both fish farms. Where parasites were present, a number of key species clearly separated the 2 groups without error. Although no clear reason can be established for the usefulness of morphometric characters in the wild vs. farmed decision, we can conjecture that the plasticity of morphological development is affected by the artificial environment, manipulated diet, etc., and there is some evidence of such whole body (e.g. Hard et al. 2000, Wintzer & Motta 2005, Janhunnen et al. 2009) and otolith (Tomas & Geffen 2003, Hussy 2008) morphological responses to captivity by other species.

From a conservation perspective it is often more relevant to distinguish individuals from several wild origins. Illegal fishing is a common issue, which can be difficult both to detect and to prove (Jensen & Vestergaard 2002). Our methods offer a tool with which to prove or disprove offered testimonies. This tool is dependent upon a multivariate approach and the availability of a broad range of biological samples. In some circumstances it may be impossible to use individual variables from a range of disciplines, for example, if fish have been eviscerated at sea, visceral parasites will be missing. Long-lived parasites, such as the nematode larvae *Anisakis simplex* s.l and *Contracaecum osculatatum*, encyst both in visceral connective tissue

and in muscle (Rokicki et al. 1993, Strømnes & Andersen 1998) and could be useful characters in eviscerated fish. In any classification study, the resources available dictate the approach taken. We have shown the potential of using whole techniques in concert as well as selecting particular available variables from a range of disciplines, both approaches being entirely reliable.

Knowing that the distribution of cod parasite species varies across our study area (Perdiguero-Alonso et al. 2008a) and that parasites alone have proved useful in distinguishing cod from the NE Atlantic (Perdiguero-Alonso et al. 2008b), we were confident that parasite assemblages would contribute significantly to a multidisciplinary traceability study. Limited by the fact that some parasites are short-lived and likely to vary annually in infection, we expected that classification by parasites alone would be greatly improved when combined with other markers.

In wild individuals, morphometric characters that distinguished regions included the size of the eye in addition to the shape of the belly. Since morphometric development is determined by very diverse factors including genetic distinctiveness of populations, presence of life age or life history variants, and number of collection grounds in any sea basin, further interpretation of these markers would require dedicated investigation.

Our expectation for otolith microchemistry results were based on published accounts of experimental manipulation of water chemistry, salinity, and temperature affecting otolith composition (Bath et al. 2000, Elsdon & Gillanders 2004, de Vries et al. 2005). Elements in the otolith are influenced in different ways, with some showing a proportional response to environmental concentrations and some influenced by individual fish physiology and temperature. Since the fingerprint is a combination of these factors, with the otolith core representing the area of spawning and the otolith edge representing the most recently inhabited area, we expected to see strong differences due to physical factors such as lower salinity in the Baltic leading to lower Sr and higher Ba in these fish because of the higher terrestrial influence. We expected that the fingerprint of Icelandic cod would also be distinct because there is little evidence that these fish move into the other basins considered in the present study.

The sensitivity of bacterial communities to environmental change has been documented for decades, and the extension of this to traceability is a straightforward one. An area of water the size of the northeastern Atlantic will have a variable physicochemical environment with changing latitude, and therefore the composition of the corresponding bacterial communities differ as a consequence of environmental changes. Biogeographical analyses and traceability of bacterial

populations are quite commonplace in terms of infection and disease. Although biogeography of bacterial communities over large spatial scales has not been extensively researched, and certainly not in relation to fish, there are examples in the published literature (Wilson et al. 2008, Smith et al. 2009) where a relationship between the types of bacteria associated with fish and the fish's harvest location have been indisputably established; their traceability is an obvious application of this technology, as ably demonstrated in the present study.

In undertaking this investigation we took into account logistical considerations to facilitate later application of the method in monitoring and policing commercial fisheries. We aimed to define a strategically simple sampling protocol that could be undertaken on board commercial vessels by non-specialists with minimal training. Fish viscera, which are usually removed at sea in commercial operations, would be simple to collect and ship. Similarly, gill clippings used for genetic analysis and mucous samples used for bacteriology are easily collected and stored in small vials without affecting the appearance or value of commercially caught fish. In defining a specific series of markers (Table 5), we have provided a number of methods of accurately placing cod from this region. For example, in determining if a fish is of wild or farmed origin it is sufficient to examine its body morphometry and otolith morphometry alongside internal and external parasites using the complete complement of markers from each of those techniques. Alternatively, one might look at our 9 microsatellite loci along with a single body metric (in this case the distance from eye orbit to snout tip; see supplement for diagram), the core concentration of magnesium from the otolith and a selection of key parasite species. We understand, however, that given the spatial and temporal limitations of this research, more specific baseline data would need to be collected and regularly updated.

