UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

Dipartimento Medicina Veterinaria Produzioni Animali

Tesi di Dottorato

"Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* species to provide specific barcode markers"

Tutor
Prof.ssa Tiziana Pepe

Tutor
Dr. Paolo Sordino

Coordinatore
Prof.Giuseppe Cringoli

Candidato
Dr. Celestina Mascolo

«Uno, ricordatevi sempre di guardare le stelle, non i piedi. Due, non rinunciate al lavoro: il lavoro dà significato e scopo alla vita, che diventa vuota senza di esso. Tre, se siete abbastanza fortunati a trovare l'amore, ricordatevi che è lì e non buttatelo via».

Stephen Hawking

Index

List of abbreviations

12S rRNA and 16S rRNA	12S and 16S Ribosomal RNA, respectively
ASFIS	List of Species for Fishery Statistics Purposes
ATP6 and ATP8	Atpase Subunit 6 and 8, respectively
BLAST	Basic Local Alignment Search Tool
bp	Base Pair
BSA	Bovine Serum Albumin
COI–III	Cytochrome C Oxidase Subunits I–III, respectively
Cytb	Cytochrome B
ddNTP	Dideoxynucleotides
DM	Ministerial Decreet
DNA	Deoxyribonucleic Acid
dNTP	Nucleoside Triphosphate
ELISA	Enzyme-Linked Immunosorbent Assay
EMA	European Medicines Agency
EU	European Union
FAO	Food and Agriculture Organization of The United Nations
FAPs	Fisheries and Aquaculture Products
FDA	Food and Drug Administration
GAO	Government Accountability Office

GC	Guanine-Cytosine content
ITS1-2	Internal Transcribed Spacer
IUCN	The International Union for Conservation of Nature
K2P	Kimura-2-Parameter
Kb	Kilobase
MHC	Major Histocompatibility Complex
*mt*DNA	Mitochondrial *DNA*
NAD1-6 and 4L	NADH Dehydrogenase Subunit 1–6 And 4 L
NCBI	National Center For Biotechnology Information
*n*DNA	Nuclear *DNA*
NGS	Next Generation Sequencing
NTS	Neurothesin
NTS	Neurotensin Gene
PCR	Polymerase Chain Reaction
Rep-origin L strand/ OL	Origin of L-Strand Replication
RFLP	Restriction Fragment Length Polymorphism,
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RT-PCR or qPCR	Reverse Transcriptase-Polymerase Chain Reaction

SNPs	Single Nucleotide Polimorfism
Tm	Melting Temperature
tRNA	Transfer RNA
US	United States
USD	United States Dollar
VNTR	Variable Number of Tandem Repeats

List of Figures

List of Figures

List of Tables

List of Tables

Abstract

The growing global diffusion of seafood for human consumption requires always more accurate sanitary and quality controls. Among the globally marketed fish, the species belonging to the family *Sparidae* are excellent food-fishes of high economic value.

The *Sparidae* family comprises about 38 genera and 159 species, some of which are highly appreciated as seafood. In Italy, the fish species of commercial interest are defined by the law (DM n. 19105 del 22 Settembre 2017). Among them, 41species belong to the Sparidae family. The identification of these species is difficult even when external characters are preserved due to their morphological similarity.
Species identification becomes even harder to achieve after industry processing, when distinctive external traits are removed. Nevertheless, despite their similarity, sparid species have different organoleptic quality that corresponds to variable prices in fish markets. Consequently, substitution of commercially important *Sparidae* species is common. The sparid specie with highest commercial value, *Dentex,* is often replaced with less expensive species or lower quality alternatives. All this premised, *Sparidae* species identification requires immediate resolutions. Research on fish mitochondrial *DNA* (*mt*DNA, mitogenome) has led to substantial advances in the fields of species authentication and population biology.
The main species' specific *DNA* sequences used as markers for fish species identification belong to the mitochondrial genes encoding ribosomal ***16S*** and ***12S*** subunits, cytochrome b (***Cytb***), and cytochrome c oxidase I (***COI***). However, current research shows that mitochondrial *DNA* markers perform well for certain species but may be less discriminating for others. Therefore, they cannot be used indiscriminately for the identification of all fish species. This situation has led to the formulation of the proposal to study and analyze the complete *mt*DNA sequence with the aim to identify mitochondrial markers or multiple marker approaches with higher and more specific discrimination capacity. At present, *mt*DNA genomes of *Sparidae* fishes are not conspicuously represented in the scientific literature, which makes difficult to understand if currently used genetic markers are the most effective for sparid species identification.
This thesis project had the aim to study and analyze Sparids' complete *mt*DNA sequence to identify mitochondrial markers or a multiple marker approach to achieve a higher and more specific discrimination capacity.

Abstract

We focused on the genera *Dentex, Pagellus* and *Pagrus*, listed in the Ministerial Decree.

1.Introduction

1.1. General overview

Fish and fish products have a central role in nutrition, as they embody an appreciated source of nutrients and micronutrients, which are essential for health benefits. The importance of fish is even higher for pregnant women and very young children, as it contains many of the vitamins and minerals needed. In fact, fish can be a crucial component of a healthy diet, as it contributes to neurodevelopment during the fetal's advance or young child's growth. Additionally, several studies highlight the beneficial effects of fish consumption in mental health and prevention of cardiovascular diseases, stroke and age-related macular degeneration (Mascolo et al., 2013; Papamichael et al., 2018; Venugopal, 2018; Venugopal, 2019; Farmery et al., 2018). Even small quantities of fish can offer essential amino acids, fats and micronutrients, such as iron, iodine, vitamin D and calcium, which are often absent in vegetable-based diets. Furthermore, seafood products contain the highest concentration of essential long-chain omega-3 polyunsaturated fatty acids of any foodstuff (Tacon & Metian, 2013). Experts agree that the positive effects of high fish intake largely compensate the potential negative effects associated with contamination or other safety risks (FAO & WHO, 2007). This consciousness has led to a substantial increase in the seafood products demand, thanks to the growing spread of healthy diets. Fish provided about 3.2 billion people with nearly 20 percent of their average per capita consumption of animal protein. In 2015, fish accounted for about 17 percent of animal protein consumed by the global population (FAO, 2018). The fish sector in the following few years is projected to record an additional growth in per capita consumption, not only in Italy, which will be matched by an increase in world production that will reach 193.9 million tons in 2026 (+ 10 percent compared to 2017) (FAO, 2018). Annual per capita fish feasting has grown gradually in developing regions (from 6.0 kg in 1961 to 19.3 kg in 2015) and in low-income food-deficit countries (from 3.4 to 7.7 kg during the same period) but is still suggestively higher in industrialized countries (24.9 kg in 2015), even if the gap is reducing (FAO, 2018) (Figure 1) (Figure 2).

Figure 1 Apparent seafood consumption per capita, average 2013-2015. Source by "The state of world fisheries and aquaculture 2018". FAO, 2018.

