JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

J. Agric. Food Chem. 2009, 57, 5665–5672 5665
DOI:10.1021/jf900520h

Arginine Kinase Peptide Mass Fingerprinting as a Proteomic Approach for Species Identification and Taxonomic Analysis of Commercially Relevant Shrimp Species

Ignacio Ortea,*,† Benito Cañas,‡ Pilar Calo-Mata,§ Jorge Barros-Velázquez,§ and José M. Gallardo†

†Instituto de Investigaciones Marinas, Consejo Superior de Investigaciones Científicas (CSIC), Eduardo Cabello 6, E-36208 Vigo, Spain, ‡Analytical Chemistry Department, Universidad Complutense de Madrid, E-28040 Madrid, Spain, and §Analytical Chemistry, Nutrition and Food Science Department, School of Veterinary Sciences, University of Santiago de Compostela, E-27002 Lugo, Spain

A proteomic approach aimed at species identification and taxonomic analysis of shrimp species of commercial interest is presented. Six different species belonging to the order Decapoda were considered. Preliminary, two-dimensional gel electrophoresis (2-DE) analysis of the sarcoplasmic proteome revealed interspecific variability in the isoelectric point (p*I*) of arginine kinase. For this reason, arginine kinase spot was selected as a potential molecular marker and subjected to tryptic digestion followed by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) peptide mass fingerprinting (PMF) analysis. Arginine kinase PMF allowed the differentiation of the six species studied. Four samples of commercial origin obtained in local markets were analyzed to validate the methodology. The PMF cluster analysis also provided information about the phylogenetic relationships in these species. The application of this methodology may be of interest for the differentiation and taxonomic analysis of shrimp species complementing DNA-based phylogenetic studies.

KEYWORDS: Species identification; shrimp; MALDI-TOF MS; PMF; arginine kinase; proteomics

1. INTRODUCTION

Crustaceans belonging to the order *Decapoda* are of notable commercial interest. This order includes the superfamily Penaeoidea, which is the most important economic resource in the world's crustacean fishery and aquafarming industry (*1*, *2*). Identification of marine species is an issue of capital relevance for the seafood industry because of global commercial requirements concerning labeling and traceability (*3*, *4*), not only for the prevention of commercial fraud but also for the assessment of safety risks derived from the inadvertent introduction of any food ingredient that might be harmful to human health (*5*). Anatomical characters are particularly difficult to use for shrimp species differentiation because of their phenotypic similarities and because of the fact that in their processing the external carapace is often removed. Therefore, fast and reliable analytical tools are necessary for distinguishing between these closely related species.

Methods for species identification are currently based on DNA or protein analysis. Mitochondrial DNA (mtDNA) has been extensively used in PCR-based studies for fish species identification (*6*). Among the mtDNA targets, the 16S rRNA gene and the cytochrome oxidase I gene have been reported to serve as good interspecific markers in some crustacean species, although most of these studies were focused on population structures, phylogeography, and phylogenetic relationships (*7*−*11*), and not on species identification. More recently, several PCR-RFLP-based methods for the detection of crustacean (*12*) and penaeid shrimps (*13*) DNA have been proposed. Protein-based methodologies have also been used for authentication purposes. Recently, proteomic tools have been applied to the identification of species in seafood products (*14*, *15*), but little effort has been made to elucidate differences among closely related seafood species using mass spectrometry (MS) (*16*−*18*), and only electrophoretic and immunological assays have been reported to date for the detection and differentiation of decapod crustaceans (*19*−*21*). Proteomics is concerned with identifying and determining the structure, expression, localization, interactions, and cellular roles of proteins. Though great advances have been made in this field in the past decade, there are but only a few papers dealing specifically with taxonomic studies and proteomics (*22*). Thus, with the emergence of molecular and bioinformatics analyses of whole genomes, selection at the molecular level is measured almost exclusively by analysis of DNA sequence variation. However, proteins are responsible for the phenotype, and selection acts on the structures that proteins build; therefore, analyses based on protein sequences have capital relevance and complement DNA-based methods (*23*).

Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) is one of the MS systems most extensively applied to proteomic studies. This technology, though limited

*Corresponding author. Tel: 0034-986-231-930. Fax: 0034-986-292-762. E-mail: nachoog@iim.csic.es.

 Published on Web 06/03/2009

Table 1. Penaeid Shrimp Populations Considered in This Study

scientific name[a]	commercial name	origin[b]	code[c]	accession number
Penaeus monodon	giant tiger prawn	IWP (Commercial)	MPN-A	EF589684
Penaeus monodon	giant tiger prawn	IWP (Malaysia)	MPN-B/C	EF589682/EF589685
			MPN-D	EF589683
Penaeus monodon	giant tiger prawn	WI (Mozambique)	MPN-E/F	FJ744153/FJ744154
Litopenaeus vannamei	Pacific white shrimp	EP (Costa Rica)	PNV-A/B/C	EF589702/EF589703/−
Fenneropenaeus indicus	Indian white prawn	WI (Mozambique)	PNI-A/B	EF589688/EF589689
			PNI-C/D	EF589690/EF589687
			PNI-E	EF589686
Farfantepenaeus notialis	Southern pink shrimp	EA (Senegal)	SOP-1/2A	EF589698/EF589697
			SOP-2C	EF589696
Pleoticus muelleri	Argentine red shrimp	SWA (Argentina)	LAA-1/2A	EF589718/EF589717
			LAA-2B/D/E	EF589716/−/−
Pleoticus muelleri	Argentine red shrimp	SWA (Commercial)	LAA-F/G	−/−
Pandalus borealis	Northern shrimp	NA (Greenland)	PRA-A/B/C	EU548069/EU548070/−
Pandalus borealis	Northern shrimp	NA (Commercial)	PRA-D	−

[a] The taxonomic classification proposed by Pérez-Farfante and Kensley (*2*) was adopted. [b] Origin abbreviations: IWP, Indo-West Pacific ocean; WI, Western Indian ocean; EP, Eastern Pacific ocean; WCP, Western Central Pacific ocean; EA, Eastern Atlantic ocean; SWA, Southern West Atlantic ocean; NA, Northern Atlantic ocean. [c] The three initial letters correspond to the FAO codes. Different letters and numbers after the FAO code indicate different individuals.

due to the need for relatively pure samples, is highly advantageous due to its relatively low cost, simplicity, and sensitivity. MALDI-TOF MS methodologies have been used in the protein profile fingerprint-based studies for species differentiation and phylogenetic relationships of microorganisms belonging to different genera and species (*24*−*26*). Moreover, the term phyloproteomics was introduced for the development and application of proteomics methodology to phylogenetic and evolutionary studies (*24*). The other main application of MALDI-TOF MS is peptide mass fingerprinting (PMF), for the identification of proteins by comparing the obtained spectra with a database (*27*). One of the advantages over protein profiling is that PMF, because of previous tryptic digestion, allows the identification of both protein and peptides, directly by database searching of the PMF spectra, and may be followed by peptide MS/MS fragmentation if results are conclusive enough. On the contrary, protein profiling does not identify any of the constituent proteins but simply provides their masses. In addition, peptide mass values can be assigned with higher precision because of the peak isotopic resolution possible for the low range (< 5000 Da) of mass spectra in common MALDI-TOF equipment. PMF has revealed itself to be a useful technique to differentiate among related marine species (*16*−*18*).