Although the techniques used here are not new, with the exception of the microbiological technique (Wilson et al. 2008, Smith et al. 2009), our integrated approach showed that even more powerful traceability can be achieved for a single fish species using multivariate analyses. Indeed, most of the techniques used here have a long and distinguished history as tools for discrimination of fishes (Campana & Thorrold 2001, Hansen et al. 2001, MacKenzie 2002). Although the variables required for classification could differ, as may the geographical scale, this multivariate approach can be readily adapted to other fish and shellfish species and can protect consumers from inappropriate labelling of the origin of seafood. Given that areas protected from fishing continue to be established, classification of individual fish to harvest locations could en-

hance enforcement of fishing bans, required for the recovery of depleted stocks (Pauly et al. 2002, Power et al. 2005, Sale et al. 2005).

It is important to note that this research considered individuals from the northeast Atlantic region only. It has been seen, however, that in at least in some respects there is little differentiation between cod from across the Atlantic Ocean (e.g. Pampoulie et al. 2008). It would be very informative to apply this approach to a greater area and attempt to distinguish eastern from western Atlantic cod.

In conclusion, our multivariate approach to classification was reliable and stable across both seasons and years. This represents a crucial step towards setting a legal standard for identifying the harvest location of a fish, be it a determination of wild or farmed origin, or the specific population of origin.

Acknowledgements. This research was funded by grants from the EU grant CODTRACE (Q5RS-2000-01697). Additional support was provided by Enterprise Ireland and University College Dublin. A debt of gratitude is owed to the many students and technical staff involved in sampling missions and laboratory work, in addition to employees of the 2 fish farms. This article is dedicated to the memory of our good friend and colleague, Johan Modin. Some ducks stay in the farmyard while others are wild and fly free... you were a wild duck.

LITERATURE CITED

- Árnason E (2004) Mitochondrial cytochrome b DNA variation in the high fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics* 166:1871–1885
- Bath GE, Thorrold SR, Jones CM, Campana SE, McLaren JW, Lam JWH (2000) Strontium and barium uptake in aragonitic otoliths of marine fish. *Geochim Cosmochim Acta* 64: 1705–1714
- Begg GA (2005) Life history parameters. In: Cadrin SX, Friedland KD, Waldman JR (eds) *Stock identification methods*. Elsevier, London
- Bentzen P, Taggart CT, Ruzzante DE, Cook D (1996) Microsatellite polymorphism and the population structure of Atlantic cod (*Gadus morhua*) in the northwest Atlantic. *Can J Fish Aquat Sci* 53:2706–2721
- Bundy A (2001) Fishing on ecosystems: the interplay of fishing and predation in Newfoundland-Labrador. *Can J Fish Aquat Sci* 58:1153–1167
- Cadrin SX, Friedland KD, Waldman JR (eds) (2005) *Stock identification methods: applications in fisheries science*. Academic Press, London
- Campana SE, Casselman JM (1993) Stock discrimination using otolith shape analysis. *Can J Fish Aquat Sci* 50: 1062–1083
- Campana SE, Thorrold SR (2001) Otoliths, increments, and elements: keys to a comprehensive understanding of fish populations. *Can J Fish Aquat Sci* 58:30–38