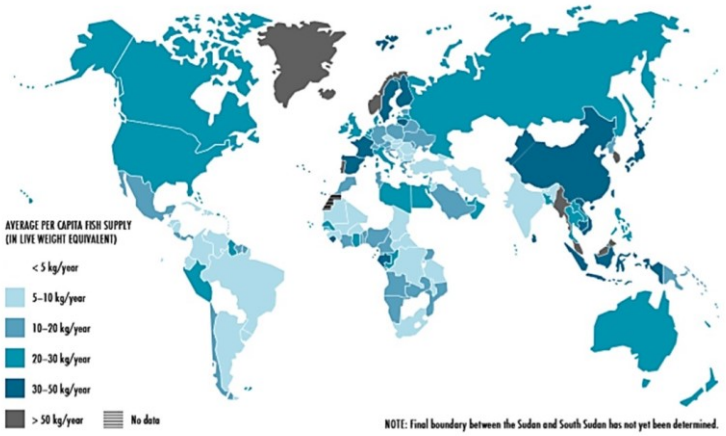
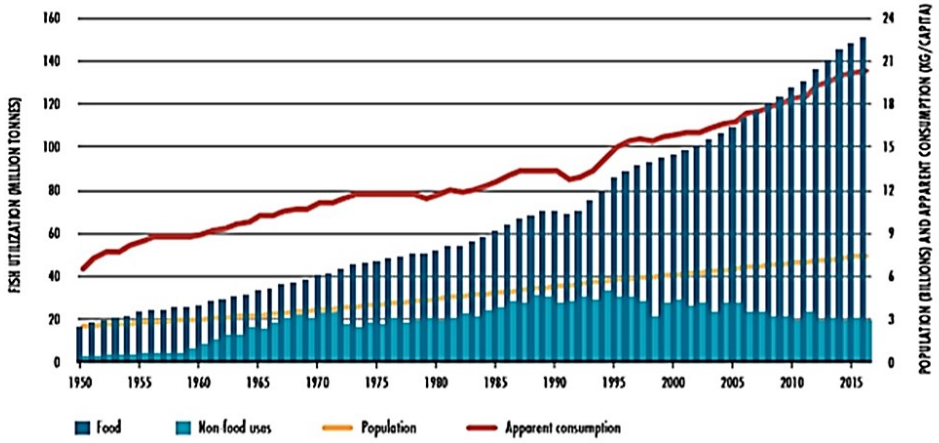


Figure 2. World fish utilization and apparent consumption. Source by "The state of world fisheries and aquaculture 2018". FAO, 2018.

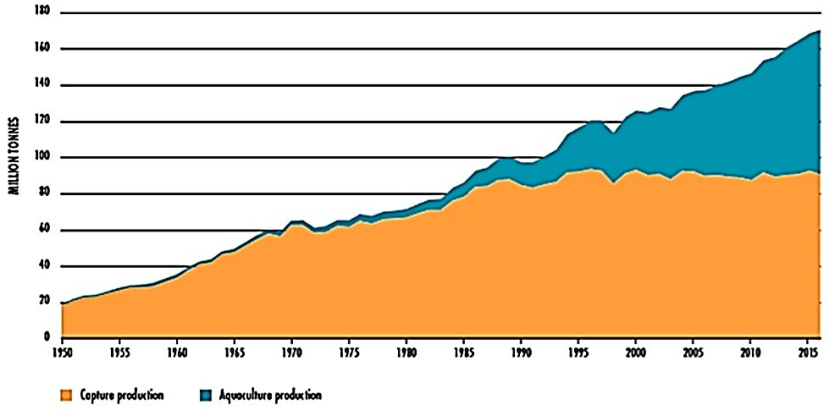


NOTE: Excludes aquatic mammals, crocodiles, alligators and caimans, seaweeds and other aquatic plants

In Italy above all, the consumption of fish products has been growing steadily of + 2 percent a year for about twenty years. In 2018, per capita consumption of 28 kg was recorded against a European average of 25 kg (EUMOFA, 2018).

Global total capture fisheries production, as derived from the FAO capture database, was 90.9 million tonnes in 2016, The 2016 global catch shows an increase of 2.0 percent over the previous year and of 10.5 percent in comparison to the 2005–2014 average (FAO, 2018).To satisfy this demand, global fishery and aquaculture production peaked at about 171 million tonnes in 2016, with aquaculture representing 47 percent of the total and 53 percent if non-food uses are excluded (FAO, 2018). The full first sale value of fisheries and aquaculture production in 2016 was valued at USD 362 billion, of which USD 232 billion was from aquaculture production (FAO, 2018) (Figure 3). Aquaculture has been responsible for the continuing remarkable growth of fish for human consumption.

Figure 3. World capture fisheries and aquaculture production. Source by "The state of world fisheries and aquaculture 2018". FAO, 2018.

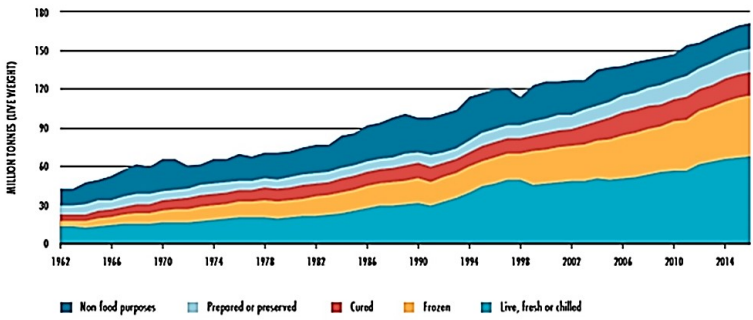


NOTE: Excludes aquatic mammals, crocodiles, alligators and caimans, seaweeds and other aquatic plants

China, which is the major producer of farmed food fish in 2016, has farmed more than the rest of the world joint every year since 1991(FAO, 2018). The other main producers in 2016 were India, Indonesia, Viet Nam,

Bangladesh, Egypt and Norway. Of the 171 million tonnes of total fish production in 2016, about 88 percent (over 151 million tonnes) was utilized for direct human feeding (FAO, 2018). Alive, fresh or chilled is habitually the most chosen and highly priced form of fish and characterises the largest share of fish for direct human consumption (45 percent in 2016), followed by frozen (31 percent) (Figure 4).

Figure 4. Utilization of world fisheries production, 1962–2016. Source by "The state of world fisheries and aquaculture 2018". FAO, 2018.



Currently, fish and fish products are also the most traded food stuffs in the world. In 2016, around 35 percent of global fish production was internationally traded in numerous forms, such as for human consumption or non-edible purposes (FAO 2018). The 60 million tonnes (live weight equivalent) of total fish and fish products exported in 2016 represent a 245 percent increase over 1976 (FAO 2018). During the same period, world trade in fish and fish products also grew significantly in value terms, with exports rising from USD 8 billion in 1976 to USD 143 billion in 2016.

1.2. Sparids: fisheries and acquaculture

In 2016 the world fish production was of 16.939.123,7 tonnes, of which about 47% comes from aquaculture and the residue from fishing. As for aquaculture, *Percoidei* represent the 11% of total production, with a quantity of fish produced equal to 8.104.243 tonnes. This group of fish constitutes the largest part of aquaculture world production after the Cypriniforms (63%) and followed by *Salmoniformes* (7%). Many species of fish caught and farmed for commercial purposes belong to the order of the Perciforms, including the tilapies (family *Cichlidae*), the sparids, the

sea bass (family *Latidae*), the seriole (family *Carangidae*), the groupers (family Serranidae), the barramundi (family *Latidae*), the cobia (family *Rachycentridae*), the perca (family *Percidae*), etc. As for aquaculture, the *Cichlidae* family (Tilapia) represents the 65% of the amount produced, while the *Sparidae* family is 6.8% with a total production of 12,817.54 tonnes (fishing and aquaculture). In 2006 the production of sparids was 528,238 tonnes but in 2016 reached 661,260.05 tonnes, with a 52% from aquaculture (339.176.05 tonnes). and the remaining 48% of production from fisheries (322.084 tonnes) (FAO FishStatPlus, 2018)

The FAO data on catches concern 57 species (ASFIS - List of Species for Fishery Statistics Purposes) (FAO, 2018a) as reported in the Table 1.