The objective of this work was to study the suitability of a proteomic approach, combining 2-DE and PMF, to discriminate among closely related shrimp species of commercial relevance. Four different samples of commercial specimens were obtained in local markets to validate the methodology. On the basis of the cluster analysis of the MALDI-TOF PMF spectra obtained, dendrograms were generated, which were validated with those obtained using DNA-based methods.

2. MATERIALS AND METHODS

2.1. Shrimp Species and Populations Considered. Specimens analyzed are shown in **Table 1**. They were collected using extractive fishing practices or from aquaculture facilities in different continents worldwide. We tested PMF's ability to differentiate among closely related species by considering six species from the order Decapoda, four of them from the Penaeidae family, *Penaeus monodon*, *Litopenaeus vannamei*, *Fenneropenaeus indicus*, and *Farfantepenaeus notialis*, one from the Solenoceridae family, *Pleoticus muelleri*, and one from the Pandalidae family, *Pandalus borealis*. Shrimp, whole animals, were frozen on board and shipped to our laboratory for analyses. Special care was taken in keeping their morphological characteristics in good shape. Whole frozen shrimps from commercial sources purchased in retail markets in Spain were also considered (**Table 1**). At least three individuals of each species were analyzed. Specimens were classified in their respective taxons according to their anatomical external features with the help of a marine biologist from the Marine Sciences Institute (Mediterranean Centre for Marine and Environmental Research, Higher Council for Scientific Research, CMIMA-CSIC, Barcelona, Spain) with expertise in penaeid shrimp taxonomy.

2.2. Extraction of Sarcoplasmic Proteins. Extraction of the sarcoplasmic proteins was performed by homogenizing 1 g of raw white muscle from each individual in two volumes of Milli-Q water, using an Ultra-Turrax blender for 3 × 15 s with interruptions of 45 s to avoid warming the samples. The extracts were then centrifuged at 30 000*g* for 15 min at 4 °C (J25 centrifuge; Beckman, Palo Alto, CA), and the supernatants were maintained at −80 °C until electrophoretic analysis. Protein concentration in the extracts was determined by the bicinchoninic acid (BCA) method (Sigma Chemical Co., USA).

2.3. 2-DE. Three gels were run per individual, and at least three individuals of each species were analyzed. First dimension native isoelectric focusing (IEF) was performed at 10 °C in a Multiphor II electrophoresis unit (Amersham Biosciences, Sweden) as described elsewhere (*17*). Briefly, protein extracts (4 mg/mL, 100−120 μg of protein) were loaded on IEF precast polyacrylamide gels (Ampholine PAGplate pH 4.0−6.5; GE Healthcare, Uppsala, Sweden) in duplicate using sample applicator paper and run under 1500 V, 50 mA, 30 W conditions, until 4000 V·h was reached. To avoid overlapping of proteins, strips with a narrow pH range (pH 4.0−6.5) were chosen. IEF strips corresponding to individual lanes were cut after the run was completed and kept at −80 °C until second dimension electrophoresis analysis was performed.

Equilibration of pH 4−6.5 IEF gel strips was carried out at room temperature as described previously (*28*). Briefly, the strips were placed for 10 min in sample buffer containing 0.75% DTT and then for another 10 min in sample buffer containing 4.5% iodoacetamide.

Preparative second dimension electrophoresis for MS analysis was performed at 15 °C in the Multiphor II electrophoresis unit (Amersham Biosciences, Sweden), using vertical SDS−PAGE (7.5%T and 3%C) gels with Tris-tricine buffer. Running conditions were 100 V, 40 mA per gel, and 150 W. Gels were stained with Coomassie brilliant blue (CBB) (GE Healthcare), which is more compatible with sample preparation for MS analysis, according to the manufacturer's instructions. Three gels were run per individual to ensure reproducibility and were analyzed by means of PDQuest 2-D Analysis Software, version 7.1.0 (Bio-Rad Laboratories, Hercules, CA).

2.4. Mass Fingerprinting of Tryptic Peptides. *2.4.1. In-Gel Digestion*. One major spot from the sarcoplasmic proteome, present in all gels, was selected as a potential molecular marker because its p*I* exhibited interspecific variability. This spot was excised from each gel and subjected to in-gel digestion with trypsin (Roche Diagnostics GmbH, Mannheim, Germany), performed overnight at 37 °C as described elsewhere (*29*). The final digestion solution was dried under vacuum and resuspended in 0.1% TFA.

