

Mass Spectrometry Characterization of Species-Specific Peptides from Arginine Kinase for the Identification of Commercially Relevant Shrimp Species

Ignacio Ortea,^{*,†} Benito Cañas,[‡] and José M. Gallardo[†]

Marine Research Institute, Spanish National Research Council (CSIC), Eduardo Cabello 6, E-36208 Vigo, Spain, and Analytical Chemistry Department, Universidad Complutense de Madrid, E-28040 Madrid, Spain

Received July 23, 2009

Abstract: The identification of commercial shrimp species is a relevant issue to ensure correct labeling, maintain consumer confidence and enhance the knowledge of the captured species, benefiting both, fisheries and manufacturers. A proteomic approach, based on 2DE, tryptic in-gel digestion, MALDI-TOF MS, and ESI-MS/MS analyses, is proposed for the identification of shrimp species with commercial interest. MALDI-TOF peptide mass fingerprint from arginine kinase tryptic digests were used for the identification of seven commercial, closely related species of Decapoda shrimps. Further identification and characterization of these peptides was performed by CID on an ESI-IT instrument, database search and *de novo* sequence interpretation, paying special attention to differential, species-specific peptides. Fisheries and manufacturers may take advantage of this methodology as a tool for a rapid and effective seafood product identification and authentication, providing and guaranteeing the quality and safety of the foodstuffs to consumers.

Keywords: Arginine kinase • MS/MS • proteomics • shrimps • species identification

Introduction

Seafood products include an extensive variety of species widely used for human nutrition, having a significant impact in food industry. Among them, crustaceans belonging to the order of Decapoda are of remarkable commercial interest. This order includes the Superfamily Penaeoidea, which is one of the most important economic resources in fishery and aquafarming industry.^{1,2} The identification of marine species is of primary relevance for the seafood industry, due to the global commercial demands concerning labeling and traceability^{3,4} for the prevention of commercial fraud, and to avoid the safety risks derived from the inadvertent introduction of food ingredients which might be harmful for human health.⁵ Anatomical features are particularly difficult to be used in penaeid shrimp species differentiation due to their phenotypic similarities and

to removal of the external carapace, which is very frequent during the manufacturing process. Therefore, it is highly recommendable to develop the analytical tools necessary to make possible the distinction between these closely related species, preventing the inadvertent or deliberate mislabeling and adulteration of these products.

Methods for species identification are currently based on DNA or protein analysis. Among the DNA targets, the mitochondrial genes 16S rRNA and cytochrome oxidase I have been used as interspecific markers for several crustacean species, although most of the studies focused on phylogeography and phylogenetic relationships,^{6–10} but not on species identification. More recently, two PCR-RFLP-based methods for the detection of crustacean¹¹ and penaeid shrimps¹² DNA have been proposed. Protein-based methodologies have also been used with authentication purposes. Proteomics tools have been applied to the identification of species in seafood products,^{13,14} but little effort has been made to elucidate differences among closely related species using mass spectrometry (MS).^{15–17} Some electrophoretic and immunological assays have been reported for the detection and differentiation of decapod crustaceans,^{18–20} but only one MS-based method has been used for the authentication of shrimp species.²¹

Arginine kinase (EC 2.7.3.3) is a monomeric phosphagen-ATP phosphotransferase, widely distributed in invertebrates²² and functionally analogous to the more thoroughly studied creatine kinase from vertebrates.²³ Arginine kinase (AK) catalyzes the reversible phosphorylation of arginine using ATP:



A comparison of AK amino acid content in two types of shrimp, *Penaeus aztecus* and *Penaeus japonicus*, revealed that, although nearly identical, some slight differences appeared, even in these closely related species.²⁴ Interspecific variability was also reported for the AK derived from four Decapoda species.²⁵ In a previous work, AK proteins proved to be a potential molecular marker for shrimp species identification, showing marked differences in the MALDI-TOF MS spectra from tryptic digests.²¹ Now, using a combination of preparative 2-DE, MALDI-TOF MS and sequencing by ESI-IT MS/MS, we have characterized the AK proteins from seven closely related species of commercial relevance, belonging to the order Decapoda. Efforts were made in peptide *de novo* sequencing, as genomes from

* To whom correspondence should be addressed. Ignacio Ortea, Marine Research Institute, Spanish National Research Council (CSIC), Eduardo Cabello 6, E-36208 Vigo, Spain; phone, 0034-986-231-930; fax, 0034-986-292-762; e-mail, nachoog@iim.csic.es.

[†] Spanish National Research Council (CSIC).

[‡] Universidad Complutense de Madrid.

members belonging to the order Decapoda still await sequencing. Particular attention was focused on the identification and characterization of several species-specific peptides for each of the species analyzed, which may be used as a tool for seafood product authentication.

Materials and Methods

Shrimp Species and Populations Considered. Specimens analyzed are shown in Table 1. They were collected using either extractive fishing practices or from aquaculture facilities in different continents worldwide. Seven different shrimp species from the order Decapoda were considered in this study; five of them were from the Penaeidae family, *Penaeus monodon*, *Litopenaeus vannamei*, *Fenneropenaeus indicus*, *Fenneropenaeus merguensis*, *Farfantepenaeus notialis*, one from the Solenoceridae family, *Pleoticus muelleri*, and one from the Pandalidae family, *Pandalus borealis*. Penaeidae and Solenoceridae families belong to the superfamily Penaeoidea, and Pandalidae family belongs to the superfamily Pandalioidea. Shrimps, whole animals, were frozen on board and shipped to our laboratory for the analyses. Special care was taken in keeping their morphological characteristics in good shape. At least six 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.

Extraction of Sarcoplasmic Proteins. Approximately 1 g of raw white muscle from each of the individuals studied was homogenized in 2 vol of Milli-Q water using an Ultra-Turrax blender for 3 × 15 s with interruptions of 45 s to avoid warming the samples. Sarcoplasmic protein extracts were then centrifuged at 30 000g for 15 min at 4 °C (J25 centrifuge; Beckman, Palo Alto, CA), and the supernatants were frozen at −80 °C until the electrophoretic analysis. Protein concentration in the extracts was determined by the bicinchoninic acid method (Sigma Chemical Co.).

2-DE. Isoelectric focusing (IEF) was performed at 10 °C in a Multiphor II electrophoresis unit (Amersham Biosciences, Sweden) as described elsewhere.¹⁶ Briefly, protein extracts (4 mg/mL, 100–120 µg of protein) were loaded in duplicate on IEF precast polyacrylamide gels with a narrow range of pH (Ampholine PAGplate pH 4.0–6.5; GE Healthcare, Uppsala, Sweden), using a sample applicator paper, and run at 1500 V, 50 mA, 30 W until 4000 Vh were 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 4.0–6.5 IEF gel strips was carried out at room temperature as described previously.²⁶ Briefly, the strips were placed for 10 min in sample buffer containing 0.75% DTT followed by 10 min incubation in sample buffer containing 4.5% iodoacetamide.

Preparative second-dimension SDS-PAGE for further MS analysis were performed at 15 °C in the Multiphor II electrophoresis unit (Amersham Biosciences, Sweden), using vertical gels (7.5%T and 3%C) 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 very compatible with sample preparation for MS analysis, according to the manufacturer's instructions. PDQuest 2-D Analysis Software Version 7.1.0 (Bio-Rad Laboratories, Hercules, CA) was used for the image analysis.

Peptide Sample Preparation for MS and MS/MS Analysis. AK spots were localized based on the data (*pI* and *M_r*) obtained by previous work,²¹ and excised from the gel taking care in maximizing the protein-to-gel ratio. Only the most intensely stained region at the center of the spot was excised to avoid extracting an excess of gel matrix. Excised pieces were subjected to *in-gel* digestion with trypsin (Roche Diagnostics GmbH, Mannheim, Germany) as described elsewhere.²⁷ Prior to nESI-IT MS analysis, the tryptic peptide samples were desalted and concentrated using in-tip reverse-phase resins (ZipTip C18, Millipore, Bedford, MA), according to the manufacturer's recommendations.

Peptide Mapping. Peptide digests were acidified with trifluoroacetic acid (TFA) and 0.8 µL aliquots were manually deposited onto the stainless steel MALDI probe and allowed to dry at room temperature. Matrix solution (0.8 µL of saturated α-cyano-4-hydroxycinnamic acid (CHCA) in 50% aqueous acetonitrile (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 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 the 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 920–3500 *m/z* range, with relative intensities greater than 5% were included in the lists. Database searches were performed against the NCBI nonredundant protein sequence database (<http://www.ncbi.nih.gov>) using the MASCOT Peptide Mass Fingerprint search program (<http://matrixscience.com>). Search parameters were monoisotopic and protonated masses, two missed cleavages allowed, 100 ppm mass tolerance, and three potential variable modifications: ACET sites (acetyl N-terminal), CysCAM (carbamidomethylcysteine) and MSO (methionine sulfoxide).