However, about 40% of the catch (312.672,49 tonnes) are designated as "porgies, seabreams nei", without indicating the different species. Two species recorded a production exceeding 10.000 tons (Silver seabream and Black seabream) while the remaining species were less important. The main farmed species in the world are: Blackhead seabream, Blackspot(=red) seabream, Common dentex, Common pandora, Common two-banded seabream, Gilthead seabream, Goldlined seabream, Japanese seabream, Pink dentex, Red porgy, Sharpsnout seabream, Silver seabream, Sobaity seabream, White seabream. Among these, only for four species (Black seabream, Gilthead seabream, Red porgy, Silver seabream) recorded a volume in tons greater than 10.000 tons in 2016. In 2007, approximately 64,32% of the 339.176,05 tonnes produced came from Asia (Silver seabream), while about 24,14% came from Europe and Africa. (Over Gilthead seabream) (FAO FishStatPlus, 2018).

Table 1. Global production by production source 2007-2016. The catches (wild and farmed) are intended for the whole world and expressed in tonnes (100 Kg) (FishstatJ). In: .www.FAO.org/fishery/statistics/software/fishstatj/en FAO.

Species	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Angolan dentex	2.488	2.780	1.198	1.499	2.161	1.592	428	341	602	448
Annular seabream	1.807	2.088	1.775	2.236	1.785	1.729	1.348	1.304	947	1.083
Axillary seabream	1.288	959	1.220	2.543	1.269	1.332	2.572	2.578	3.079	2.686
Black musselcracker	4	6	7	5	3	3	5	7	8	10
Black seabream	7.956	6.121	10.585	9.484	9.008	7.333	7.585	9.317	7.057	11.461
Blackhead seabream	3.998	2.635	2.477	2.983	2.023,1	2.502	1.547	1.510,21	1.733	1.772,99
Blackspot(-red) seabream	1.891	1.891	1.662,35	1.466,2	1.315,04	1.268,5	1.618,13	1.793	1.807	1.519
Bluespotted seabream									45	221
Bogue	33.362	37.706	36.267	34.173	39.643	40.697	39.312	39.560	36.836	40.850
Carpenter seabream	294	250	323	345	444	410	584	606	557	726
Common dentex	1.920	2.370	1.625,05	1.616	1.222	1.333,64	1.890	1.466,36	1.403,13	1.397,14
Common pandora	6.321	7.246	13.223	14.716	11.412,1	22.154,2	15.714,5	14.779,3	14.185	13.901,04
Common Two-banded seabream	435	435,1	527	510,8	452	478	624	851	887	
Congo dentex	559	558	123	94	179	109	94	35	128	1.539
Crimson seabream										
Dagger(head) breams nei[1]	194	234	216	227	252	303	272	272	229	279
Dentex nei[1]	28.534	30.723	37.871	35.065 F	31.527 F	27.420 F	25.280 F	22.820	19.374	29.742
Gilthead seabream	132.015	137.712,35	144.079,17	150.668,28	141.888,57	150.109,26	165.007,13	166.319,59	175.770,47	195.410,42
Goldlined seabream	8	12	12	4	4	31	3	6	2	23
Goldsilk seabream	68	72,1	70	46	63	76	74	111	99 F	97 F
Halfbara seabream	1.555	1.498	1.527	1.502	1.352	1.981	1.838	1.855	1.628	1.575
Japanese seabream			10	14	209,3	2				
Jobfish										
Jobhead porgy						9	8	12	13	16
Karanteen seabream	78	37	48	35	66	70	52	65	57	66
King soldier bream	4.918	6.829	5.693	5.779	6.490	6.897	5.877	4.830	5.268	5.096
Large-eye dentex	1.665	1.362	1.867	1.897	2.240	2.150	3.524	3.479	3.657	5.020
Pacific porgy						41				3
Pandoras nei	5.083	3.776	4.775	6.300	5.130	4.222	3.870	4.669	4.800	4.875
Panga seabream	2.305	1.880	2.183	2.249	1.903	891	1.805	1.718	1.483	1.747
Pargo breams nei[1]	4.502	9.476	6.783	7.610	9.154	6.695,02	8.354	10.027,19	4.997	6.116
Pinfish	41	27				20	61	73	43	44
Pink dentex						2	1	77	93	63
Porgies	416	339	361	335	395	276	274	321	328 F	225 F
Porgies, seabreams nei[1]	259.271	246.512,26	271.249	283.137,04	290.899,3	291.243	295.025,11	301.013,88	304.933	312.672,49
Red pandora	118.58	10.036	8.948	8.202	6.847	7.178	6.825	6.850	9.347	10.071
Red porgy	11.400	14.561	15.455	14.083	13.389,8	11.434,88	11.155,5	9.826,3	9.648,6	11.733,2
Red steenbras	1.890	3	5	3	2	2	-	-	-	-
Saddled seabream	1.890	1.683	1.851	1.758	1.616	1.412	1.287	1.555	1.446	1.632
Salema	4.774	4.176	3.857	4.236	3.185	3.185	3.526	4.101	4.414	4.661
Sand steenbras	2.738	2.638	4.226	3.891 F	3.948	3.441 F	3.513 F	3.652	3.076	1.887
Santer seabream	163	159	77	79	71	137	151	176	163 F	160 F
Sargo breams nei[1]	4.388	4.463	4.782	5.778,4	5.299,3	4.404,3	3.727,24	4.057,14	3.013,42	3.955,42
Scup	4.212	2.375	3.879	4.659	6.618	6.739	8.091	7.196	7.693	7.135
Sharpsnout seabream	159	50	51	41	273	411,1	315,1	578,2	372,4	252,8
Sheepshead	746	782	825	751	671	775	922	775	603	615
Silver seabream	101.194	105.401	105.749	98.529	91.625	84.100	82.062	88.592	92.746	95.127
Sobaity seabream	316F	268	210	759,49	670,09 F	684 F	658 F	116,89	111	106
South American silverporgy	26	82	14	26	60	40	31	32	18	18
Steenbrasses nei[1]	3				0 0					
Stumpnoses nei[1]	334	614	1.053	870 F	670 F	460	520 F	586	580	580
Twobar seabream	3.495	3.933	3.604	3.896	4.542	4.995	5.532	5.683	6.560	7.041
Western Atlantic seabream		25	23	25	22	21	15	14	22	1
White seabream	2.491	2.658	4.243,22	4.373,4	3.983,58	3.452,64	3.786,83	3.398,79	3.802,97	4.093,05
White steenbras			-	-	-	-	-	-	-	-
White stumpnose	24	51	50	78	99	106	89	82	87	102
Yellowback seabream										
Yellowfin seabream	480	544	658	828	773	842	654	657	556	663
TOTAL TONNES	653.683	659958,81	707.444,79	720.504,61	706.556,18	706.560,54	717.336,54	729.270,85	740.474,99	791.404,51

" nei " = Not Elsewhere Included; " F " = FAO estimate from available sources of information.

1.2.1. Aquaculture of EU species (Mediterranean-Atlantic)

- **1.2.1.1 *Dentex* (Common dentex)**

In 2016 the total catches of the Dentice[1] were around 139.7,11 tonnes (FAOFishStatPlus, 2018), with 1.353 tonnes from fishing and 44,11 tonnes from aquaculture. Most of the fish (90% of world production) comes from the Mediterranean Sea, from country like Greece, Libya, Italy, Tunisia and Spain. The farmed fishes originate only from two countries: Turkey has a production of 43 tonnes, while Croatia 1.11 tonnes.

- **1.2.1.2. *Dentex gibbosus* (Pink dentex)**

The catches of the Dentice gibboso[1] are reported by FAO only in the Mediterranean, especially in Croatia and Turkey. Specifically, the fishing was 2 tonnes (in Croatia alone) and the production for aquaculture was 61 tonnes for Turkey (FAOFishStatPlus, 2018).