- Campana SE, Chouinard GA, Hanson JM, Frechet A (1999) Mixing and migration of over-wintering Atlantic cod (*Gadus morhua*) stocks near the mouth of the Gulf of St. Lawrence. *Can J Fish Aquat Sci* 56:1873–1881
- Carr SM, Snellen AJ, Howse KA, Wroblewski JS (1995) Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus morhua*) from bay and offshore locations on the Newfoundland continental shelf. *Mol Ecol* 4:79–88
- Christensen V, Guénette S, Heymans JJ, Walters CJ, Watson R, Zeller D, Pauly D (2003) Hundred-year decline of north Atlantic predatory fishes. *Fish Fish* 4:1–24
- Cook RM, Sinclair A, Stefansson G (1997) Potential collapse of North Sea cod stocks. *Nature* 385:521–522
- de Vries MC, Gillanders BM, Elsdon TS (2005) Facilitation of barium uptake into fish otoliths: influence of strontium concentration and salinity. *Geochim Cosmochim Acta* 69: 4061–4072
- DEFRA (2004) DNA tests reveal fish culprits. *Mar Pollut Bull* 48:1012–1015
- Elsdon TS, Gillanders BM (2004) Fish otolith chemistry influenced by exposure to multiple environmental variables. *J Exp Mar Biol Ecol* 313:269–284
- Galley EA, Wright PJ, Gibb FM (2006) Combined methods of otolith shape analysis improve identification of spawning areas of Atlantic cod. *ICES J Mar Sci* 63:1710–1717
- Gibb FM, Gibb IM, Wright PJ (2007) Isolation of Atlantic cod (*Gadus morhua*) nursery areas. *Mar Biol* 151:1185–1194
- Hansen MM, Kenchington E, Nielsen EE (2001) Assigning individual fish to populations using microsatellite DNA markers. *Fish Fish* 2:93–112
- Hard JJ, Berejikian BA, Tezak EP, Schroder SL, Knudsen CM, Parker LT (2000) Evidence for morphometric differentiation of wild and captive reared adult coho salmon: a geometric analysis. *Environ Biol Fishes* 58:61–73
- Higgins RM (2005) Optimising classification analyses for the traceability of individual fish within the north-eastern Atlantic Ocean. PhD dissertation, University College Dublin
- Hussy K (2008) Otolith shape in juvenile cod (*Gadus morhua*): ontogenetic and environmental effects. *J Exp Mar Biol Ecol* 364:35–41
- Hutchings JA (2000) Collapse and recovery of marine fishes. *Nature* 406:882–885
- Hutchinson WF, van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). *Proc Biol Sci* 270:2125–2132
- ICES (2005) Spawning and life history information for North Atlantic cod stocks. ICES Cooperative Research Report, no. 274
- Jackson JBC, Kirby MX, Berger WH, Bjorndal KA and others (2001) Historical overfishing and the recent collapse of coastal ecosystems. *Science* 293:629–637
- Jacquet JL, Pauly D (2008) Trade secrets: renaming and mislabeling of seafood. *Mar Policy* 32:309–318
- Janhunen M, Peuhkuri N, Piironen J (2009) Morphological variability among three geographically distinct arctic charr (*Salvelinus alpinus* L.) populations reared in a common hatchery environment. *Ecol Freshw Fish* 18: 106–116
- Jensen F, Vestergaard N (2002) Moral hazard problems in fisheries regulation: the case of illegal landings and discard. *Resour Energy Econ* 24:281–299
- Jonsdóttir ODB, Imsland AK, Danielsdóttir AK, Thorsteinsson V, Naevdal G (1999) Genetic differentiation among Atlantic cod in south and south-east Icelandic waters: synaptophysin (Syn I) and haemoglobin (HbI) variation. *J Fish Biol* 54:1259–1274
- Jonzen N, Cardinale M, Gardmark A, Arrhenius F, Lundberg P (2002) Risk of collapse in the eastern Baltic cod fishery. *Mar Ecol Prog Ser* 240:225–233