Peptide Sequencing by ESI-IT MS/MS. Peptide digests were analyzed online by LC-ESI-IT-MS/MS using a LC system model

Table 1. Penaeid Shrimp Species Considered in the Study

scientific name ^a	commercial name	origin ^b
<i>Penaeus monodon</i>	Giant tiger prawn	IWP and WI
<i>Litopenaeus vannamei</i>	Pacific white shrimp	EP
<i>Fenneropenaeus indicus</i>	Indian white prawn	WI
<i>Fenneropenaeus merguensis</i>	Banana prawn	WCP
<i>Farfantepenaeus notialis</i>	Southern pink shrimp	EA
<i>Pleoticus muelleri</i>	Argentine red shrimp	SWA
<i>Pandalus borealis</i>	Northern shrimp	NA

^aThe taxonomic classification proposed by Pérez-Farfante and Kensley²⁵ was adopted. ^bOrigin 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.

SpectraSystem P4000 (Thermo-Finnigan, San Jose, CA) coupled with an IT mass spectrometer model LCQ Deca XP Plus (Thermo-Finnigan). The separation was performed on a 0.18 mm $\times$ 150 mm BioBasic-18 RP column (ThermoHypersil-Keystone) using 0.5% acetic acid in water and 0.5% acetic acid in ACN as mobile phases A and B, respectively. A 90 min linear gradient from 5 to 60% B, at a flow rate of 1.4–1.7 μ L/min, was used. Peptides were detected using survey scans from 300 to 1300 amu (3 μ scans), followed by a data-dependent Zoom-Scan (5 μ scans) and MS/MS scan (5 μ scans), using an isolation width of 3 amu and a normalized collision energy of 35%. Fragmented masses were set in dynamic exclusion for 3 min after the fragmentation event and singly charged ions were excluded from MS/MS analysis.

Off-line analyses by nESI-IT were performed using an ion trap mass spectrometer, model LCQ Deca XP Plus (Thermo-Finnigan) equipped with a nanospray interface. Peptides were previously desalted by microcolumns ZipTip C18 eluted with 10 μ L of 70% MeOH/0.5%CH₃COOH. PicoTips borosilicate glass needles with 1 μ m orifice (New Objective, Woburn, MA) were filled with 3–5 μ L of sample and used as emitters.

MS/MS spectra were searched using SEQUEST (Bioworks 3.1 package, thermo-Finnigan) against the nr.fasta database (NCBI Resources, NIH, Bethesda, MD). The following constraints were used for the searches: tryptic cleavage after Arg and Lys, up to two missed cleavage sites, and tolerances $\pm$ 1.5 Da for precursor ions and $\pm$ 0.8 Da for MS/MS fragment ions. The variable modifications allowed were methionine oxidation, carbamidomethylation of Cys, and acetylation of the N-terminus.

De novo sequencing was performed by manual interpretation of the ion series in the spectra. BLAST program was used for homology searches between *de novo* obtained sequences and those in the nr-NCBI database.

Results and Discussion

2-DE. Sarcoplasmic protein extracts from 6 individuals for each of the 7 species under study were subjected to 2-DE. Gels were Coomassie stained and analyzed with the image software. To ensure reproducibility of the experiments, 3 gels were run per individual. In accordance with previously reported data,²¹ arginine kinase isoforms run at 40 kDa with an isoelectric point (pI) between 4.68 and 5.25. The 2-DE patterns for the different AKs are shown in Figure S-1 in the Supporting Information and pI values for all the AK spots studied are compiled in Table S-1 in the Supporting Information. Differences in the AK spots pI values pointed to the presence of amino acid substitutions in their sequences; consequently, characterization of each of the AK isoforms was performed by mass spectrometry.

MALDI-TOF MS Analysis of AK Spots. After preparative 2-DE, AK spots were *in-gel* digested with trypsin, and the peptides analyzed by MALDI-TOF MS. The complete list of the peaks detected can be found in Table S-2 in the Supporting Information. Mass coincidences between experimental data and *in silico* produced tryptic digests from the AKs present in protein databases were apparent: 24 peak masses were coincident with theoretical peptides from the *P. monodon* allergen Pen m 2 (gil27463265), seven with peptides from an AK found in *L. vannamei* (gil115492980), another seven with peptides from *Marsupenaeus japonicus* AK (gil1708615); four more peaks matched peptides from *Galapaganus conwayensis* (gil209402936); two peaks were present in the AK from *Solenopsis invicta* (gil192337578); other two were present in the AK from *Chasmagnathus granulata* (gil7243761) and one was contained in the AK sequences from *Periplaneta americana* (gil50428904), *Cimberis pilosa* (gil228014712), *Apis mellifera* (gil58585146), *Homarus gammarus* (gil585342), *Lepidophthalmus louisianensis* (gil171473157), *Artemia franciscana* (gil25453072), *Pterapion* sp.

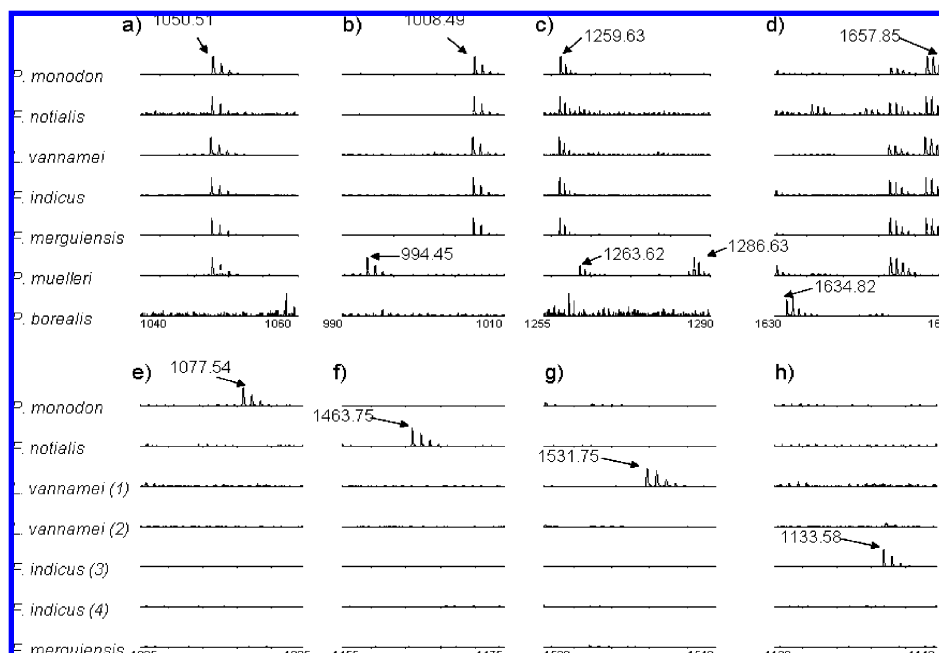


Figure 1. Regions of the MALDI-TOF MS peptide maps from AK trypsin digested spots, showing peaks specific to (a) superfamily Penaeoidea (1050.51 Da); (b, c, and d) family Penaeidae (1008.49, 1259.63, and 1657.85 Da), family Solenoceridae (994.45, 1263.62, and 1286.63 Da) or family Pandalidae (1634.82 Da); (e) *P. monodon* (1077.54 Da); (f) *F. notialis* (1463.75 Da); (g) *L. vannamei* from Costa Rica (1531.75 Da); (h) *F. indicus* from Mozambique (1133.58 Da). Populations from (1) Costa Rica; (2) Argentina; (3) Mozambique and (4) Madagascar.