1.2.1.3. *Diplodus puntazzo* (Sharpsnout sea bream)

In the last 10 years there have not been many FAO data on catch trends for Sarago pizzuto[1], but only in the last 5 years there has been an increase in catches, both in fishing (95 tonnes) and in aquaculture (157,8 tonnes) (FAOFishStatPlus, 2018). Between the leading countries in the fishing trade, there is Italy with 70 tonnes in 2016, while for the aquaculture there is Greece with 140,8 tonnes.

1.2.1.4. *Diplodus sargus* (White sea bream)

The total production of Sarago maggiore[1] in 2016, reached 4.093 tonnes, of which 4.078 deriving from fisheries, the remaining 15,05 were obtained through aquaculture (FAOFishStatPlus, 2018). According to the FAO data, the wild catches of this species came from Egypt, Greece, Spain, Libya and Tunisia and aquaculture production derived from Spain, Portugal, Italy and Greece.

[1] denomination in Italian according to D.M. MIPAAF, 22 Settembre 2017.

1.2.1.5. *Diplodus vulgaris* (Common two-banded seabream)

In the FAO data of the last 10 years, the Sarago fasciato[1] is constantly present only in relation to the fishing catches, which in 2016 was 887 tonnes, while for aquaculture the data go back to 2010 (11,8 tonnes) and 2008 (18,1 tonnes) (FAOFishStatPlus, 2018). Most of the fish comes from the Mediterranean Sea (Tunisia, Italy and Israel) and a small portion from the Atlantic Ocean (Portugal, Spain and Neherlands). Regarding the, this species was registered only in Italy and in Turkey.

1.2.1.6. *Pagellus bogaraveo* (Blackspot=red sea bream)

In 2016 the total production of Pezzogna[1] was 1,539 tonnes (1,341 tonnes from fishing and 178 tonnes from aquaculture) mainly fished in the North-East Atlantic (Spain and Portugal) and a small part in the Mediterranean and Central-Eastern Atlantic. Portugal (627 tonnes) and Spain (206 tonnes) are the main producers. As for aquaculture in the decade 2006-2016 the Pezzogna was farmed only in Spain and Portugal, while in 2016 only in Spain (172 tonnes) (FAO FishStatPlus, 2018).

1.2.1.7. *Pagellus erythrinus* (Common pandora)

Between 2006 and 2016, the production of Fragolino[1] deriving from fishing ranges from 9.347 to 13.901 tonnes (FAO FishStatPlus, 2018), while the aquaculture production has progressively decreased from 197 tonnes in 2006 to 0,04 in 2016. The countries where this species is coming from are: Egypt, Italy, Croatia, Algeria, Slovenia, Morocco, Libya, Lebanon, Malta, Cyprus, Israel, France, Spain, Tunisia. Most of the wild Fragolino comes from the Mediterranean Sea (12.254 tonnes), mainly from Italy, Libya, Spain and Tunisia, while a small share comes from Atlantic countries (1.647 tonnes), especially Morocco. Regarding aquaculture, there was a small production in some fishfarm in Greece (127,27 tonnes in 2014) and in Cyprus (0.23 tonnes in 2016).

1.2.1.8. *Pagrus* (Red porgy)

In 2016 the total catches of Pagro[1] were 11,733.2 tonnes of which 3.266 were farmed and 8.467 were caught (FAO FishStatPlus, 2018). This species is fished mainly in the Western Atlantic (Argentina and Brazil) and in the Mediterranean Sea (Egypt, Greece, Turkey). The Pagro is considered the most suitable species to be farmed in the Mediterranean

countries, in fact the major (and only) world producers, are Greece and Turkey.

1.2.1.9. *Sparus aurata* (Gilthead sea bream)

The first FAO statistics regarding the production of Orata[1], are related to Italy in 1970, with a production of 10 tonnes. Since then, production has grown rapidly and in 2016 statistics included 16 countries: Albania, Algeria, Bosnia and Herzegovina, Croatia, Cyprus, Dominican Republic, France, Greece, Israel, Italy, Libya, Malta, Montenegro, Morocco, Palestine, Portugal, Slovenia, Spain, Tunisia, Turkey, Yugoslavia SFR. The production systems of Sea bream are varied and range from extensive production (valliculture in Italy and production in the lagoons in Egypt) or semi-intensive in the lagoons (Portugal and southern Spain) to highly intensive farming systems on land, on the coasts (Greece and Turkey) and in open sea cages (Cyprus, Italy and Spain). The total catches in 2016 of Sea bream were 195.410,42 tonnes, of which 185.980,42 from aquaculture and 9.430 deriving from fishing (FAO-FishStatPlus, 2018). Production took place mainly in the Mediterranean Sea (156.545,42 tonnes in 2016), although a small production was also present in the Atlantic Ocean (4.132,97 tonnes Portugal, Spain and France in 2016). Between the major producer countries are Turkey (37,21%), followed by Greece (31,32%), Tunisia (7,7%) and Spain (6,46%).

1.2.2. Aquaculture of non-EU species

- #### 1.2.2.1. *Acanthopagrus schlegelii* (Blackhead sea bream)

The catches of the *A. schlegelii*[2] have progressively increased between 2006 (3.858 tonnes) and 2016 (1.772,99 tonnes) (FAO-FishStatPlus, 2018). About aquaculture, FAO data mention Japan, Korea and Taiwan. In 2016 the quantity produced was 1.544,99 tonnes, with Taiwan being the main producer (FAO-FishStatPlus, 2018). The *A. schlegelii* is farmed in

1 denomination in Italian language according to D.M. MIPAAF, 22 Settembre 2017.

2 does not have a denomination in Italian language according to D.M. MIPAAF, 22 Settembre 2017.

lakes with brackish water. Data on fisheries, on the other hand, concern only Taiwan, which in 2016 produced 228 tonnes.

- **1.2.2.2. *Pagrus auratus* (Silver seabream)**

The Pagro indopacifico[1] is very popular in Australia. FAO data on total production in 2016 are 95.127 tonnes divided into 72.286 tonnes for aquaculture and 22.841 tonnes for fishing. Among the major producers we find: Australia, Japan, Taiwan and New Zeland. Even though the largest aquaculture productions took place in Japan and Korea (FAO-FishStatPlus, 2018).

- **1.2.2.3. *Pagrus major* (Japanese seabream)**

The largest amount of Pagro maggiore[1] is fished in the Northern Pacific. As for aquaculture, this species was the first of the *Sparidae* family to be farmed intensively. The first data on production go back to 1957 in Japan (FAO-FishStatPlus, 2018). The production has grown rapidly in the 1970s and, within a decade, the stock quantity surpassed the amount fished in the Northern Pacific. The *Pagrus major* farming continued to increase in the 80s and 90s, but then decreased at the beginning of 2000. The latest data go back to 2011 (Taiwan - 209.3 tonnes) and 2012 (Cypro - 2 tonnes) when there was only acquaculture production.

- **1.2.2.4. *Sparidentex hasta* (Sobaity sea bream)**

Sparidentex hasta[2] is the only species of the genera Sparidentex. Data on fishing for this species have been reported since 2000 in the Persian Gulf. The quantity fished is small, around 102 tonnes (FAO-FishStatPlus, 2018). In 2016 about 4 tonnes were obtained from aquaculture and the only producers are Kuwait and Bhrain.