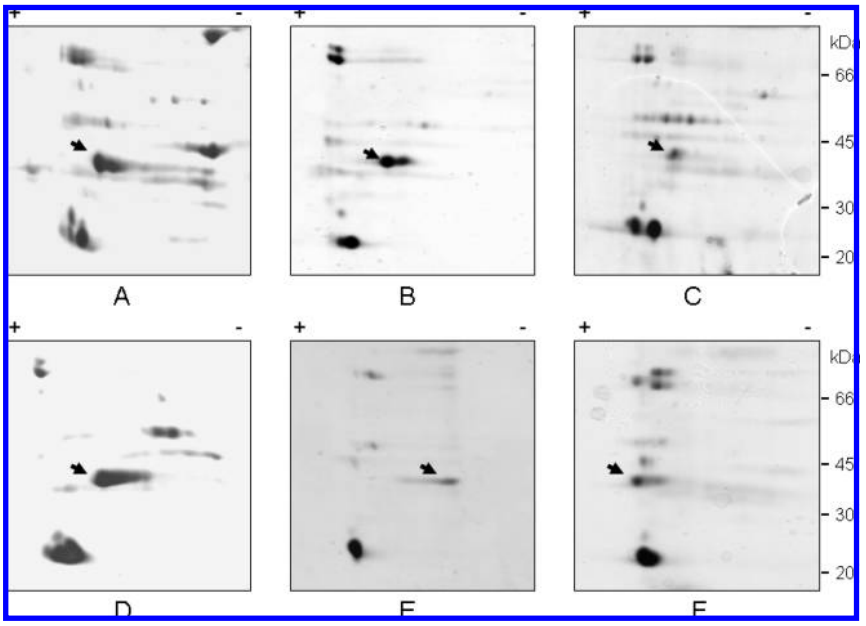

Figure 1. 2-DE patterns representative of the six *Penaeoidea* species tested. Three gels were run per individual. One hundred to 120 μg of protein were resolved first on a p*I* range of 4.0−6.5 and separated on a 7.5% polyacrylamide gel. (**A**) *P. monodon*; (**B**) *L. vannamei*; (**C**) *F. indicus*; (**D**) *F. notialis*; (**E**) *P. muelleri*; (**F**) *P. borealis*. The protein spots selected for MALDI-TOF analysis are highlighted.

2.4.2. PMF. A 1 μL aliquot of the final sample solution was manually deposited onto a stainless steel MALDI probe and allowed to dry at room temperature. Then, matrix solution (0.8 μL of saturated CHCA in 50% aqueous ACN and 0.1% TFA) was added and again allowed to dry at room temperature.

Mass spectra were obtained using a Voyager DE STR MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA) operating in the reflector, delay extraction, and positive-ion mode. Laser intensity was set just above the ion generation threshold. The values for the MS parameters were as follows: low mass gate, m/z 500 Da; delay time, 350 ns; accelerating voltage, 20,000 V; and grid voltage, 68.5%. Mass spectra were acquired by accumulating 150 laser shots in the m/z range from 850 to 3500. External close calibration with Calibration Mixture 2 of the Sequazyme Peptide Mass Standards Kit (Applied Biosystems) was used. Mass spectra were baseline corrected, and data lists containing monoisotopic m/z values were extracted from mass spectral data with the specific software of the instrument (Data Explorer, version 4.0.0.0). Signals within the 920 to 3000 m/z range, with relative intensities greater than 5%, were included in the lists. Mass lists were used for protein and peptide identification using MASCOT Peptide Mass Fingerprint. The experimental data were compared with calculated peptide mass, by applying appropriate cleavage rules to the entries in a sequence database, and the peptides and proteins wich best match the data were probability-based scored (*27*). Search parameters were as follows: NCBInr database; all entries; allowing up to 2 missed cleavages; peptide tolerance, 100 ppm (ppm); and monoisotopic and protonated mass (MH^+) values.

2.4.3. MS/MS Analysis. In order to confirm identification, several peaks obtained by MALDI-TOF were subjected to MS/MS analysis, using an IT mass spectrometer model LCQ Deca XP Plus (Thermo-Finnigan, San Jose, CA), and the sequences inferred were searched against the NCBInr protein sequences database using the basic local alignment search tool (BLAST) (http://www.ncbi.nlm.nih.gov).

2.5. PMF Cluster Analysis. Peptide mass fingerprints of all the specimens were compared by cluster analysis. Each peak list was integrated in a present (1) and absent (0) peak matrix. If the mass difference between peaks m_i in peak list i and m_j in peak list j was less than 200 ppm, then the two mass peaks were assumed to be the same, allowing this error in exact measurement of individual peak masses. A distance matrix was generated by the Windist software (from Winboot package, freely available at www.irri.org), using the Ochiai similarity coefficient (*30*). The distance coefficient $d_{i,j}$ between peak lists i and j is computed by subtracting the similarity coefficient $S_{i,j}$ from 1 (eq 1).

$$d_{i,j} = 1 - S_{i,j} \quad (1)$$

The similarity measure $S_{i,j}$ between peak lists i and j is defined as follows in eq 2:

$$S_{ij} = \frac{a_{ij}}{\sqrt{n_i n_j}} \quad (2)$$

where n_i and n_j are the total number of peaks identified in the individual peak lists, and $a_{i,j}$ is the number of peaks in common between peak lists i and j. The value of $d_{i,j}$ is limited to the domain $0 \le d_{i,j} \le 1$. When the peak lists i and j are indistinguishable, $d_{i,j} = 0$; when the peak lists are completely dissimilar, $d_{i,j} = 1$.

The matrix of peak list−peak list distances was represented as a dendrogram with the MEGA software (*31*) using the neighbor-joining method (*32*), with 1000 bootstrap replicates. This involved resampling the sites by the random replacement of the columns of the data table, to obtain 1000 distance matrices, which were represented in a consensus tree (*33*).

2.6. DNA Analysis. DNA was extracted from 250 mg of muscle of each shrimp specimen by means of a commercial kit (DNeasy Tissue kit, QIAGEN, Darmstadt, Germany). PCR amplification was carried out on a MyCycler thermal cycler (BioRad Laboratories, Hercules, CA, USA), as previously described (*13*). The primers used for the amplification and sequencing of the 16S rRNA-tRNA^{Val} genes were 5′-AATATGGC-TGTTTTTAAGCCTAATTCA-3′ and 5′-CGTTGAGAAGTTCGTTG-TGCA-3′. Sequencing was performed in both directions. Prior to sequencing, the PCR products were purified by means of the ExoSAP-IT kit (GE Healthcare). Direct sequencing was performed with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems). The same primers used for PCR were also employed for the sequencing of both strands of the PCR products. Sequencing reactions were analyzed in an automatic sequencing system (ABI 3730XL DNA Analyzer, Applied Biosystems) provided with the POP-7 system. All nucleotide sequences were aligned using ClustalX 1.8 (*34*) and then manually adjusted. Forward and reverse strands were combined and sequences checked for errors. Phylogenetic analysis was conducted in MEGA 4.0 (*31*). Phylogenetic relationships were estimated using neighbor-joining analysis (*32*). Boot-strapping (1000 replicates) was performed to assess the confidence level at each branch.