Table 2. Concise Summary of Peptide Sequence Data Obtained by MS/MS for AKs from the Shrimp Species Studied Useful for Species Identification^a

<i>m/z</i> (2) observed	(<i>M</i> + <i>H</i>) ⁺ matched	Peptide Sequence	Modification	Position	MALDI observed	SOP	PRA	GIT	PNV (1)	PNV (2)	PNV (3)	PNV (4)	PBA	LAA
752.30 (2+)	1503.798	ADAAVIEKLEAGFK	N-Acyl	2-15	Y (PMF)									
766.22 (2+)	1531.83	VDAAVIEKLEAGFK	N-Acyl	2-15	Y									
660.20 (2+)	1319.69	VSSLSLSELEGLK		152-164										
675.22 (2+)	1349.71	VSSLSLSELEGLK		152-164										
685.20 (2+); 457.07 (3+); 343.23 (4+)	1369.76	YLSKIDFDKLLK		30-40	Y (PMF)									
685.25 (2+); 457.14 (3+)	1369.73	YLSKEVFDQLK		30-40	Y									
692.33 (2+)	1383.78	YLTKEVFDKLLK		30-40										
699.20 (2+); 466.43 (3+)	1397.76	YLTKEIFDQLK		30-40										
532.18 (2+)	1063.57	AVFDQLKDK		34-42										
539.16 (2+)	1077.58	AVFDQLKEK		34-42	Y (PMF)									
539.14 (2+)	1077.58	ALFDQLKDK		34-42	Y									
603.28 (2+)	1205.68	ALFDQLKDKK		34-43	Y									
601.18 (2+)	1217.59	GYTFPLTGMSK	MSO	165-175										
616.09 (2+)	1231.60	GYTFPLTGMTK	MSO	165-175										
617.11 (2+)	1233.58	GYTFPLTGMSK	MSO	165-175										
525.59 (2+)	1050.52	FLQAANAQR	CysCAM	194-202	Y (PMF)									
817.76 (2+)	1634.81	SFLVWVNEEDQLR		217-229	Y									
829.26 (2+); 553.14 (3+)	1657.82	TFLVWVNEEDHLR		217-229	Y (PMF)									
860.09 (2+); 573.71 (3+)	1718.80	TFLVWVNEEDHLR	CysCAM	217-229	Y (PMF)									
759.85 (2+); 506.73 (3+)	1518.82	LEEYAGKYNLQVR		297-309	Y (PMF)									
766.80 (2+)	1532.83	LEEYAGKYNLQVR		297-309	Y									

^a *m/z*: mass/charge. MALDI observed; (Y): peptide found by MALDI-TOF MS; (PMF): peptide matched by peptide mass fingerprint with an AK entry in protein databases. (■) denotes the presence of a peptide, and (□) the absence. CysCAM, carboxymethylated cysteine; MSO, methionine sulfoxide; N-Acyl, acetyl N-terminal. SOP, *F. notialis*; PRA, *P. borealis*; GIT, *P. monodon*; PNV, *L. vannamei*; PNI, *F. indicus*; PBA, *F. merguensis*; LAA, *P. muelleri*. (1) Costa Rica population; (2) Argentina population; (3) Mozambique population; (4) Madagascar population.

(gil228014764), *Cancer antennarius* (gil171473177), and *Hemigrapsus oregonensis* (gil171473159).

MALDI-TOF MS peptide maps containing specific peaks are shown in Figure 1. Highlighted are those used for the definition and differentiation of the species studied. According to the results presented in Table S-2, 11 peaks with an *m/z* of 1050.51 (Figure 1a), 1507.73, 1604.81, 1683.82, 1796.95, 1807.92, 2015.98, 2031.99, 2033.98, 2190.05, and 2636.32 were present in all the species of the superfamily Penaeoidea. Several differences were also obtained among the Penaeoidea species, which may allow the classification of shrimps into two groups (Figure 1b–d): the family Penaeidae, presenting four specific peaks at *m/z* 1008.49, 1259.63, 1657.85, and 2599.31 Da, and the family Solenoceridae, with other 12 specific peaks at 994.45, 1263.62, 1286.63, 1424.68, 1442.73, 1532.78, 1541.89, 1718.85, 1789.89, 2315.25, 2644.36, and 2772.53 Da. Within the Penaeidae family, species-specific peaks were found (Figure 1e,f): one peak with an *m/z* of 1077.54 was specific to *P. monodon*; other peak with an *m/z* of 1463.75 was specific to *F. notialis*, while two peaks at 2285.06 and 2472.38 Da were specific for *F. merguensis*. Seventeen specific peaks were found for *P. borealis*, making the pattern very different from those found in the other species, and therefore to the family Pandalidae (e.g., peak at 1634.82 Da in Figure 1d). In addition, intraspecific variations, depending on the origin of the population considered, were found in *L. vannamei* and *F. indicus*: peaks with *m/z* 1148.53, 1369.71, 1513.78, 1531.75 (Figure 1g), 1761.92, 2271.18, 2287.23, 2431.23, 2548.35, and 3379.73 were found in *L. vannamei* from Costa Rica, while peaks with *m/z* 1503.71, 1631.90, 1775.89, 2301.21, and 3390.74 were found in *L. vannamei* from Argentina, being the last one an specific marker found only in this *L. vannamei* population. Differential peaks were found as well for two populations of *F. indicus*: 1133.58 (Figure 1h), 2301.21 and 2426.31 Da in the specimens coming from Mozambique, while 1622.80, and 2619.97 in those from Madagascar. The information provided by these specific peaks and the combination of the presence or absence of other peaks in the different groups of these species (Table S-2) is enough to clearly differentiate between the two superfamilies, the three families and the seven species studied. Furthermore, intraspecific differences described for *L. vannamei* and *F. indicus* may be used to differentiate populations from these two species.

MS/MS Characterization of Peptides from AK. Tryptic digests were further analyzed by MS/MS using a combination of LC-ESI-IT and nESI-IT. The spectra produced were used for database search and for manual *de novo* amino acid sequence interpretation. Results of the sequencing efforts are shown in Tables 2 and S-3; in the most extended, presented in the Supporting Information, a summary of the 131 peptide sequences obtained is presented, and Table 2 is a concise summary with the masses and sequences used for shrimp species identification. Peaks with coincident molecular weight were previously found by MALDI-TOF MS analysis for 54 peptides and 35 were previously assigned by PMF.

Several peptides were found to be specific for each of the superfamilies. Those with *M_r* 1147.60, 1050.52, 1667.82, 1683.82, 1796.94, 792.43, 2033.94, 1719.77, 1139.58, and 1304.68 Da were specific to the superfamily Penaeoidea. Specific peptides were found for each of the families included in this superfamily: peptides at 2599.26 and 1657.82 Da for the Penaeidae family, and peptides at 1397.76, 1403.67, 1718.80, 1442.80, 1130.59, 1286.70, 2772.40, 1016.55, 1789.97, and 1532.83 Da for the Solenoceridae family.

Within the Penaeidae family, species-specific peptides were also found. Examples are peptides at 3057.50, 3073.50, 3089.49, 3105.49, and 1984.93 Da for *P. monodon*; 1205.68 Da for *F. merguensis*; 2639.15 Da for *F. notialis*.

As previously revealed using MALDI-TOF spectra, intraspecific differences were found for *L. vannamei* and *F. indicus*: peptides at 1531.83, 1659.92, 1201.58, 1513.82, and 1761.94 Da were specific for *L. vannamei* from Costa Rica; 1383.78 and 3390.66 Da for *L. vannamei* from Argentina; 1215.60 and 1231.60 Da for *F. indicus* from Mozambique; and 1063.57 Da, was specific for *F. indicus* from Madagascar. Another pair of isobaric peptides, with a *M_r* of 1369.76 Da and sequences YLSKIDFDKLLK and YLSKEVFDQLK was found specific for *L. vannamei* from Costa Rica, and to *F. indicus* from Mozambique, respectively.

Several examples of the fragmentation spectra from species-specific peptides, together with the *m/z* of the precursor ions, are shown in Figures 2–4. The peak with *m/z* 525.59 (charge 2+) was chosen because it corresponds to one of the 10 peptides specific for the Penaeoidea superfamily (Figure 2). A pair of homologous peptides serves to differentiate the Pe-

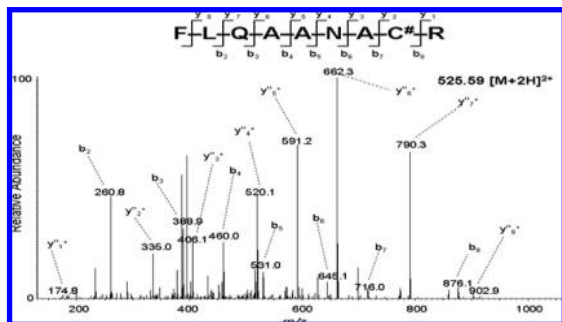


Figure 2. MS/MS spectrum of the AK peptide FLQAANAC*R. This peptide is present in all the species belonging to the superfamily Penaeoidea. C*, carboxymethylated cysteine.