- Kurlansky M (1999) Cod: a biography of the fish that changed the world. Vintage UK, London
- Larsen G, Hemmingsen W, MacKenzie K, Lysne DA (1997) A population study of cod, *Gadus morhua* L., in northern Norway using otolith structure and parasite tags. *Fish Res* 32:13–20
- MacKenzie K (2002) Parasites as biological tags in population studies of marine organisms: an update. *Parasitology* 124:S153–S163
- MacKenzie K, Abaunza P (2005). Parasites as biological tags. In: Cadrin SX, Friedland KD, Waldman JR (eds) Stock identification methods. Elsevier, London
- McCarthy T (2002) Is fish farming safe? *Time*, November 25, 2002
- Meng HJ, Stocker M (1984) An evaluation of morphometrics and meristics for stock separation of Pacific herring (*Clupea harengus pallasii*). *Can J Fish Aquat Sci* 41:414–422
- Nielsen EE, Hansen MM, Schmidt C, Meldrup D, Grønkjær P (2001) Population of origin of Atlantic cod. *Nature* 413:272
- O'Driscoll RL, Schneider DC, Rose GA, Lilly GR (2000) Potential contact statistics for measuring scale-dependent spatial pattern and association: an example of northern cod (*Gadus morhua*) and capelin (*Mallotus villosus*). *Can J Fish Aquat Sci* 57:1355–1368
- Palma J, Andrade JP (2002) Morphological study of *Diplodus sargus*, *Diplodus puntazzo*, and *Lithognathus mormyrus* (Sparidae) in the Eastern Atlantic and Mediterranean Sea. *Fish Res* 57:1–8
- Pampoulie C, Stefánsson MÖ, Danilowicz B, Daniélsdóttir AK (2008) Recolonisation history and large scale dispersal in the open sea: the case study of the North Atlantic Cod (*Gadus morhua* L.). *Biol J Linn Soc* 94:315–329
- Pauly D, Christensen V, Guenette S, Pitcher TJ and others (2002) Towards sustainability in world fisheries. *Nature* 418:689–695
- Pauly D, Alder J, Bennett E, Christensen V, Tyedmers P, Watson R (2003) The future for fisheries. *Science* 302:1359–1361
- Payne RW (2002). The guide to GenStat, release 6.1. Lawes Agricultural Trust, Rothamsted Experimental Station, Harpenden
- Pepin P, Carr SM (1993) Morphological, meristic, and genetic analysis of stock structure in juvenile Atlantic cod (*Gadus morhua*) from the Newfoundland Shelf. *Can J Fish Aquat Sci* 50:1924–1933
- Perdiguero-Alonso D, Montero FE, Raga JA, Kostadinova A (2008a) composition and structure of the parasite faunas of cod, *Gadus morhua* L. (teleostei: gadidae), in the north east Atlantic. *Parasit Vectors* 1, Issue 23
- Perdiguero-Alonso D, Montero FE, Kostadinova A, Raga JA, Barrett J (2008b) random forests, a novel approach for discrimination of fish populations using parasites as biological tags. *Int J Parasitol* 38:1425–1434
- Power AM, Balbuena JA, Raga JA (2005) Parasite infra-communities as predictors of harvest location of bogue (*Boops boops* L.): a pilot study using statistical classifiers. *Fish Res* 72:229–239
- Primmer CR, Koskinen MT, Piironen J (2000) The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. *Proc Biol Sci* 267:1699–1704
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rokicki J, Valter ED, Myjak P (1993) *Contraecum osculatum* (Nematoda, Anisakidae) in cod *Gadus morhua* L. from the Polish coast of the Baltic. *Acta Parasitol* 38:33–35
- Roldan MI, Perrotta RG, Cortey M, Pla C (2000) Molecular and morphologic approaches to discrimination of variability patterns in chub mackerel, *Scomber japonicus*. *J Exp Mar Biol Ecol* 253:63–74
- Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St. Lawrence. *Fish Res* 43:79–97
- Sale PF, Cowen RK, Danilowicz BS, Jones GP and others (2005) Critical science gaps impede use of no-take fishery reserves. *Trends Ecol Evol* 20:74–80
- Smith CJ, Danilowicz BS, Meijer WG (2009) Bacteria associated with the mucus layer of *Merlangius merlangus* (whiting) as biological tags to determine harvest location. *Can J Fish Aquat Sci* 66:713–716
- Stefánsson MO, Sigurdsson T, Pampoulie C, Daniélsdóttir AK, Thorgilsson B, Ragnarsdóttir A, Gislason D, Coughlan J, Cross T F, Bernatchez (2009) Pleistocene genetic legacy suggests incipient species of *Sebastes mentella* in the Irminger Sea. *Heredity* 102:514–524
- Stransky C, Baumann H, Fevolden SE, Harbitz A and others (2008) Separation of Norwegian coastal cod and northeast Arctic cod by outer otolith shape analysis. *Fish Res* 90: 26–35
- Strømmes E, Andersen K (1998) Distribution of whaleworm (*Anisakis simplex*, Nematoda, Ascaridoidea) L3 larvae in three species of marine fish; saithe (*Pollachius virens* (L.)), cod (*Gadus morhua* L.) and redfish (*Sebastes marinus* (L.)) from Norwegian waters. *Parasitol Res* 84:281–285
- Swain DP, Foote CJ (1999) Stocks and chameleons: the use of phenotypic variation in stock identification. *Fish Res* 43: 113–128
- Swain DF, Frank KT, Maillet G (2001) Delineating stocks of Atlantic cod (*Gadus morhua*) in the Gulf of St. Lawrence and Cabot Strait areas using vertebral number. *ICES J Mar Sci* 58:253–269
- Templeman WT (1981) Vertebral numbers in Atlantic cod, *Gadus morhua*, of the Newfoundland and adjacent areas 1947–1971, and their use in delineating cod stocks. *J Northwest Atl Fish Sci* 2:21–45
- Thorrold SR, Latkoczy C, Swart PK, Jones CM (2001) Natal homing in a marine fish metapopulation. *Science* 291: 297–299
- Tomas J, Geffen AJ (2003) Morphometry and composition of aragonite and vaterite otoliths of deformed laboratory reared juvenile herring from two populations. *J Fish Biol* 63:1383–1401
- Waldman JR (1999) The importance of comparative studies in stock analysis. *Fish Res* 43:237–246
- Williams HH, MacKenzie K, McCarthy AM (1992) Parasites as biological indicators of the population biology, migrations, diet, and phylogenics of fish. *Rev Fish Biol Fish* 2:144–176
- Wilson B, Danilowicz BS, Meijer WG (2008) The diversity of bacterial communities associated with Atlantic Cod *Gadus morhua*. *Microb Ecol* 55:425–434
- Wintzer AP, Motta PJ (2005) Diet-induced phenotypic plasticity in the skull morphology of hatchery-reared Florida largemouth bass, *Micropterus salmoides floridanus*. *Ecol Freshw Fish* 14:311–318