3. RESULTS AND DISCUSSION

3.1. Selected Spot Identification. Protein profiles observed after 2-DE analysis revealed some differences among species (**Figure 1**). One major spot, corresponding to a protein with a relative molecular mass (*Mr*) or 40 kDa was present in all gels. For this protein, p*I* variations among species were observed,

indicating the potential presence of aminoacid substitutions in its sequence. Table S-1 in Supporting Information reports the mean p*I* (±RSD) for the spot under study, calculated for each species by means of PDQuest 2-D Analysis Software. For this reason, this spot was selected for MS analysis, digested in situ with trypsin, and the produced peptides analyzed by MALDI-TOF-MS. Complete peak lists obtained from all of the specimens analyzed are shown in Table S-2 in the Supporting Information. MASCOT database search was used for the identification of this spot and the protein assigned as Arginine Kinase (accession nos. 115492980, 1708615, 25453078, 25453073 or 585342) or as the allergen Pen m2 (accession no. 27463265), described before as arginine kinase (*35*). MASCOT scores were between 95 and 303, substantially above the significance level, and the sequence coverage between 31 and 62%. Identified peptide sequences are compiled in Table S-3 in Supporting Information. Although arginine kinase unambiguous matches were returned from database searches, not all of the peaks within the PMF could be identified, with the remaining unidentified peaks arising from post-translationally modified peptides, incorrect cleavage, adducts, and so forth. In addition, crustaceans are poorly characterized at the genome and proteome levels; therefore, the protein under study may be highly homologous to the protein found in the database, but, as there is no complete identity, unmatched peptides will exist (although still allowing the PMF identification of closely related species such as those on which we are focused). Arginine kinase was also confirmed by MS/MS analysis. **Figure 2** shows several of the MS/MS spectra corresponding to arginine kinase peptides found in all of the specimens belonging to the Penaeidae and Solenoceridae families (**Figure 2a** and **b**) (sequences FLQAANAC$^{\#}$R and GTRGEHTEAEGGIYDISNK) or to all specimens belonging to the Pandalidae family (**Figure 2c**) (sequence LVDDHFLFVSGDR).

3.2. MALDI-TOF MS PMF. To assess the ability of the MALDI-TOF MS PMF spectra to differentiate shrimp families and species, different species from the Penaeidae family (*P. monodon*, *L. vannamei*, *F. indicus* and *F. notialis*), the Solenoceridae family (*P. muelleri*), and the Pandalidae family (*P. borealis*) were analyzed. Representative spectra from each of the species analyzed in this study are shown in **Figure 3**. While there were differences among the spectra, some of the peaks detected were common to all specimens belonging to one or more species, suggesting that species-specific peptides were present and that these may be considered for species identification. Thus, a peak at *m*/*z* 1077.6 was present only in *P. monodon*; peaks at *m*/*z* 1531.8, 1761.9, 2271.0, and 2287.2 were only present in *L. vannamei*; and a peak at *m*/*z* 1719.9 was present only in *F. notialis*. Family-specific peptides were detected when comparing the spectra from the specimens of each family. Thus, peaks at *m*/*z* 1008.5, 1657.8, and 2599.2 were specific to the Penaeidae family, while peaks at *m*/*z* 994.5, 1263.7, 1286.7, 1532.8, 1718.8, 1790.0, and 2644.2 were specific to the Solenoceridae family. PMF results of arginine kinase from *P. borealis* were completely different, with several specific peaks at *m*/*z* 1196.5, 1634.7, 1690.8, 1757.8, 1766.8, 1812.8, 1828.8, 2126.0, and 2350.0.

3.3. Cluster Analysis of Argine Kinase PMF. To asses quantitatively if arginine kinase PMF could be used to distinguish among different families and species, a neighbor-joining cluster analysis of all the spectra was performed, and a dendrogram was derived to represent the results graphically. The presence or absence of a peak was considered as a taxonomical characteristic with two states: peak presence (1) or peak absence (0) (Table S-4 in Supporting Information). The number of considered peaks in the spectra ranged from 17 to 48 (average was 30 peaks). The matrix of peak presence/absence was constructed with a total of

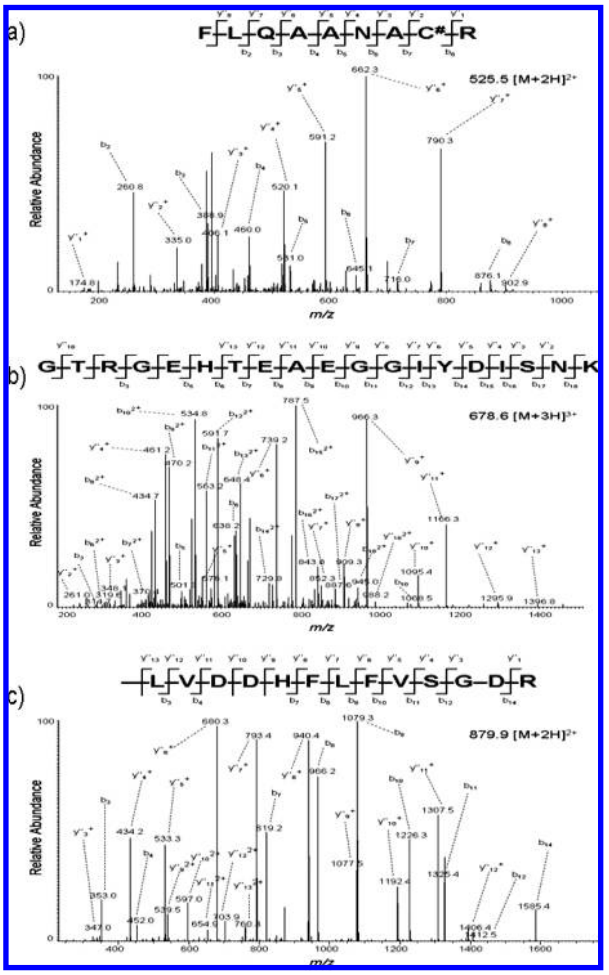

Figure 2. Ion-trap MS/MS spectrum of several arginine kinase peptides obtained after trypsin digestion of the spot selected in 2-DE gels. (**a**) FLQAANAC$^{\#}$R; (**b**) GTRGEHTEAEGGIYDISNK, which are present in the selected spot of all *P. monodon*, *L. vannamei*, *F. indicus*, *F. notialis*, and *P. muelleri* analyzed; and (**c**) LVDDHFLFVSGDR, which is present in the selected spot of all *P. borealis* analyzed. C$^{\#}$: carboxamidomethilated cysteine.

147 possible peaks. Accordingly, **Table 2** shows the genetic distance among specimens. **Figure 4A** illustrates the neighbor-joining tree obtained from arginine kinase PMF cluster analysis. The numerical similarity between two specimens or two clusters is the similarity value of the point at which the two branches diverge.