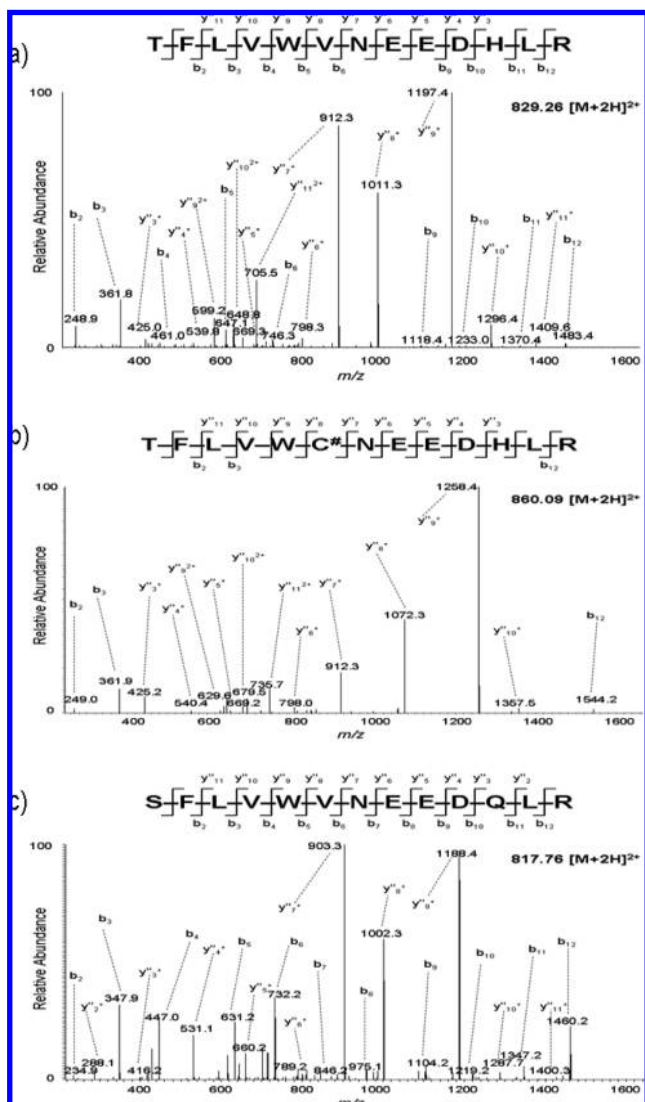


Figure 3. MS/MS spectra of the AK differential peptides: (a) TFLVWVNEEDHLR, (b) TFLVWVC*NEEDHLR, and (c) SFLVWVNEEDQLR, which are present in the Penaeidae, Solenoceridae and Pandalidae families, respectively. C*, carboxymethylated cysteine.

naeidae and Solenoceridae families: the peak at m/z 829.26 (2+), which is only present in Penaeidae, corresponds to the sequence TFLVWVNEEDHLR (Figure 3a), whereas the peak at 860.09 (2+), only present in Solenoceridae, corresponds to TFLC*WVNEEDHLR (Figure 3b). Both are absent in the Pandalidae family. A substitution of Val for Cys in the sequence is

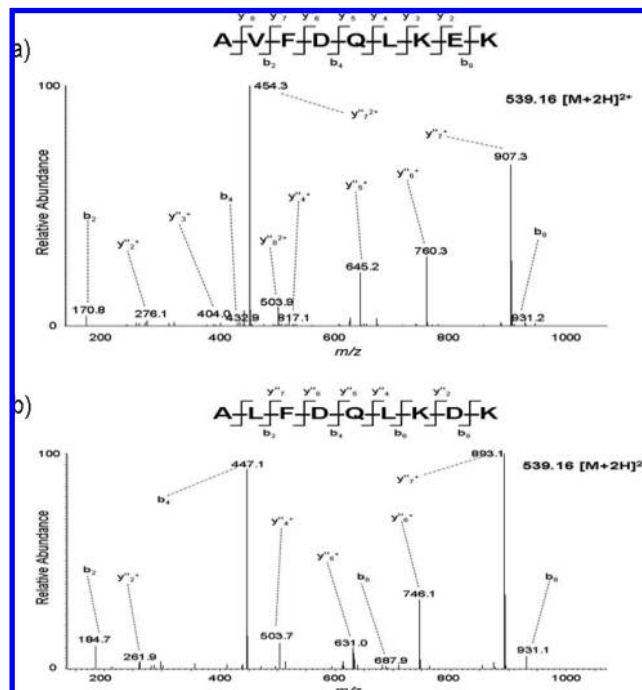


Figure 4. MS/MS spectra of the isobaric peptides (a) AVFDQLKEK and (b) ALFDQLKDK, which are present in AK from *P. monodon* and *F. merguensis*, respectively.

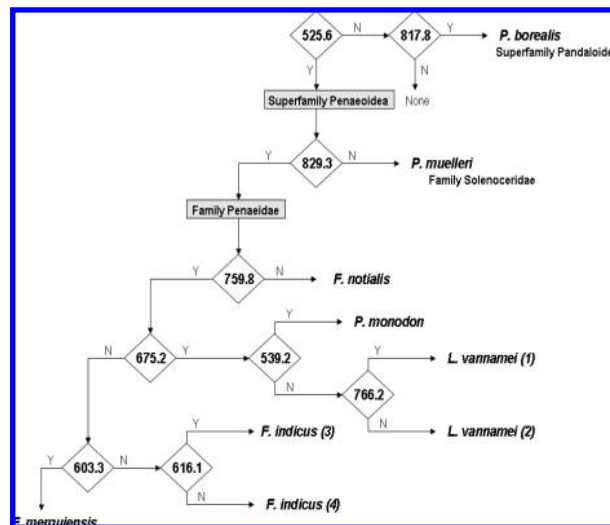


Figure 5. Flow diagram showing the systematic identification of shrimp species using nine differential peptides from AK. m/z of the doubly charged ions were 525.6, 817.8, 829.3, 759.8, 675.2, 539.2, 766.2, 603.3, and 616.1. "Y" denotes the presence and "N" the absence of a particular peptide.

responsible for the difference in the mass of the precursors and for the presence of different fragmentation diagnostic peaks at m/z 1011.3, 1197.4 (for the 829.26 precursor) and 1072.3, 1258.4 (for the 860.09 precursor).

Only one *P. borealis* specific peptide could be completely characterized by manual *de novo* sequencing: SFLVWVNEEDQLR, with a M_r of 1634.81 Da. This peak was chosen because it was only found in the Pandalidae family species (Figure 3c). This sequence has two substitutions with respect to that present in the Penaeidae species: a Thr for a Ser, and a His for a Gln.

The peak with a M_r of 1077.58 Da was found in *P. monodon* and *F. merguensis* samples, but it happened to be a mixture

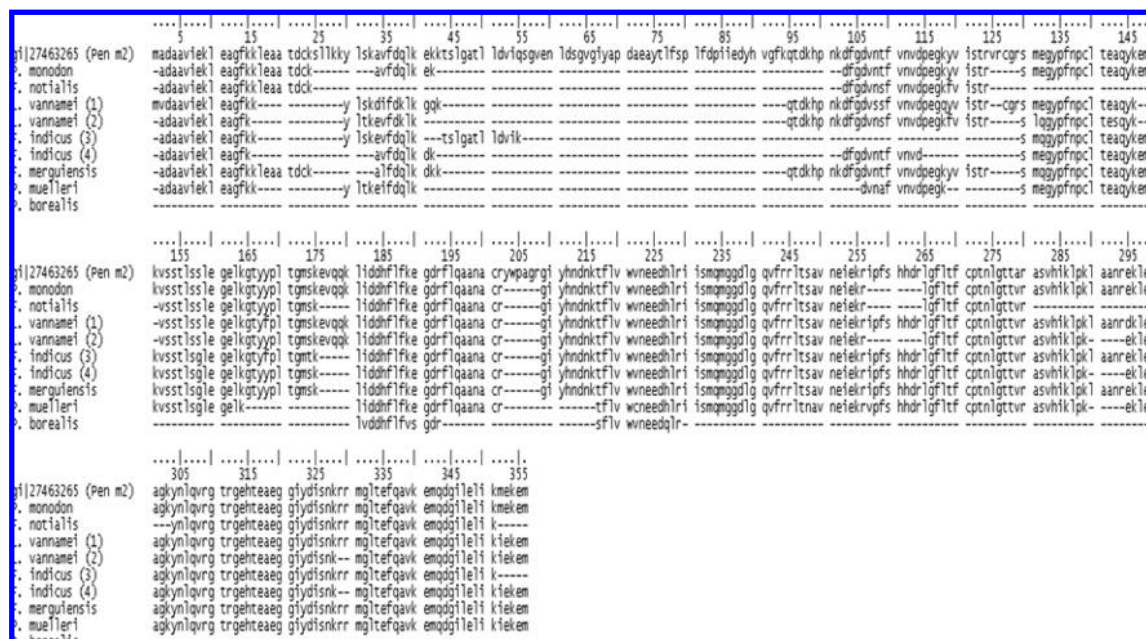


Figure 6. Alignment of the AK amino acid sequences, as obtained by MS/MS analysis, for all the shrimp species studied. The sequences were aligned with the allergen Pen m 2 sequence from *P. monodon* (accession number gi|27463265), which presents the highest degree of homology with them. (–) not determined amino acid. (1) Costa Rica; (2) Argentina; (3) Mozambique and (4) Madagascar populations.

of two isobaric peptides, with specific sequences for each of the species: AVFDQLKEK and ALFDQLKDK, respectively. There are a Val and a Glu for *P. monodon* and a Leu and an Asp for *F. merguensis*, in positions 2 and 8. Fragmentation spectra are shown in Figure 4; as expected, these two substitutions are responsible for the different diagnostic fragmentation peaks at m/z 454.3, 760.3, 907.3 for *P. monodon* (Figure 4a) and at m/z 447.1, 746.1, 893.1 for *F. indicus* (Figure 4b). As the MS/MS fragmentation patterns are different for each sequence, these peptides can be used in a multiple reaction monitoring (MRM) experiment to differentiate both shrimp species.