1.3. The *Sparidae* family

Classe: Osteichtyeis

Subclass: Actinopterygii

Superorder: Teleostei

Order: Perciformes

Suborder: Percoidei

Family: *Sparidae*

1.3.1. Systematics

The *Sparidae* family belongs to the order of the Perciforms that is composed by 20 suborders, 160 families, about 1.540 genera and over 10.000 species. About 52 families have a single genus, 23 a single species and 21 have 100 or more species (Nelson, 2006). The suborder of Percoidei is the largest of the 20 Perciformes suborder and includes 79 families, including the *Sparidae* family, 540 genera and about 3.176 species (Nelson, 2006).

The *Sparidae* family includes about 115 species divided in 33 genera (Nelson, 2006) although, according to Fishbase, the species are 159 and the genera 38 (http://www.fishbase.org/Nomenclature/FamilySearchList.php?). Based on FAO ASFISS list (FAO. 2018a.) 107 species of *Sparidae* are globally commercialized with different trade designations, of these 41 species are commercialized in Italy (D.M. MIPAAF, 31 January 2008; Italian Republic, 2008).

1.3.2. Classification

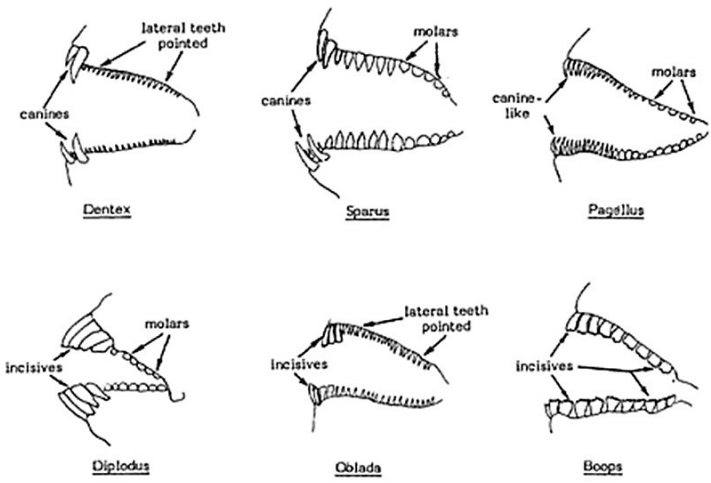
The sparids classification is mostly based on morphology: the teeths, the number of spiny and soft rays of the dorsal fin, form and livery (Bauchot & Hureau, 1986). In particular, the teeth have been typically used by many authors to distinguish the genres (Jordan & Fesler, 1893; Munro, 1949; Akazaki, 1962) and to outline a classification in subfamilies (Figure 5). In fact, Smith (1938) grouped the 33 genera into four subfamilies that can be distinguished by their different teeth, related to the very variable diet of the species:

- *Boopsinae*: mainly herbivores, sometimes predating small invertebrates, they have compressed incisiviform teeth;
- *Denticinae*: they have a piscivorous diet, so they are provided with large canines in the external part of the mouth and small conical teeth on the inside;
- *Pagellinae*: they are carnivorous and eat preferably small invertebrates. They haven't got canines but, externally they have got conical teeth while internally they have got molars, both of small size;
- *Sparinae*: they are carnivorous and prey on crustaceans, molluscs and small fishes, due to strongly rounded molars, located in the back part of the jaw, replaced in the external part by wider teeth.

Later Akazaki (1962) identified two further new subfamilies:

- *Diplodinae:* they have got 6-8 front teeth and obliquely oriented incisors;
- *Pagrinae:* they have got two series of molars, 4 canines on the upper jaw and 4- 6 canines on the lower jaw. The cranium is also provided with flakes extending into the interorbital region.

Figure 5. Subfamilies classification based on dentition. Source by Fao, 1981.



The two new subfamilies described by Akazaki (1962) were not considered valid by Smith (1986). Finally, in 1991 Fiedler (1991) carried out a further revision of the family, recognizing as valid only three subfamilies: *Sparinae*, *Denticinae* and *Pagellinae*.

1.3.3 Morphology

The sparids have got a mostly elongated body, somehow compressed laterally. They have a silvery colour, characterized by reflections ranging from pink, yellow to gray and darker longitudinal and transverse stains and bands. The maximum length recorded is 2 meters (*Petrus rupestris*). Ctenoid scales, large and rough, are present throughout the body except for the area of the mouth and the head; the operculum and the preoperculum are characterized by scales but do not show depressions or spikes (Carpenter & Niem, 2001). The mouth, small and arranged horizontally or ventrally, is extended forward and has a strong dentition with evident heterodontids: incisiviform, caniniform, molariform and villiform teeth's (Tortonese, 1975). The dorsal fin is unique and elongated, generally composed of 10-13 spiny rays in the front half and 10-15 soft rays in the back; the anal fin has three spiny rays and 8-14 soft rays; the pectoral fins are long and pointed; the caudal fin is biloba.

1.3.4. Phylogenetic position

The studies, realized with the use of molecular markers, have highlighted that the phylogenetic relationships of the *Sparidae* family and subfamilies need a revision. Orrel (2002), Orrel & Carpenter (2004) and Chiba (2009), used different type of molecular markers, they have demonstrated that the *Sparidae* family is not monophyletic, unless it includes the genera *Spicara*, currently attributed to the family *Centracanthidae*, and that nobody of the subfamilies classified as such appears monophyletic. This important similarity has already been highlighted by other authors (Regan, 1913; Smith & Smith, 1986; Johnson, 1993) who reported a morphological comparison of the maxillary and premaxillary joints. Many others morphological studies (Day, 2000; Day, 2002; Hanel & Strumbauer, 2000) pointed that the success of the evolutionary radiation of sparids, more than other closely related families of teleosts, appears to be linked to the high adaptive plasticity of mouthparts and teeth, which allowed a specialization in food strategies and predation Consequently, the morphological similarities could be more related to independent processes of evolutionary

convergence (Duftner et al., 2007) and reflect the trophic ecology of the species (Antoniucci et al., 2009), which does not reproduce phylogenetic closeness. For this reason, the dental characteristics alone can not be decisive in delineating a precise classification.

1.3.4. Reproduction

Within the *Sparidae* family, the individual species vary considerably with regards to their reproductive strategy. Hermaphroditism is frequent, with proteogenic species (eg. *Tanuta*) that are females in the first part of their life, and proterandric species (*Bream, Saragus, Ramblack, Mormora, Salpa*) that are males in the first part of their life, with interest in some species of the entire population while in others only a few individual are involved. There are also gonochoric species (*Dentex*), that is, of separate sexes. Additionally, in some species an immature intersexual gonad is developed (rudimentary hermaphroditism). Buxton & Garrat consider these species as "late gonochorics", referring to the period of growth in which there are separate sexes to distinguish them from "true gonochoric" species where no hermaphrodite tissue is present at any stage of development. Lastly, in *Pagellus bogaraveo* a simultaneous hermaphroditism was identified, as to say the presence of male and female gonads (Williamson, 1911, Le Gall, 1929; Buxton & Garrat, 1990). The causes on the evolution of hermaphroditism as a reproductive style could be to maximize reproductive potential (Williams, 2018) and thus their individual opportunities. In Sparids, and in fish in general, progressive hermaphroditism (in which an individual reproduces first in one sex and then in the other) is considered the most common form of hermaphroditism (Choat & Robertson, 1975; Shapiro, 1981; Buxton & Garrat, 1990). However, sequential hermaphroditism is considered beneficial when an individual reproduces in one sex at a moment in life and in the other sex in another. (Ghiselin, 1969). As a rule, the spawning period must meet the physiological needs of the offspring. The spawning period is influenced, in addition to the photoperiod and the termoperiod, by a variety of environmental parameters, such as predation, competition, and food availability.