Given the general lack of information on protein expression in shrimp, the possibility of intraspecific variation due to amino acid substitutions and/or post-translational modifications (PTMs) has been considered. However, the presence or absence of a particular peak due to PTMs is insignificant compared to number of peak mass differences caused by interspecific aminoacidic substitutions, although it might be necessary to test a larger number of specimens from each of the species to detect the presence of intraspecific differences, due to PTMs or sequence variations. Nevertheless, in the analytical conditions used in this work, the interspecific peptide variation is greater than the intraspecific variation, as indicated by the distances from cluster to cluster (**Figure 4** A), and good enough for unambiguous species identification.

The dendrogram revealed some relevant results. First of all, this methodology allowed the classification of all specimens considered in six clusters, each of them corresponding to each

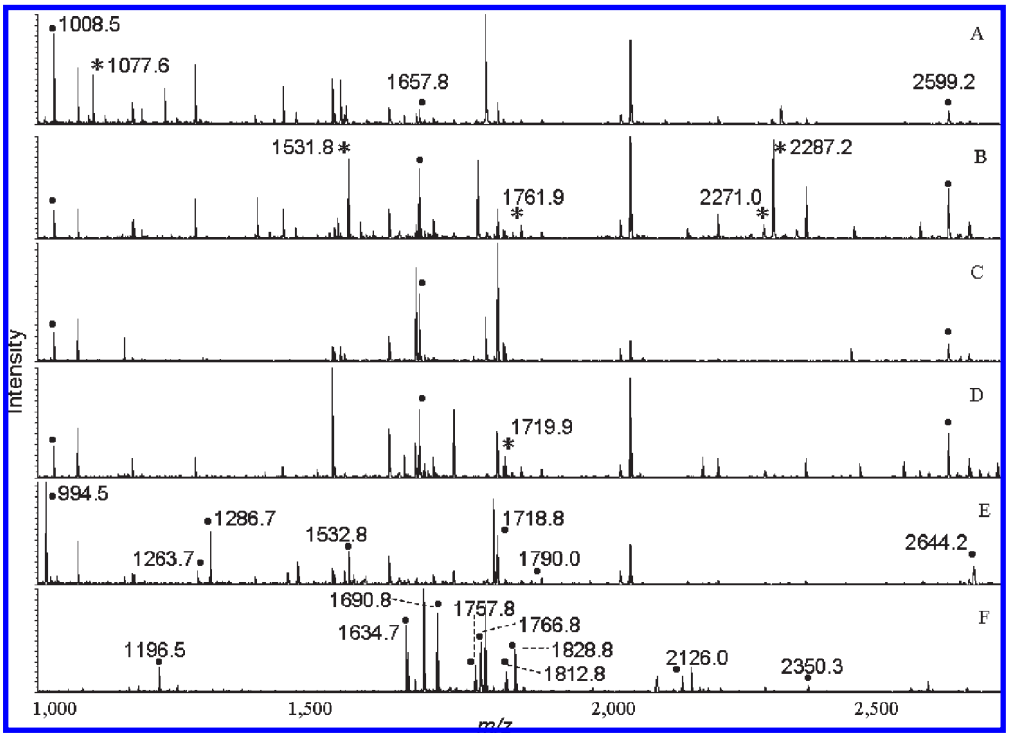

Figure 3. Arginine kinase PMF for representative individuals from 6 different shrimp species. (**A**) *P. monodon*; (**B**) *L. vannamei*; (**C**) *F. indicus*; (**D**) *F. notialis*; (**E**) *P. muelleri*; (**F**) *P. borealis*. Family specific peptides (•) and species-specific peptides (*) are highlighted.

species. The first cluster corresponded to *P. monodon* specimens, with similarity values of ~0.85 but a bootstrap value of 52 (**Figure 4** A). All three *F. notialis* specimens were tightly clustered, exhibiting similarity values of ~0.9 and a bootstrap support of 82. Another cluster was formed by *L. vannamei* specimens, with a similarity value over ~0.85 and bootstrap values over 90. All five *F. indicus* specimens were clustered with similarity values over 0.8, although the bootstrap value was below 50. The fifth cluster was formed by *P. muelleri* specimens, with similarity values over 0.8 and the highest bootstrap value of 100. Finally, the sixth cluster included *P. borealis* specimens with high similarity and a bootstrap value of 100. The first five clusters could be strongly grouped (indicated by the highest bootstrap support, 100) into two bigger clusters with similarity values of ~0.7, one corresponding to the Penaeidae family (bootstrap value of 91) and the other corresponding to the Solenoceridae family (bootstrap value of 100) (**Figure 4A**). These clustering results showed the reproducibility of PMF spectra for measurements of different individuals of a particular species. Commercial samples were also effectively identified after measuring their level of similarity against the rest of the specimens.

Topology resulting from the PMF cluster analysis (**Figure 4A**) was highly concordant with a dendrogram derived from an analysis of 16S rRNA-tRNA^{Val} mithocondrial genes (**Figure 4B**) and with previously described taxonomical levels in penaeid shrimp determined by independent validated methods (*8*−*11*, *13*). Remarkably, a detailed comparison with such previous phylogenetic trees, constructed on the basis of mtDNA data, revealed only slight differences. This result highlights the accuracy of the information provided by the arginine kinase PMF method described in this work, as compared to the genetic studies based on either 16S rRNA or cytochrome oxidase I mitochondrial genes.

The efficiency of the arginine kinase PMF methodology to differentiate between the three families and the six shrimp species and its usefulness in the identification of four commercial samples has been demonstrated. At first sight, the proteomic approach described in this work seems to be as time-consuming and complicated as the usual genetic analyses, such as RFLP, which usually takes two days. However, once we know the p*I* and the M_w of the protein of interest and thus its position in the 2-DE gels, we can obtain the PMF spectra, avoiding the more complicated steps of this proteomic approach such as identification and sequencing of the spot. Proteomic techniques may be automated, producing fast, reproducible, and sensitive results and allowing high-throughput analysis of foodstuffs. In addition, this methodology can be applied to species that are poorly characterized at the genomic and proteomic levels in the databases, avoiding the disadvantages of DNA sequencing. Sarcoplasmic proteins are usually more heat-stable than other myofibrillar proteins; thus, the selection of this fraction of proteins seems to be an asset to the applicability of this proteomic approach on products treated at high temperatures. However, heated samples were not included in this study. For this purpose, further studies have to be made.