In addition to species-specific peptides, there are others which are always present in several species, being always absent in others. Combining the presence or absence in the samples of carefully selected peptides among those shown in Table 2, unequivocal shrimp species differentiation is clearly possible. Figure 5 presents one of the possible schematic flow diagrams which can be used for the systematic discrimination among the seven species studied and even, among the different geographical populations for *L. vannamei* and *F. indicus*.

The AK amino acid sequences from the different shrimp considered, as deduced from MS/MS data, were aligned together with the sequence Pen m 2 from *P. monodon*,²⁵ accessible in the protein databases, which presents the highest homology with those found in the present work (Figure 6). As AKs are highly conserved, only small differences are observed, but these differences may be used to distribute the species into different groups.

Conclusions

Proteomics methodology was used to find sequence similarities and differences among the arginine kinases from seven different shrimp species belonging to the order Decapoda. A high degree of homology was found among the AK sequences, but the sequence variations found are very useful to distinguish

the different members of these species, providing (i) a selective differentiation between the superfamilies Penaeoidea and Pandaloeidea and between the families Penaeidae, Solenoceridae, and Pandalidae; and (ii) an unequivocal identification of several shrimp species. Because of AK interspecific variability and the high concentration at which it is present in the muscle from shrimp, we can forecast that they could be used as a good biomarker for species identification. The identification and characterization of specific peptides, as done by this work, is the first step toward the designing of fast and cheap detection analyses. Identified species-specific peptides can be used to prepare antibody-based, easy-to-use kits for the sensitive detection of each of the species.

Acknowledgment. We thank Lorena Barros and Eva María Rodríguez for their excellent technical assistance. We also acknowledge members from CETMAR for their helpful collaboration in the collection of specimens for this study. This work was supported by the INIA National Food Program; Spanish Ministry for Education and Science (Project CAL-03-030-C2-2) and by the PGIDIT Research Program in Marine Resources (Project PGIDIT04RMA-261004PR), Xunta de Galicia.

Supporting Information Available: Results and Discussion, including Figure S-1, and Tables S-1, S-2 and S-3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Holthius, 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) 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.
- (7) 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.
- (8) 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.
- (9) 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.
- (10) 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.
- (11) 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.
- (12) Pascoal, A.; Barros-Velázquez, J.; Cepeda, A.; Gallardo, J. M.; Calomata, 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.
- (13) Piñeiro, C.; Barros-Velázquez, J.; Vázquez, J.; Figueras, A.; Gallardo, M. Proteomics as a tool for the investigation of seafood and other marine products. *J. Proteome Res.* **2003**, *2*, 127–135.
- (14) Martinez, I.; Friis, T. J. Application of proteome analysis to seafood authentication. *Proteomics* **2004**, *4*, 347–354.
- (15) López, J. L.; Marina, A.; Alvarez, G.; Vázquez, J. Application of proteomics for fast identification of species-specific peptides from marine species. *Proteomics* **2002**, *2*, 1658–1665.
- (16) 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.
- (17) Carrera, M.; Canas, B.; Pineiro, C.; Vazquez, 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.
- (18) An, H.; Marshall, M. R.; Otwell, W. S.; Wei, C. I. Species identification of raw and boiled shrimp by a urea gel isoelectric focusing technique. *J. Food Sci.* **1989**, *54*, 233–236.
- (19) Civera, T.; Parisi, E. The use of sodium dodecyl sulfate polyacrylamide gel electrophoresis for the identification of raw and cooked shrimps and crabs. *Ital. J. Food Sci.* **1991**, *3*, 149–157.
- (20) Shanti, K. N.; Martin, B. M.; Nagpal, S.; Metcalfe, D. D.; Rao, P. V. Identification of tropomyosin as the major shrimp allergen and characterization of its IgE-binding epitopes. *J. Immunol.* **1993**, *151*, 5354–5363.
- (21) Ortea, I.; Cañas, B.; Calomata, P.; Barros-Velázquez, J.; Gallardo, J. M. Arginine kinase peptide mass fingerprinting as a proteomic approach for species identification and taxonomic analysis of commercially relevant shrimp species. *J. Agric. Food Chem.* **2009**, *57*, 5665–5672.
- (22) Morrison, J. F. Arginine kinase and other invertebrate guanidine kinases. In *The Enzymes*, 3rd ed.; Boyer, P. D., Ed.; Academic Press: New York, 1973; Vol. 8, pp 457–486.
- (23) Watts, D. C. Creatine kinase (adenosine 5'-triphosphate creatine phosphotransferase. In *The Enzymes*, 3rd ed.; Boyer, P. D., Ed.; Academic Press: New York, 1973; Vol. 8, pp 383–455.
- (24) France, R. M.; Sellers, D. S.; Grossman, S. H. Purification, characterization, and hydrodynamic properties of arginine kinase from Gulf shrimp (*Penaeus aztecus*). *Arch. Biochem. Biophys.* **1997**, *345*, 73–78.
- (25) 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.
- (26) Piñeiro, C.; Barros-Velázquez, J.; Sotelo, C. G.; Pérez-Martín, 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.
- (27) 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.

PR900663D

SUPPORTING INFORMATION:

RESULTS AND DISCUSSION

2-DE

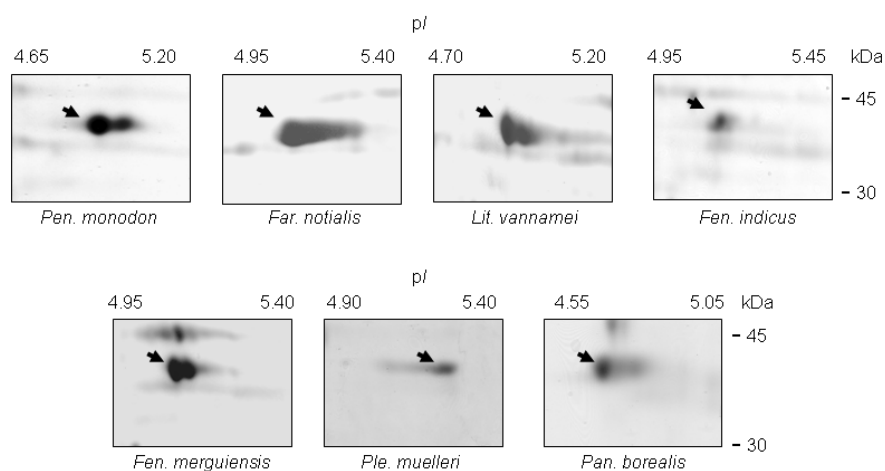


Figure S-1. Cropped regions of representative 2-DE gels showing the separated arginine kinase from all shrimps studied. *Pen*, *Penaeus*; *Far*, *Farfantepenaeus*, *Lit*, *Litopenaeus*; *Fen*, *Fenneropenaeus*; *Ple*, *Pleoticus*; *Pan*, *Pandalus*.

Table S-1. Mean pI of the analyzed spot, calculated from the 2-DE patterns of all individuals analysed of each species.

species	pI $\pm$ RSD
<i>Penaeus monodon</i>	4.99 $\pm$ 0.6
<i>Farfantepenaeus notialis</i>	5.22 $\pm$ 0.6
<i>Litopenaeus vannamei</i>	4.98 $\pm$ 0.8
<i>Fenneropenaeus indicus</i>	5.18 $\pm$ 0.3
<i>Fenneropenaeus merguiensis</i>	5.16 $\pm$ 1.0
<i>Pleoticus muelleri</i>	5.25 $\pm$ 1.1
<i>Pandalus borealis</i>	4.68 $\pm$ 1.6

MALDI-TOF MS analysis

Table S-2. Complete information of the MALDI-TOF MS analysis of arginine kinase spots in the seven shrimp species considered.