1.3.5. Habitat and geographic distribution

The *Sparidae* family is characterized for the most part by marine fishes, only a few species, such as *Archosargus probatocephalus* live in fresh or salty water (Nelson, 2006). They are demersal and mainly coastal, distributed in all oceans, in both tropical and temperate waters, at depths between 0 and 250 meters (Bargelloni et al., 2003). Occasionally they can drive into the lagoons or form nurseries (Carpenter & Niem, 2001), such as in the Mediterranean Sea, where some species conduct themselves like cyclical migratory, entering the lagoons after metamorphosis and living in this environment for the first stages of the life cycle and then return offshore the following year (October-November). The forces that regulate these migrations could be intrinsic (reproduction) or linked to variations of abiotic factors, such as fluctuations in temperature and salinity. Frequently in juvenile stages they are gregarious, but adults are more commonly solitary and may even go deeper than the platform and the continental slope (Carpenter & Niem, 2001).

Sparids are very common in the Mediterranean Sea and often constitute most of the coastal fish fauna. Sparids are widespread, their maximum concentration is along the Southern and Western coasts of Africa. With over 44 species, the Cape region in South Africa contains the highest concentration of *Sparidae*. They are the only fishes of the *Sparoidea* superfamily to be present in the western Atlantic Ocean (Hudson, 1998). To the north-west, in the cold waters of the Benguela stream, off Namibia, the number of *Sparidae* species is drastically reduced and then increased again in the warmer waters of Angola and the Gulf of Guinea. To the north-east of the Cape region, the number of *Sparidae* species decreases in the warm waters of the Aghulas stream up to only 12 species in the Red Sea. With relation to the distribution of sparids, the Western Pacific Ocean is less important. Only 21 species were found from Northern Japan to Southern Australia and Southern New Caledonia. While the sparids are completely missing in the central Pacific, there are two other species of *Sparidae* (*Archesargus pourtalesii* and *Calamus brachisomus*) in the Eastern Pacific.

1.4. Seafood frauds

1.4.1. Universal notion of food frauds.

Food adulteration is not a modern phenomenon and, probably, it is as old as the food processing. In the 'modern' scientific era, the first to report this problem was the analytical chemist Frederick Accum (Accum, 1820) who did a treatise on adulteration of food and culinary poisons, published in 1820. The sophistication of methods used to commit fraud, on the contrary, constantly changes, according to the progress of production systems. Currently, there is no legislative or definite definition of food fraud. A complete definition generally used by different bodies has been defined by the Food Fraud Initiative of Michigan University *"Food Fraud: a collective term encompassing the deliberate and intentional substitution, addition, tampering or misrepresentation of food, food ingredients or food packaging, labeling, product information or false or misleading statements made about a product for economic gain that could impact consumer health"* (Spink & Moyer, 2011).

Food fraud is mostly done for economic reasons, so that the US Food and Drug Administration (FDA) added to "Food Fraud" the sub-category of the "Economically Motivated Adulteration" (EMA), defined as the *"fraudulent, intentional substitution or addition of a substance in a product for the purpose of increasing the apparent value of the product or reducing the cost of its production"*. If food fraud is frequently inspired by economic motivation, its impact may lead to a tangible risk for public health. In fact, any tipes of fraud result in a modification of the identity and/or purity of the original ingredient by substituting, diluting, or modifying it by physical or chemical means and, as the fraudulent ingredient is unknown, such a manipulation may pose toxicological or hygienic risks or allergen risks to the consumer (Moore et al., 2012). To date, the estimate costs of food frauds to the world economy are around $49 billion for year (US) (NSF international, 2013). Distinct types of food fraud have been recognized, as described in Table 2.

Table 2. Different types of food fraud (Spink, 2007; Spink, 2009b).

Tipe of food fraud	Definition	Example	Potential public health
Adulteration /dilution/ replacement	A component of the finished product is fraudulent	Substitution with an alternate ingredient this can include misrepresentation of geographic, botanical, varietal, or animal origin, or the additional of a non-food grade substance. A component of the finished product is fraudolent.	Fraudulent component
Tampering/alteration	Legitimate product and packaging are used in a fraudulent way	Change the chemical-physical and/or organoleptic characteristic of a product, due to spontaneous degenerative processes or to poor management or conservation e.g. changed expiry information.	Fraudulent packaging information
Over-run	Legitimate product is made in excess of production agreements	Under-reporting of production	Fraudulent product is distributed outside of regulated or controlled supply chain
Theft	Legitimate product is stolen and passed off as legitimately procured	Stolen products are co-mingled with legitimate products.	Fraudulent product is distributed outside of regulated or controlled supply chain
Diversion	The sale or distribution of legitimate products outside of intended markets	Relief food redirected to markets where aid is not required	Shortages or delays of relief food to needy populations
Simulation	Illegitimate product is designed to look like but not exactly copy the legitimate product	"Knock-offs" of popular foods not produced with same food safety assurances	Fraudulent product of lesser quality
Counterfeit	All aspects of the fraudulent product and packaging are fully replicated	Copies of popular foods not produced with same food safety assurances	Fraudulent product

1.4.2. Tipe of seafood frauds in the world

In the draft report "*The food crisis, fraud in the food chain and the control thereof* "of EU Parliament, has reported that the products which face more risk include fish, olive oil and organic foods.

In the February 2009 U.S. Government Accountability Office (GAO) has done a report on seafood fraud (Shames, 2009). Nevertheless, this report was intended to focus on economic risks, and only briefly cited possible public health risks (a toxic pufferfish incident, in which it was mislabeled as monkfish, and allergen risks due to mislabeling) The study of Spink & Moyer (Spink & Moyer, 2011) builds on the GAO Seafood Fraud Report (Shames, 2009) by exploring more of the potential EMA, food-associated public health risks, and contains additional comments in the right hand column (Table 3).

Table 3. GAO seafood fraud report: types of food fraud and potential public health risk (Shames, 2009).

GAO seafood fraud report detail			New comment potential public health food risk
Fraud type	**Description**	**Cause and motivation**	
Transshipment	Transferring cargo among different transports and countries	Avoid tariffs or anti-dumping duties	Compromised storage, handling, and traceability (in the event of a recall)
Over-treating	Adding more ice or water than allowed by regulation	Increase profits by including more weight for ice than fish	Water may include pathogens or chemicals (for example, if ice was made from pond water)
Species substitution	Substituting less costly species and misrepresenting them as more expensive species	Increase profits due to cost differential	Misrepresented species may be toxic or cause allergic reactions
Short-weighting	Package labels state weights higher than packaged contents	Increase profits due to weight differential	None
Other mislabeling or misrepresentation	Misrepresenting country of origin, ingredients, so on	Generally, avoid costs and maximize profits	Undeclared allergens, toxins from banned locals (for example, ciguatoxin-prone reefs), weight increased added through other unknown materials, so on