A great advantage of the proteomic approach is that it is the first step toward designing easy and cheap detection analyses: the identification of specific peptides, as done by this work, can be used to identify not only taxons but also proteins of interest. These diagnostic peptides can be used in fast, cheap, and easy-to-use kits based on antibodies against the specific peptides. The selection of arginine kinase is particularly interesting because in addition to discriminating species, it is also an allergen (*35*), and therefore, an analysis targeting this protein would have a double application: in species identification, if the peptide selected discriminates at the species level, and in food safety, if the peptide selected discriminates for common peptides for several relevant species. This may permit the detection of the allergen in certain products and help the

Table 2. Distance Matrix between the Arginine Kinase PMF Peak Lists of All Different Specimens[a], Calculated by Means of the Ochiai Similarity Coefficient

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1														
2	0.277													
3	0.158	0.221												
4	0.444	0.176	0.332											
5	0.958	0.956	0.958	0.956										
6	0.957	0.955	0.957	0.955	0.186									
7	0.955	0.952	0.955	0.952	0.212	0.235								
8	0.960	0.958	0.960	0.958	0.269	0.224	0.249							
9	0.963	0.961	0.963	0.961	0.289	0.217	0.268	0.153						
10	0.961	0.958	0.961	0.958	0.280	0.268	0.294	0.373	0.332					
11	1.000	1.000	1.000	1.000	0.487	0.449	0.449	0.456	0.367	0.545				
12	1.000	1.000	1.000	1.000	0.484	0.475	0.376	0.477	0.455	0.545	0.310			
13	1.000	1.000	1.000	1.000	0.499	0.437	0.434	0.472	0.368	0.480	0.188	0.337		
14	0.956	0.953	0.956	0.953	0.543	0.500	0.509	0.363	0.407	0.505	0.579	0.626	0.583	
15	0.947	0.944	0.947	0.944	0.497	0.446	0.370	0.401	0.405	0.489	0.606	0.473	0.570	0.294
16	0.947	0.944	0.947	0.944	0.372	0.318	0.235	0.281	0.330	0.449	0.498	0.432	0.470	0.382
17	0.953	0.950	0.953	0.950	0.441	0.356	0.359	0.360	0.338	0.510	0.490	0.350	0.440	0.450
18	0.955	0.952	0.955	0.952	0.427	0.308	0.308	0.317	0.268	0.495	0.449	0.445	0.406	0.434
19	1.000	1.000	1.000	1.000	0.452	0.350	0.412	0.362	0.243	0.428	0.401	0.470	0.399	0.487
20	1.000	1.000	1.000	1.000	0.430	0.359	0.387	0.342	0.200	0.464	0.332	0.419	0.312	0.494
21	1.000	1.000	1.000	1.000	0.408	0.337	0.364	0.322	0.263	0.444	0.417	0.426	0.415	0.438
22	1.000	1.000	1.000	1.000	0.704	0.639	0.682	0.548	0.500	0.666	0.569	0.685	0.579	0.500
23	1.000	1.000	1.000	1.000	0.567	0.559	0.606	0.587	0.585	0.624	0.658	0.677	0.631	0.649
24	1.000	1.000	1.000	1.000	0.673	0.585	0.620	0.533	0.492	0.591	0.581	0.657	0.570	0.512
25	1.000	1.000	1.000	1.000	0.629	0.536	0.571	0.511	0.468	0.625	0.537	0.641	0.549	0.579
26	1.000	1.000	1.000	1.000	0.627	0.586	0.636	0.580	0.608	0.713	0.652	0.639	0.625	0.714
27	1.000	1.000	1.000	1.000	0.755	0.626	0.693	0.689	0.637	0.732	0.686	0.684	0.677	0.613
28	1.000	1.000	1.000	1.000	0.707	0.574	0.640	0.641	0.628	0.725	0.678	0.635	0.669	0.691

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
15														
16	0.211													
17	0.298	0.204												
18	0.325	0.190	0.119											
19	0.465	0.426	0.422	0.346										
20	0.472	0.397	0.362	0.291	0.151									
21	0.405	0.367	0.371	0.332	0.243	0.227								
22	0.591	0.591	0.603	0.586	0.513	0.467	0.526							
23	0.623	0.581	0.590	0.535	0.574	0.580	0.556	0.437						
24	0.555	0.624	0.574	0.591	0.503	0.510	0.444	0.250	0.347					
25	0.606	0.534	0.554	0.479	0.479	0.435	0.443	0.291	0.259	0.185				
26	0.702	0.574	0.545	0.490	0.629	0.542	0.608	0.488	0.186	0.446	0.304			
27	0.590	0.641	0.544	0.518	0.627	0.596	0.637	0.456	0.428	0.400	0.371	0.377		
28	0.632	0.632	0.579	0.550	0.618	0.585	0.628	0.479	0.372	0.384	0.319	0.318	0.179	

[a] Codes: 1, PRA A; 2, PRA C; 3, PRA B; 4, PRA D; 5, MPN B; 6, MPN D; 7, MPN C; 8, MPN E; 9, MPN F; 10, MPN A; 11, PNV A; 12, PNV B; 13, PNV C; 14, PNI E; 15, PNI D; 16, PNI C; 17, PNI B; 18, PNI A; 19, SOP 1, 20, SOP 2A; 21, SOP 2C; 22, LAA 2B; 23, LAA 2A; 24, LAA 1; 25, LAA D; 26, LAA E; 27, LAA F; 28, LAA G.

manufacturer to identify, in real-time, the existence of a contamination in a production line.

The use of a mathematical approach for the comparison of PMF spectra, instead of performing the visual spectral evaluation used in other species differentiation studies (*16*, *17*), eliminates subjectivity bias providing a quantitative measure of how well two spectra match. The phylogenetic relationships derived from the PMF cluster analysis also provided information for the grouping of species belonging to the order Decapoda. To our knowledge, the results reported here represent the first application of a MALDI-TOF MS PMF method to infer taxonomic relationships in any type of organism and are consistent with the phylogenies established with mtDNA data. This study opens the way to more in depth studies with a wider range of shrimp species and provides the basis for further investigations on phyloproteomics. Since proteins constitute the biological machinery, and selection acts on the structures formed by them, analyses based on protein sequences should be considered as a complementary tool to mtDNA-based methods.