exp mass	exp mr	calc mr	delta	peptide	modification	Organism	Position	SOP	PRA	GIT	PNV	PNi	PBA	LAA
994.4523	993.445	993.4781	-0.03	VPFSSHDR	CysCAM	<i>Galapaganus conwayensis</i>	189-196							
1008.49	1007.482	1007.494	-0.01	IPFSHHR		<i>Penaeus monodon</i>	257-264							
1050.506	1049.499	1049.508	-0.01	FLQAANACR		<i>Penaeus monodon</i>	194-202							
1077.544	1076.537	1076.587	-0.05	AVFDQLKEK		<i>Penaeus monodon</i>	34-42							
1133.576														
1141.561												(3)		
1147.597	1146.589	1146.607	-0.02	LIDDHFLFK		<i>Penaeus monodon</i>	181-189							
1148.534											(1)			
1149.568	1148.561	1148.619	-0.06	EIFDQLKTR		<i>Solenopsis invicta</i>	33-41				(1)			
1158.583														
1164.651	1163.644	1163.595	0.04	RIPFSSHHR		<i>Litopenaeus vannamei</i>	256-264							
1196.634														
1205.667	1204.66	1204.682	-0.02	AVFDQLKEKK		<i>Penaeus monodon</i>	34-43							
1259.63	1258.623	1258.688	-0.07	LTSAVNEIEKR		<i>Penaeus monodon</i>	246-256							
1263.618														
1286.634														
1369.71	1368.702	1368.765	-0.06	YLSKDIFDKLK		<i>Marsupenaeus japonicus</i>	29-39				(1)			
1415.777	1414.769	1414.789	-0.02	RLTSAVNEIEKR		<i>Penaeus monodon</i>	245-256							
1424.679														
1438.708	1437.701	1437.74	-0.04	RRMGLTEYDAVK		<i>Periplaneta americana</i>	329-340							
1442.727	1441.719	1441.764	-0.04	LKGTIFYPLTGMSK		<i>Cimberis pilosa</i>	113-125							
1463.751	1462.743	1462.837	-0.09	RLVHAVNEIEKR		<i>Apis mellifera</i>	244-255							
1503.715	1502.708	1502.798	-0.09	ADAAVIEKLEAGFK	N-Acyl	<i>Penaeus monodon</i>	2-15				(2)			
1507.731	1506.723	1506.7	0.02	EGDRFLQAANACR	CysCAM	<i>Penaeus monodon</i>	190-202							
1513.781	1512.774	1512.826	-0.05	LAANRDKLEEVAGK		<i>Marsupenaeus japonicus</i>	289-302				(1)			
1518.732	1517.724	1517.82	-0.10	LEEYAGKYNLQVR		<i>Penaeus monodon</i>	297-309							
1525.779	1524.772	1524.855	-0.08	KYLTKEIFDQLK		<i>Solenopsis invicta</i>	28-39							
1531.753											(1)			
1532.782	1531.774	1531.836	-0.06	LEEVAAYKNLQVR		<i>Galapaganus conwayensis</i>	229-241							
1541.891	1540.884	1540.857	0.03	LAANREKLEEVAAK		<i>Homarus gammarus</i>	290-303							
1604.814	1603.806	1603.8	0.01	LIDDHFLFKEGDR		<i>Penaeus monodon</i>	181-193							
1606.776														
1622.799	1621.792	1621.85	-0.06	LQAGFKKLEAATDCK		<i>Marsupenaeus japonicus</i>	9-23					(4)		
1631.899	1630.892	1630.893	-0.01	ADAAVIEKLEAGFKK	N-Acyl	<i>Penaeus monodon</i>	2-16				(2)			
1634.824														
1638.826														
1651.839	1650.832	1650.822	0.01	IISMQMGGDLGQVFR		<i>Penaeus monodon</i>	230-244							
1657.851	1656.843	1656.826	0.02	TFLVWVNEDHLR		<i>Penaeus monodon</i>	217-229							
1666.815														
1667.81	1666.802	1666.817	-0.01	IISMQMGGDLGQVFR	MSO	<i>Litopenaeus vannamei</i>	230-244				(1)			
1683.816	1682.809	1682.812	-0.01	IISMQMGGDLGQVFR	2 MSO	<i>Penaeus monodon</i>	230-244							
1690.872														

Table S-2. Continued.

exp mass	exp mr	calc mr	delta	peptide	modification	Organism	Position	SOP	PRA	GIT	PNV	PNi	PBA	LAA
1718.85	1717.843	1717.788	0.05	TFLVWCNEEDHLR	CysCAM	<i>Galapaganus conwayensis</i>	149-161							
1719.841	1718.834	1718.775	0.06	GEHTEAEGGIYDISNK		<i>Litopenaeus vannamei</i>	313-328							
1758.86														
1761.917	1760.91	1760.942	-0.03	DKLEEVAGKYNLQVR		<i>Marsupenaeus japonicus</i>	294-308				(1)			
1766.903														
1775.891	1774.884	1774.958	-0.07	EKLEEVAGKYNLQVR		<i>Penaeus monodon</i>	295-309				(2)			
1789.895	1788.887	1788.862	0.03	KTFLVWCNEEDHLR		<i>Lepidophthalmus louisianensis</i>	88-101							
1796.95	1795.942	1795.929	0.01	LGFLTFCPNLGTTVR	CysCAM	<i>Litopenaeus vannamei</i>	265-280							
1807.92	1806.913	1806.923	-0.01	IISMQMGGDLGQVFRR		<i>Penaeus monodon</i>	230-245							
1810.85	1809.843	1809.873	-0.03	FLQAANACRYWPTGR	CysCAM	<i>Chasmagnathus granulata</i>	194-208							
1812.826														
1828.812														
1839.965	1838.957	1838.913	0.04	IISMQMGGDLGQVFRR	2 MSO	<i>Litopenaeus vannamei</i>	230-245							
1875.918	1874.91	1874.876	0.03	GEHTEAEGGIYDISNKR		<i>Penaeus monodon</i>	313-329							
2015.983	2014.976	2014.946	0.03	SLQGYFPNCLTEAQYK	CysCAM	<i>Artemia franciscana</i>	130-146							
2031.996	2030.989	2030.977	0.01	GEHTEAEGGIYDISNKR		<i>Penaeus monodon</i>	313-330							
2033.982	2032.975	2032.945	0.03	GTRGEHTEAEGGIYDISNK		<i>Penaeus monodon</i>	310-328							
2080.958														
2126.042														
2142.023														
2190.052	2189.045	2189.046	-0.01	GTRGEHTEAEGGIYDISNKR		<i>Penaeus monodon</i>	310-329							
2271.181											(1)			
2285.058														
2287.228											(1)			
2315.253														
2346.164	2345.156	2345.173	-0.02	DHLRIISMQMGGDLGEVYRR		<i>Pterapion sp.</i>	64-83							
2350.241														
2426.309												(3)		
2431.226	2430.219	2430.134	0.08	DFGDVSSFVNVDPEGQYVISTR		<i>Marsupenaeus japonicus</i>	102-123				(1)			
2472.377	2471.369	2471.197	0.17	DFGDVNTFVNVDPEGKYVISTR		<i>Penaeus monodon</i>	103-124							
2531.266														
2548.347	2547.34	2547.355	-0.02	RMGLTEFQAVKEMQDGILQLIK		<i>Marsupenaeus japonicus</i>	329-350				(1)			
2599.311	2598.303	2598.262	0.04	GIYHNDNKTLFWVNEEDHLR		<i>Penaeus monodon</i>	209-229							
2619.972	2618.964	2619.199	-0.23	SLQGYFPNCLTESQYKEMEAK	CysCAM	<i>Litopenaeus vannamei</i>	130-151					(4)		
2636.316	2635.308	2635.297	0.01	LIDDFHFLKEGDRFLQAANACR	CysCAM	<i>Penaeus monodon</i>	181-202							
2644.363	2643.356	2643.118	0.24	SMEGYFPNCLTETQYKEMESK	2 MSO	<i>Cancer antennarius</i>	130-151							
2772.532	2771.524	2771.397	0.13	VPFSHHDRGLGFLTFCPNLGTTVR	CysCAM	<i>Galapaganus conwayensis</i>	189-212							
2786.384	2785.376	2785.412	-0.04	IPFSHHDRGLGFLTFCPNLGTTVR	CysCAM	<i>Litopenaeus vannamei</i>	257-280							
3379.732	3378.724	3378.612	0.12	QTDKHPNKDFGDVSSFVNVDPEGQYVISTR		<i>Marsupenaeus japonicus</i>	94-123				(1)			
3390.738	3389.731	3389.579	0.15	YWPAAGRIYHNDNKTLVWCNEEDHLR	CysCAM	<i>Hemigrapsus oregonensis</i>	75-101				(2)			
3420.675	3419.668	3419.589	0.08	YWPTGRGIYHNDNKTLVWCNEEDHLR	CysCAM	<i>Chasmagnathus granulata</i>	203-229							

exp mass, observed protonated monoisotopic masses in Da; exp Mr, observed neutral masses in Da; calc Mr, calculated neutral masses in Da; delta, exp Mr - calc Mr; ; (■) denotes the presence of a peak, and () the absence. CysCAM, carboxymethylated cysteine; MSO, methionine sufoxide; N-Acyl, acetyl N-terminal. SOP, *F. notialis*; PRA, *Pandalus borealis*, GIT, *Penaeus monodon*; PNV, *Litopenaeus vannamei*; PNI, *Fenneropenaeus indicus*; PBA, *Fenneropenaeus merguensis*; LAA, *Pleoticus muelleri*. (1) Costa Rica population; (2) Argentina population; (3) Mozambique population; (4) Madagascar population.