Seafood fraud can occur during any sector of the supply chain, from the boat or farm to the trade distribution, and in different forms, including species substitution, improper labelling, plus hiding out the true origin of seafood products, or adding extra breading, water or glazing to seafood products to increase their apparent weight (Warner et al., 2016; Upton, 2015). Among them, existing reports indicate that the substitution with different species, particularly the case of expensive species replaced with cheaper alternatives but also the occurrence of threatened species retailed as more sustainable, are currently the most conspicuous types of fraud (Warner et al., 2016; Xiong, et al., 2016b; Johnson, 2014). Occasionally species substitution may occur involuntarily and not encompass economic gains; for instance, errors may occur during fishing operations when catch is erroneously labelled on board or at landings, generally due to morphological similarity of species in mixed fisheries, or when a big volume of catch is managed rapidly for transmitting in a short time (Garcia-Vazquez et al., 2012). Though fishers may mislabel seafood, especially if they caught it illegally, the literature proposes that mislabelling is most often done by distributors and the final seafood retailer (e.g., fishmongers and restaurants) in order to improve profits (Pardo et al., 2016). Particularly, deliberate mislabelling is simpler when seafood is sold processed. Substitution rates are enormously depending on

the species, since some are often sold mislabelled, though others infrequently or never (Bénard-Capelle et al., 2015). In some cases, the grade of mislabelling is also dependent on the time of the year in which some species are purchased (Warner et al., 2015). The fraudulent activity of mislabelling obviously generates consumer economic loss. Nevertheless, it can occasionally create a hazard for consumers if an innocuous fish is substituted with a species associated with some types of food poisoning or exposure to certain allergens (Johnson, 2014). For this reason, in the last years seafood mislabelling has been internationally recognized as a significant issue and media attention has constantly focused on this topic, highlighting the alarming number of reported cases. Aquaculture has been playing an increasing role in seafood fraud; in fact, consumers across the world may eat several popular farmed fish without even knowing it, since they are frequently camouflaged as wild-caught, higher-value fish (Warner et al., 2016). Oceana, a non-profit international organization focused on oceans conservation, carried out a broad investigation on seafood mislabelling worldwide, analysing more than 200 studies published since 2005, to identify general trends in seafood fraud (https://usa.oceana.org/publications/reports/deceptive-dishes-seafood-swaps-found-worldwide). This investigation covered 55 countries with a total of 25.700 samples of seafood analysed for mislabelling (US and EU accounted for three-quarters of the studies and cases). About 65% of the studies reviewed contain strong evidence of economically motivated adulteration of seafood products. The 16% of the species identified as substitutes are considered to have some level of elevated conservation risk (either threatened or close to becoming threatened with extinction in the near future) by the IUCN. Most of them (nearly 12% of all the species substituted) are considered critically threatened, threatened or vulnerable. An alarming percentage of 58% of the samples identified as substitute species, demonstrated a species-specific health risk to consumers including parasites, environmental chemicals and aquaculture drugs, and other natural toxins such as histamine, ciguatera, gempylotoxin or tetrodotoxin. To date, only US seafood fraud seems to represent approximately 15% (40% for some species) of the total analysed seafood products. Regarding EU, a concrete diminution of cases has been observed during the last few years and, after the application of legal provisions pointed at preventing illegal fishing (IUU) and improving transparency and accountability in the seafood supply chain, seafood fraud rates have apparently begun to decrease since 2011, primarily, in those EU countries

where the rules are enforced, and for those products covered by the legal provisions (Warner et al., 2016). The preliminary data of the EU suggest that catch documentation, traceability and consumer labeling is feasible and effective for fighting seafood fraud.

1.4.4. Sparids and frauds

The family of sparids includes numerous genera and contains numerous species that are commercially valuable for the quality of their meat, but also species of inferior value that can be easily confused. This confusion originates because, within the various genera, the species are very similar to each other or even like species of different families. For this reason, within the *Sparidae* family, different substitution scenarios can occur:1) concerning species belonging to the same genera, 2) relates to similar species but belonging to different genera; 3) is linked to the possibility of confusing very similar species but belonging to different families. Fraud for substitution of similar species is very frequent because the demand in the European market of precious species is so high that it is addressed to species imported from non-EU countries, from the Atlantic and Indo-Pacific Oceans to encounter domestic needs. Fortunately, the manifest heterodontia (Figure 5) that characterizes the different genera of the sparids can be used for the identification of those species that can be easily confused or deliberately replaced. In general, it is easy (for skilled operators) to unmask substitution fraud between species of a different genera than those that imply the use of species belonging to the same genera. It is even easier to identify the substitution frauds of species belonging to different families.

The unmask of the fraud is possible through the macroscopic examination of the dental table for the species belonging to the genera *Dentex*, especially for the *Dentex macrophthalmus* and *Dentex angolensis*, from the species belonging to the genera *Pagellus, Pagrus, Diplodus, Spondyliosoma*, while to recognize with certainty the other species belonging to the generas *Boops, Sarpa* it is necessary to use the microscope because the differences in their teeth are imperceptible to the naked eye.

1.4.4.1 Replacement among species belonging to the same genera

Dentex: it is possible to differentiate the *Dentex dentex* (Dentice) from the other species belonging to the same genera because it has got caniniform

teeth that are quite linear and quite high compared to the *D. barnardi* and *D. canariensis* where they are much robust and of bigger diameter. The *Dentex macrophthalmus* is characterized by tinny, protruding and slightly arched caniniforms, while the teeth of the *Dentex gibbosus* are thinner even if they are less delicate than those of the *Dentex dentex* (Malandra & Renon, 1998; Manzoni, 1993) (Table 4).

Pagrus: even if the most important and frequent species on the Italian market are all named "Pagro" (*P. pagrus, P. caeruleostictus, P. auriga*), it may be useful, for its merit, to recognize with certainty the *Pagrus auriga*. In fact, in this species the conical teeth (also called premolars) are always present in lower numbers (1-2 for emiarcata) compared to the other porgy. In addition, the tip of all canine and conical teeth is dark gray, while in the others it is always clear (Malandra & Renon, 1998; Manzoni, 1993) (Table 5).

Pagellus: differentiating the *Pagellus erythrinus* from the *Pagellus bellottii* is possible because the first has the upper developed conical teeth and the lower ones are almost V-shaped. Through a dorsal view of the mandibular profile, it can be noted that the mandible is bilobed and on each side symmetrically protrudes a canine laterally. In *P. bellottii* the canines, always equidistant from each other, are slimmer almost transparent and through an observation under the microscope, you can highlight their typical hook shape (especially the lower ones) (Malandra & Renon, 1998; Manzoni, 1993). (Table 6).

Table 4. Species belonging to the Dentex genera of commercial importance in Italy (DM n. 19105 del 22 Settembre 2017). Revised from FAO webpage (http://www.FAO.org/fishery/species/search/en).