ABBREVIATIONS USED

ACN, acetonitrile; BCA, bicinchoninic acid; BLAST, basic local alignment search tool; CBB, Coomassie brilliant blue; CHCA, α-cyano-4-hydroxycinnamic acid; Da, Dalton (molecular mass); 2-DE, two-dimensional gel electrophoresis; DTT, dithiothreitol; IEF, isoelectric focusing; MALDI-TOF, matrix-assisted laser desorption/ionization-time-of-flight; *Mr*, relative molecular mass; MS, mass spectrometry; MS/MS, tandem mass spectrometry; *m*/*z*, mass-to-charge ratio; PCR, polymerase chain reaction; p*I*, isoelectric point; PMF, peptide mass fingerprinting; ppm, parts per million; PTM, post-translational modifications; RFLP, restriction fragment length polymorphism; RSD, relative standard deviation; SDS−PAGE, sodium dodecyl

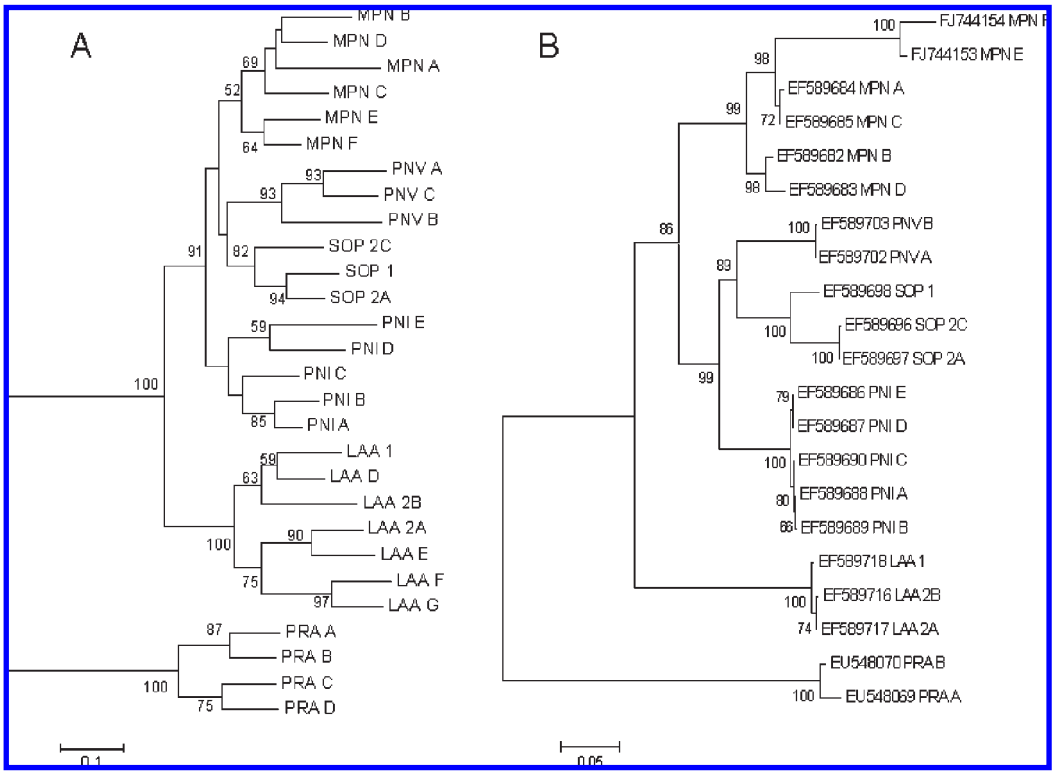

Figure 4. Topologies resulting from (**A**) arginine kinase PMF cluster analysis; (**B**) phylogenetic analysis of the nucleotide sequences of 16S rRNA-tRNA^{Val} mithocondrial genes. Both were obtained by means of the neighbor-joining method; bootstrap values >50 are shown. MPN, *P. monodon*; PNV, *L. vannamei*; SOP, *F. notialis*; PNI, *F. indicus*; LAA, *P. muelleri*; PRA, *P. borealis*; as referred to in **Table 1**.

sulfate−polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; Tris, tris(hydroxymethyl)aminomethane.

ACKNOWLEDGMENT

We thank Lorena Barros for her excellent technical assistance and Dr. Vanesa Díaz for her helpful assistance. We also acknowledge members from CETMAR for their helpful collaboration in the collection of specimens for this study. The manuscript benefited from critical comments by three anonymous reviewers.

Supporting Information Available: p*I* of the analyzed spot, calculated from the 2-DE patterns of all the individuals of each species studied; complete peaks lists obtained from arginine kinase PMF of all specimens analyzed in this study; complete information of the MALDI-TOF analysis showing Decapoda arginine kinase matching peptide sequences; and matrix of peaks presence (1 = present; 0 = absent) of all different specimens. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Holthuis, L. B. *FAO Species Catalogue: Shrimps and Prawns of the World. An Annotated Catalogue of Species of Interest to Fisheries*; FAO Fish. Synop. (125); FAO: Rome, Italy, 1980; Vol. 1, pp 1−271.
(2) Pérez-Farfante, I., Kensley, B. *Penaeoid and Sergestoid Shrimps and Prawns of the World. Keys and Diagnoses for the Families and Genera*; Éditions du Muséum National d'Histoire Naturelle: Paris, France, 1997.
(3) Council Regulation (EC) No. 104/2000 of the European Parliament of Dec. 17, 1999 on the common organization of the markets in fishery and aquaculture products.
(4) Royal Decree 121/2004 of Jan. Twenty-three (BOE No. 31, Feb. 5, 2004) from Ministry of Agriculture, Fisheries and Food, Spain.
(5) Civera, T. Species identification and safety of fish products. *Vet. Res. Commun.* **2003**, *27*, 481–489.
(6) Rasmussen, R. S.; Morrissey, M. T. DNA-based methods for the identification of commercial fish and seafood species. *Compr. Rev. Food Sci. Food Saf.* **2008**, *7*, 280–295.
(7) Maggioni, R.; Rogers, A. D.; Maclean, N.; Incao, F. Molecular phylogeny of Western Atlantic Farfantepenaeus and Litopenaeus shrimp based on mitochondrial 16S partial sequences. *Mol. Phylogenet. Evol.* **2001**, *18*, 66–73.
(8) Baldwin, J. D.; Bass, A. L.; Bowen, B. W.; Clark, W. H. Molecular phylogeny and biogeography of the marine shrimp *Penaeus*. *Mol. Phylogenet. Evol.* **1998**, *10*, 399–407.
(9) Quan, J.; Zhuang, Z.; Deng, J.; Dai, J.; Zhang, Y.-p. Phylogenetic relationships of 12 Penaeoidea shrimp species deduced from mitochondrial DNA sequences. *Biochem. Genet.* **2004**, *42*, 331–345.
(10) Lavery, S.; Chan, T. Y.; Tam, Y. K.; Chu, K. H. Phylogenetic relationships and evolutionary history of the shrimp genus *Penaeus s. l.* derived from mitochondrial DNA. *Mol. Phylogenet. Evol.* **2004**, *31*, 39–49.
(11) Voloch, C. M.; Freire, P. R.; Russo, C. A. M. Molecular phylogeny of penaeid shrimps inferred from two mitochondrial markers. *Genet. Mol. Res.* **2005**, *4*, 668–674.
(12) Brzezinski, J. L. Detection of crustacean DNA and species identification using a PCR-restriction fragment length polymorphism method. *J. Food Prot.* **2005**, *68*, 1866–1873.
(13) Pascoal, A.; Barros-Velázquez, J.; Cepeda, A.; Gallardo, J. M.; Calo-Mata, P. A polymerase chain reaction-restriction fragment length polymorphism method based on the analysis of a 16S rRNA/tRNA^{Val} mitochondrial region for species identification of commercial penaeid shrimps (*Crustacea: Decapoda: Penaeoidea*) of food interest. *Electrophoresis* **2008**, *29*, 499–509.
(14) Piñeiro, C.; Barros-Velázquez, J.; Vázquez, J.; Figueras, A.; Gallardo, J. M. Proteomics as a tool for the investigation of seafood and other marine products. *J. Proteome Res.* **2003**, *2*, 127–135.
(15) Martínez, I.; Friis, T. J. Application of proteome analysis to seafood authentication. *Proteomics* **2004**, *4*, 347–354.
(16) López, J. L.; Marina, A.; Álvarez, G.; Vázquez, J. Application of proteomics for fast identification of species-specific peptides from marine species. *Proteomics* **2002**, *2*, 1658–1665.