MS/MS Characterization of Peptides from AK.

Table S-3. Complete information of peptide sequence data obtained by MS/MS for AKs from the shrimp species studied

m/z (z) observed	(M+H) ⁺ matched	Peptide Sequence	Modification	Position	MALDI observed	SOP	PRA	GIT	PNV (1)	PNV (2)	PNI (3)	PNI (4)	PBA	LAA
546.08 (3+)	1636.86	MVDAAVLEKLEAGFK	MSO	1-15	Y (PMF)									
429.49 (2+)	858.45	ADAAVIEK	N-Acyl	2-9										
443.38 (2+)	886.48	VDAAVIEK	N-Acyl	2-9										
752.30 (2+)	1503.798	ADAAVIEKLEAGFK	N-Acyl	2-15										
766.22 (2+)	1531.83	VDAAVIEKLEAGFK	N-Acyl	2-15		Y								
816.37 (2+); 544.49 (3+)	1631.89	ADAAVIEKLEAGFKK	N-Acyl	2-16		Y (PMF)								
830.39 (2+)	1659.92	VDAAVIEKLEAGFKK	N-Acyl	2-16										
840.70 (3+)	2520.29	ADAAVIEKLEAGFKKLEAATDCK	N-Acyl;CysCAM	2-24										
840.72 (2+)	1680.86	LEAGFKKLEAATDCK	CysCAM	10-24										
518.10 (2+)	1035.51	KLEAATDCK	CysCAM	16-24										
685.20 (2+); 457.07 (3+); 343.23 (4+)	1369.76	YLSKDIFDKLK		30-40	Y (PMF)									
685.25 (2+); 457.14 (3+)	1369.73	YLSKEVFDQLK		30-40	Y									
692.33 (2+)	1383.78	YLTKEVFDKLK		30-40										
699.20 (2+); 466.43 (3+)	1397.76	YLTKEIFDQLK		30-40										
410.48 (2+)	820.45	AVFDQIK		34-40										
417.52 (2+)	834.46	ALFDQLK		34-40										
532.18 (2+)	1063.57	AVFDQLKDK		34-42										
539.16(2+)	1077.58	AVFDQLKEK		34-42	Y (PMF)									
539.14 (2+)	1077.58	ALFDQLKDK		34-42	Y									
596.14 (2+)	1191.66	DIFDKLKGQK		34-43										
603.26 (2+)	1205.68	ALFDQLKDKK		34-43	Y									
615.99 (2+)	1230.72	TSLGATLLDVIK		44-55										
901.02 (3+); 676.12 (4+)	2701.27	QTDKHPNKDFGVDVNTFVNVDPEGK		95-118										
1127.09 (3+)	3379.61	QTDKHPNKDFGDVSSFVNVDPEGQYVISTR		95-124	Y (PMF)									
848.33 (4+)	3390.66	QTDKHPNKDFGDVNSFVNVDPEGKFVISTR		95-124	Y									
855.90 (4+)	3420.67	QTDKHPNKDFGVDVNTFVNVDPEGKYVISTR		95-124	Y									
869.70 (2+)	1738.78	DFGDVNSFVNVDPEGK		103-118										
876.77 (2+)	1752.80	DFGDVNTFVNVDPEGK		103-118										
1216.22 (2+); 811.14 (3+)	2431.13	DFGDVSSFVNVDPEGQYVISTR		103-124	Y (PMF)									
814.75 (3+)	2442.18	DFGDVNSFVNVDPEGKFVISTR		103-124										
824.71 (3+)	2472.19	DFGDVNTFVNVDPEGKYVISTR		103-124	Y (PMF)									
702.48 (2+)	1403.67	DVNAFVNVDPEGK		106-118										
808.78 (3+)	2424.04	CGRSMEGYFPNPCLTEAQYK	CysCAM ₁₂₇ ;MSO;CysCAM ₁₃₉	127-146										
1016.43 (2+)	2031.95	SLQGYFPNPCLTESQYK	CysCAM	130-146	Y									
1017.30 (2+); 678.47 (3+)	2033.91	SMQGYFPNPCLTEAQYK	CysCAM	130-146	Y									
1017.77 (2+)	2034.89	SMEGYFPNPCLTEAQYK	CysCAM	130-146										
1025.25 (2+); 683.96 (3+)	2049.90	SMQGYFPNPCLTEAQYK	MSO;CysCAM	130-146										
1025.85 (2+); 684.30 (3+)	2050.89	SMEGYFPNPCLTEAQYK	MSO;CysCAM	130-146										
880.39 (3+)	2639.15	SMEGYFPNPCLTEAQYKEMEAK	CysCAM;MSO ₁₄₈	130-151										
885.69 (3+)	2655.14	SMEGYFPNPCLTEAQYKEMESK	MSO ₁₃₁ ;CysCAM	130-151										

Table S-3. Continued.

m/z (z) observed	(M+H) ⁺ matched	Peptide Sequence	Modification	Position	MALDI observed	SOP	PRA	GIT	PNV (1)	PNV (2)	PNI (3)	PNI (4)	PBA	LAA
891.27 (3+)	2671.14	SMEGYFPNPCLTEAQYKEMESK	MSO ₁₃₁ ;CysCAM;MSO ₁₄₈	130-151										
983.35 (2+); 655.810(3+)	1965.97	EMEEKVVSSTLSGLELELK		147-164										
991.33 (2+); 661.15 (3+)	1981.96	EMEEKVSSTLSGLELELK	MSO	147-164										
660.20 (2+)	1319.69	VSSTLSGLELELK		152-164										
675.22 (2+)	1349.71	VSSTLSSLELELK		152-164										
844.87 (3+)	2532.28	VSSTLSSLELELKGYFYPLTGMSK		152-175										
850.12 (3+)	2548.29	VSSTLSSLELELKGYFYPLTGMSK	MSO	152-175	Y									
850.15 (3+)	2548.29	VSSTLSSLELELKGYTYPLTGMSK		152-175	Y									
1282.73 (2+); 855.33 (3+)	2564.28	VSSTLSSLELELKGYTYPLTGMSK	MSO	152-175										
1054.03 (3+); 790.92 (4+)	3160.61	VSSTLSSLELELKGYFYPLTGMSKEVQVK	MSO	152-180										
1059.47 (3+); 794.86 (4+)	3176.60	VSSTLSSLELELKGYTYPLTGMSKEVQVK	MSO	152-180										
601.18 (2+)	1201.58	GYFYPLTGMSK		165-175										
608.15 (2+)	1215.60	GYFYPLTGMTK		165-175										
608.99 (2+)	1217.59	GYFYPLTGMSK	MSO	165-175										
609.06 (2+)	1217.58	GYTYPLTGMSK		165-175										
616.09 (2+)	1231.60	GYFYPLTGMTK	MSO	165-175										
617.11 (2+)	1233.58	GYTYPLTGMSK	MSO	165-175										
879.92 (2+)	1758.80	(-)LVDDHFLFVSGDR		179-193	Y									
574.20 (2+); 383.12 (3+)	1147.60	LIDDHFLFK		181-189	Y (PMF)									
802.82 (2+); 401.91 (4+)	1604.80	LIDDHFLFKEGDR		181-193	Y (PMF)									
879.46 (3+); 659.83 (4+)	2636.30	LIDDHFLFKEGDRFLQAANACR	CysCAM	181-202	Y (PMF)									
754.29 (2+); 503.14 (3+)	1507.71	EGDRFLQAANACR	CysCAM	190-202	Y (PMF)									
525.59 (2+)	1050.52	FLQAANACR	CysCAM	194-202	Y (PMF)									
480.64 (2+)	960.44	GIYHNDNK		209-216										
867.03 (3+); 650.62 (4+)	2599.26	GIYHNDNKTFVLVWVNEEDHLR		209-229	Y (PMF)									
817.76 (2+)	1634.81	SFLVWVNEEDQLR		217-229	Y									
829.26 (2+); 553.14 (3+)	1657.82	TFLVWVNEEDHLR		217-229	Y (PMF)									
860.09 (2+)/ 573.71 (3+)	1718.80	TFLVWCNEEDHLR	CysCAM	217-229	Y (PMF)									
826.29 (2+); 551.13 (3+)	1651.82	IISMQMGGDLGQVFR		230-244	Y (PMF)									
834.49 (2+); 556.47 (3+)	1667.82	IISMQMGGDLGQVFR	MSO ₂₃₃ or MSO ₂₃₅	230-244	Y (PMF)									
842.25 (2+); 561.82 (3+)	1683.82	IISMQMGGDLGQVFR	MSO ₂₃₃ ;MSO ₂₃₅	230-244	Y (PMF)									
603.25 (3+)	1807.93	IISMQMGGDLGQVFR		230-245	Y (PMF)									
613.80 (3+)	1839.92	IISMQMGGDLGQVFR	MSO ₂₃₃ ;MSO ₂₃₅	230-245	Y (PMF)									
548.16(2+)	1095.53	MGGDLGQVFR	MSO	235-244										
474.64 (2+)	948.48	GGDLGQVFR		236-244										
472.52 (2+); 472.53 (3+)	1415.79	RLTSADVNEIEKR		245-256	Y (PMF)									
481.54 (3+)	1442.80	RLTNAVNEIEKR		245-256	Y									
552.10 (2+)	1103.58	LTSADVNEIEK		246-255										
565.79 (2+)	1130.59	LTSADVNEIEK		246-255										
630.24 (2+); 420.43 (3+)	1259.68	LTSADVNEIEKR		246-256	Y (PMF)									
643.86 (2+); 429.42 (3+)	1286.70	LTSADVNEIEKR		246-256	Y									
504.65 (2+); 336.76 (3+)	1008.49	IPFSHHDR		257-264	Y (PMF)									

Table S-3. Continued.

m/z (z) observed	(M+H) ⁺ matched	Peptide Sequence	Modification	Position	MALDI observed	SOP	PRA	GIT	PNV (1)	PNV (2)	PNI (3)	PNI (4)	PBA	LAA
924.77 (3+)	2772.40	VPFSHHDRLGFLTFCPTNLGTTVR	CysCAM	257-280	Y (PMF)									
929.45 (3+); 697.37 (4+)	2786.42	IPFSHHDRLGFLTFCPTNLGTTVR	CysCAM	257-280	Y (PMF)									
720.96 (2+)	1440.72	LGFLTFCPTNLGT	CysCAM	265-277										
898.85 (2+)	1796.94	LGFLTFCPTNLGTTVR	CysCAM	265-280	Y (PMF)									
496.79 (2+); 331.38 (3+)	992.61	ASVHIKLPK		281-289										
505.07 (3+)	1513.82	LAANRDKLEEVAGK		290-303	Y (PMF)									
509.97(3+)	1527.84	LAANREKLEEVAGK		290-303										
572.60 (3+); 572.39 (4+)	2287.24	LAANRDKLEEVAGKYNLQVR		290-309	Y									
767.70 (3+); 576.04 (4+)	2301.26	LAANREKLEEVAGKYNLQVR		290-309	Y									
501.65 (2+); 334.89 (3+)	1002.54	EKLEEVAGK		295-303										
494.58 (2+)	988.52	DKLEEVAGK		295-303										
508.66 (2+)	1016.55	EKLEEIAGK		295-303										
587.80 (3+)	1761.94	DKLEEVAGKYNLQVR		295-309	Y (PMF)									
888.42 (2+); 592.62 (3+)	1775.95	EKLEEVAGKYNLQVR		295-309	Y (PMF)									
597.36 (3+)	1789.97	EKLEEIAGKYNLQVR		295-309	Y									
759.85 (2+); 506.73 (3+)	1518.82	LEEVAGKYNLQVR		297-309	Y (PMF)									
766.80 (2+)	1532.83	LEEIAGKYNLQVR		297-309	Y									
396.53 (2+)	792.43	YNLQVR		304-309										
1017.19 (2+); 678.49 (3+); 509.24 (4+)	2033.94	GTRGEHTEAEGGIYDISNK		310-328	Y (PMF)									
548.18 (4+)	2190.04	GTRGEHTEAEGGIYDISNKR		310-329	Y (PMF)									
587.34 (4+); 469.91 (5+)	2346.14	GTRGEHTEAEGGIYDISNKRR		310-330	Y									
860.32 (2+); 573.94 (3+)	1719.77	GEHTEAEGGIYDISNK		313-328	Y (PMF)									
625.92 (3+); 469.66 (4+)	1875.87	GEHTEAEGGIYDISNKR		313-329	Y (PMF)									
508.77 (4+)	2031.98	GEHTEAEGGIYDISNKRR		313-330	Y (PMF)									
427.00 (3+)	1279.67	RMGLTEFQAVK		330-340										
432.55 (3+)	1295.68	RMGLTEFQAVK	MSO	330-340										
645.98 (4+)	2581.34	RMGLTEFQAVKEMQDGILELIK	MSO ₃₃₁ ;MSO ₃₄₂	330-351										
562.15 (2+)	1123.57	MGLTEFQAVK		331-340										
570.16 (2+)	1139.58	MGLTEFQAVK	MSO	331-340										
1197.10 (2+); 798.42 (3+)	2393.24	MGLTEFQAVKEMQDGILELIK		331-351										
1205.19 (2+); 803.64 (3+)	2409.23	MGLTEFQAVKEMQDGILELIK	MSO ₃₃₁ or MSO ₃₄₂	331-351										
1213.06 (2+); 809.01 (3+)	2425.24	MGLTEFQAVKEMQDGILELIK	MSO ₃₃₁ ;MSO ₃₄₂	331-351										
691.75 (4+)	2763.46	MGLTEFQAVKEMQDGILELIKIEK		331-354										
695.45 (4+)	2779.45	MGLTEFQAVKEMQDGILELIKIEK	MSO ₃₃₁	331-354										
699.51 (4+)	2795.46	MGLTEFQAVKEMQDGILELIKIEK	MSO ₃₃₁ ;MSO ₃₄₂	331-354										
1008.49 (3+); 756.66 (4+)	3023.55	MGLTEFQAVKEMQDGILELIKIEKEM		331-356										
1013.82 (3+); 760.61 (4+)	3039.55	MGLTEFQAVKEMQDGILELIKIEKEM	MSO ₃₃₁ or MSO ₃₄₂ or MSO ₃₅₆ two of MSO ₃₃₁ or MSO ₃₄₂ or MSO ₃₅₆	331-356										
1019.05 (3+); 764.50 (4+)	3055.54	MGLTEFQAVKEMQDGILELIKIEKEM	MSO ₃₃₁ or MSO ₃₄₂ or MSO ₃₅₆	331-356										
765.14 (4+)	3057.50	MGLTEFQAVKEMQDGILELIKMEKEM	MSO ₃₃₁ or MSO ₃₄₂ or MSO ₃₅₆	331-356										
1024.50 (3+); 768.79 (4+)	3071.54	MGLTEFQAVKEMQDGILELIKIEKEM	MSO ₃₃₁ ;MSO ₃₄₂ ;MSO ₃₅₆ MSO ₃₃₁ or MSO ₃₄₂ or MSO ₃₅₆ ;MSO ₃₅₆	331-356										
1025.14 (3+); 769.09 (4+)	3073.50	MGLTEFQAVKEMQDGILELIKMEKEM		331-356										

Table S-3. Continued.

<i>m/z</i> (z) observed	(M+H) ⁺ matched	Peptide Sequence	Modification	Position	MALDI observed	SOP	PRA	GIT	PNV (1)	PNV (2)	PNI (3)	PNI (4)	PBA	LAA
1030.56 (3+)	3089.49	MGLTEFQAVKEMQDGILELIKMEKEM	MSO ₃₃₁ or MSO ₃₄₂ ;MSO ₃₅₂ ;MSO ₃₅₆	331-356										
1035.39 (3+); 777.03 (4+)	3105.49	MGLTEFQAVKEMQDGILELIKMEKEM	MSO ₃₃₁ ;MSO ₃₄₂ ;MSO ₃₅₂ ;MSO ₃₅₆	331-356										
644.66 (2+)	1288.67	EMQDGILELIK		341-351										
652.71 (2+)	1304.68	EMQDGILELIK	MSO	341-351										
558.82 (3+)	1674.90	EMQDGILELIKIEK	MSO	341-354										
640.22 (3+)	1918.98	EMQDGILELIKIEKEM		341-356										
975.86 (2+); 650.88 (3+)	1950.98	EMQDGILELIKIEKEM	MSO ₃₄₂ ;MSO ₃₅₆	341-356										
662.22 (3+)	1984.93	EMQDGILELIKMEKEM	MSO ₃₄₂ ;MSO ₃₅₂ ;MSO ₃₅₆	341-356										

m/z: mass/charge. MALDI observed; (Y): peptide found by MALDI-TOF MS; (PMF): peptide matched by peptide mass fingerprint with an AK entry in protein databases. (■)

denotes the presence of a peptide, and () the absence. CysCAM, carboxymethylated cysteine; MSO, methionine sufoxide; N-Acyl, acetyl N-terminal. SOP, *F. notialis*; PRA, *Pandalus borealis*, GIT, *Penaeus monodon*; PNV, *Litopenaeus vannamei*; PNI, *Fenneropenaeus indicus*; PBA, *Fenneropenaeus merguensis*; LAA, *Pleoticus muelleri*. (1) Costa Rica population; (2) Argentina population; (3) Mozambique population; (4) Madagascar population.