(17) Carrera, M.; Cañas, B.; Piñeiro, C.; Vázquez, J.; Gallardo, J. M. Identification of commercial hake and grenadier species by proteomic analysis of the parvalbumin fraction. *Proteomics* **2006**, *6*, 5278–5287.
(18) Carrera, M.; Cañas, B.; Piñeiro, C.; Vázquez, J.; Gallardo, J. M. De novo mass spectrometry sequencing and characterization of species-specific peptides from nucleoside diphosphate kinase B for the classification of commercial fish species belonging to the family Merlucciidae. *J. Proteome Res.* **2007**, *6*, 3070–3080.
(19) An, H.; Wei, C. I.; Zhao, J.; Marshall, M. R.; Lee, C. M. Electrophoretic identification of fish species used in surimi products. *J. Food Sci.* **1989**, *54*, 253–257.
(20) Civera, T.; Parisi, E. The use of sodium dodecyl sulphate polyacrylamide gel electrophoresis for the identification of raw and cooked shrimps and crabs. *Ital. J. Food Sci.* **1991**, *3*, 149–157.
(21) Shanti, K. N.; Martin, B. M.; Nagpal, S.; Metcalfe, D. D.; Subba Rao, P. V. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J. Immunol.* **1993**, *151*, 5354–5363.
(22) Navas, A.; Albar, J. P. Application of proteomics in phylogenetic and evolutionary studies. *Proteomics* **2004**, *4*, 299–302.
(23) Karr, T. L. Application of proteomics to ecology and population biology. *Heredity* **2008**, *100*, 200–206.
(24) Conway, G. C.; Smole, S. C.; Sarracino, D. A.; Arbeit, R. D.; Leopold, P. E. Phyloproteomics: species identification of *Enterobacteriaceae* using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J. Mol. Microbiol. Biotechnol.* **2001**, *3*, 103–112.
(25) Dickinson, D. N.; La Duc, M. T.; Haskins, W. E.; Gornushkin, I.; Winefordner, J. D.; Powell, D. H.; Venkateswaran, K. Species differentiation of a diverse suite of *Bacillus* spores by mass spectrometry-based protein profiling. *Appl. Environ. Microbiol.* **2004**, *70*, 475–482.
(26) Mazzeo, M. F.; Sorrentino, A.; Gaita, M.; Cacace, G.; Di Stasio, M.; Facchiano, A.; Comi, G.; Malorni, A.; Siciliano, R. A. Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the discrimination of food-borne microorganisms. *Appl. Environ. Microbiol.* **2006**, *72*, 1180–1189.
(27) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567.
(28) Piñeiro, C.; Barros-Velazquez, J.; Sotelo, C. G.; Perez-Martin, R. I.; Gallardo, J. M. Two-dimensional electrophoretic study of the water-soluble protein fraction in white muscle of Gadoid fish species. *J. Agric. Food Chem.* **1998**, *46*, 3991–3997.
(29) Jensen, O. N.; Wilm, M.; Shevchenko, A.; Mann, M. Sample Preparation Methods for Mass Spectrometric Peptide Mapping Directly from 2-DE Gels. In *2-D Proteome Analysis Protocols*; Link, A. J., Ed.; Humana Press: Totowa, NJ, 1999; pp 513−530.
(30) Ochiai, A. Zoogeographical studies on the Soleoid fishes found in Japan and its neighbouring regions. *Bull. Jpn. Soc. Sci. Fish.* **1957**, *22*, 526–530.
(31) Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596–1599.
(32) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.
(33) Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. *CBMS-NSF Regional Conference Series in Applied Mathematics*, Monograph 38, SIAM: Philadelphia, PA, 1982.
(34) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
(35) Yu, C.-J.; Lin, Y.-F.; Chiang, B.-L.; Chow, L.-P. Proteomics and immunological analysis of a novel shrimp allergen, Pen m 2. *J. Immunol.* **2003**, *170*, 445–453.

Received February 13, 2009. Revised manuscript received May 12, 2009. Accepted May 15, 2009. This work was supported by the National Food Program of the INIA (Spanish Ministry for Education and Science) (Project CAL-03-030-C2-2) and by the PGIDIT Research Program in Marine Resources (Project PGIDIT04RMA261004PR) of the Xunta de Galicia (Galician Council for Industry, Commerce and Innovation). I.O. is supported by the FPU program (AP-2004-5826) and B.C. is supported by the RyC program, under the auspices of the Spanish MICINN.