Specie	Eye	Teeth	Gill rakes	Fins					Colour
D. angolensis (Dentice atlantico) 	Eye diameter greater than of the interorbital space and of the suborbital space which is 1/5 or 1/6 of the head	Several rows of canine-like teeth, outer row the strongest with 4 to 6 better developed anterior teeth in each jaw, the uppers visible when mouth is closed.	First arch 9 or 10 lower and 6 to 9 uppers.	DXII 9/10(increasing in length up to the 4°or 5° and subequal thereafter longest spine 44 to 51% of head length).	AIII 8/9	C	P	VI 5	Colour red with silvery reflections, head darker and belly lighter; a small dark area above the insertions of pectoral fins; dorsal and anal fins red except on their bases; pelvic fins light-coloured; pectoral fins and caudal fin reddish.
D. barnardi (Dentice atlantico) 	**Eye diameter greater than of the interorbital space.**	**Several rows of canine-like teeth, outer row the strongest with 4 to 6 very well-developed anterior teeth in each jaw.**	**First arch 14 to 16 lower and 8 to 11 uppers**	**DXII 9/10 (first two very short; those following 3°-4°-5° very long and filamentous)**	**AIII 8**	**C.**	**P 15/16**	**V I 5**	**Colour reddish with silvery reflections. A large red-brownish spot on the posterior end of the base of the dorsal fin, and a spot equal to the base of the pectoral fins. Soft portion of the dorsal fin with dark spots more or less aligned. reddish tail with black hem.**
D. canariensis (Dentice atlantico) 	Eye diameter smaller than width of suborbital space in large individual	Several rows of canine-like teeth, outer row the strongest with 4 to 6 very well-developed anterior teeth in each jaw.	First arch 10 to 13 lower and 6 to 9 uppers	DXII 9/10 (first two very short; those following 3°-4°-5° very long and filamentous)	AIII 8/9	C.	P	V I 5	Colour reddish with silvery reflections. Belly lighter and head darker. A dark red spot posteriorly on base of dorsal fin extending well beyond scaly sheath. A dark area at pectoral fin axils; aligned dark spots on soft portion of dorsal fin. Caudal fin dark red, very finely edged with black. In some individuals, a greenish yellow band between eyes.
D. dentex (Dentice) 	**Eye small, suborbital space wide.**	**Several rows of canine-like teeth, outer row by far the strongest with 4 to 6 very well-developed anterior teeth in each jaw.**	**First arch 9 or 10 lower and 8 or 9 uppers.**	**DXI-11/12 (increasing in length from the 1° to the 4° or 5°and subequal thereafter);**	**AIII 7/9**	**C 18**	**P 14/15**	**VI 5**	**In the young specimens greyish, spotted with black on back and upper sides, becoming pink with sexual maturity; old individuals are bluish grey, and the dark spots become diffuse with age.**
D. gibbosus (Dentice gibboso) 	Eye diameter about equal to width of suborbital space	Several rows of canine-like teeth, outer row much the strongest with 4 to 6 better developed anterior teeth in each jaw.	First arch 8 to 10 lower and 6 to 8 uppers. .	DXII 10/11 (first two very short; those following very long and filamentous)	AIII 7/9	C 17	P 14	VI 5	Colour reddish with bluish silvery reflections; belly lighter and head darker. A small black spot behind posterior end of dorsal fin; a brownish black spot at pectoral fin axils; a dark area at upper angle of opercle; 1 or 2 dark lines on soft part of dorsal fin. Caudal fin red, edged with black. Large individuals are often tinged wine red and spotted with black on head (males) or greyish (females).
D. macrophtalmus (Dentice occhiuto) 	**Eye very large, its diameter greater than snout length; suborbital space slight.**	**Numerous rows of canine-like teeth, outer row the strongest, with 4 well developed anterior teeth (in upper jaw, visible when mouth is closed) and 10 small anterior teeth (clearly smaller than the upper canines) in lower jaw.**	**First arch 17 to 20 lower and 9 to 12 uppers.**	**DXII 10/11 (increase in length from 1° to 4°/5° then decrease)**	**AIII 8**	**C 9-12-9**	**P 16/17**	**VI 5**	**Colour of body and fins reddish, lateral line more bright red; spinous portion of dorsal fin whitish at base. Anal fin edged with white; inferior margin of lower caudal fin lobe white. The coloration becomes more intense during the spawning season.**
D. marocccanus (Dentice rosa) 	Eye diameter greater than of the interorbital space	Several rows of canine-like teeth, outer row much the strongest with 4 to 6 very well-developed anterior teeth in each jaw, the uppers visible when mouth is closed.	First arch 9 to 12 lower and 7 to 9 uppers.	DXII 10/11	AIII 8/9	C	P	VI 5	Colour light red with silvery reflections; head darker and fins pinkish; distal part of dorsal and anal fins more intensely reddish; fork of caudal fin edged with dark red; a very small dark spot above pectoral fin insertion. Males display a more intense coloration

Table 5. Species belonging to the Pagrus genera of commercial importance in Italy (DM n. 19105 del 22 Settembre 2017). Revised from Fao webpage (http://www.FAO.org/fishery/species/search/en).

Specie	Eye	Teeth	Gill rakes	Fins					Colour
P. africanus (Pagro atlantico) 		Jaws provided with large caniniform teeth, 4 upper and 6 lower ones, followed by various series of teeth smaller and blunt conical that become molariforms.		DXII 10/11(first two very developed).	AIII 8/9	C	P	VI 5	Rose with silvery reflections, sometimes with blue specks on the sides. Presence of a reddish-brown stain at the armpit of the pectoral fins above the base of the fins; pinkish uneven fins with orange margin.
P. auratus (Pagro indopacifico) 		**Jaws provided with large caniniform teeth followed by various series of molariforms.**		**D (first two very short; those following)**	**A**	**C**	**P**	**V I 5**	**Colour from golden rose to red, to reddish bronze, with blue specks on the upper part of the hips, margin of the blackish caudal fin; tail with inferior lobe with white margin in the distal third.**
P. auriga (Pagro) 		Anterior teeth canine-like, 4 in upper and 6 in lower jaw, followed by blunter teeth that become progressively molar-like and are arranged in 2 or 3 rows. Behind the row of large canine-like teeth there are some smaller teeth	Gill rakers on first arch 10 or 11 lower and 6 to 8 uppers.	DXI 11/12 (first two very short; those following 3°-4°-5° very long and filamentous),	AIII 8/9	C 17.	P 15/ 16	V I 5	Colour pink with silvery reflections and 4 or 5 dark red alternatingly broad and narrow cross bars. Adults of a more intense wine red with the cross bars less well visible than in the young. Head dark between nape and corner of mouth; hind edge of opercle very dark. Dorsal fin pink with some black on the membranes separating the filamentous spines and with orange on distal parts of soft rays. Anal fin similar in colour to the dorsal. Pectoral fins pinkish orange. Pelvic fins wine red edged with black. Caudal fin greyish at base, pink or orange edged with black distally.
P. caeruleostictus (Pagro) 		**Anterior teeth canine-like, 4 in upper and 6 in lower jaw, followed by blunter teeth that become progressively molar-like and are arranged in 2 or 3 rows. Behind the row of large canine-like there are some smaller teeth.**	**First arch 10 to 13 lower and 6 or 7 uppers.**	**D XI/XII 9/12 (the first 2 spines always very short, the 3° to 5° longest, filamentous in the young).**	**AIII 8/9**	**C 16**	**P 15**	**VI 5**	**Colour pink with silvery reflections and large bluish black spots on back and sides. Head darker, particularly on the interorbital space. A dark spot at bases of last dorsal soft rays extending on to the sheath of the fin but becoming lighter with age. Caudal fin pinkish, the fork edged with black; other fins bluish or pinkish. Old individuals very often with numerous irregular dark spots on head and back; old males with yellow on the head during the reproduction season.**
P. major (Pagro maggiore) 		Anterior teeth canine-like, followed by blunter teeth that become progressively molar-like.		DXII 10/12 (first two very developed)	AIII 7/9	C	P 14	VI 5	Colour reddish with bluish silvery reflections.
P. Pagrus (Pagro) 		**Both jaws anteriorly with large canine-like teeth, 4 in upper and 6 in lower jaw, followed by smaller and blunter canine-like teeth that become progressively molar-like toward the posterior third of jaws; the 2 outer rows of strong teeth are flanked in the region anterior to the molars by several rows of very small teeth.**	**Gill rakers short, lower limb of first arch with 8 to 10, 6 to 8 on upper limb.**	**DXII 9/12**	**AIII 8/9**	**C 23/ 27**	**P 15**	**VI 5**	**Colour: pink with silvery reflections, lighter on belly; head dark from nape to angle of mouth; sometimes, fine blue dots present on upper sides, particularly conspicuous in young individuals; often a somewhat darker area at pectoral fin axils; caudal fin dark pink, with both tips white; other fins pinkish**

Table 6. Species belonging to the Pagellus genera of commercial importance in Italy (DM n. 19105 del 22 Settembre 2017). Revised from Fao webpage (http://www.FAO.org/fishery/species/search/en).

Specie	Eye	Teeth	Gill rakes	Fins					Colour
P. acarne (Pagello) 		Both jaws with pointed teeth anteriorly and molar-like ones posteriorly. An inner band of numerous slightly smaller, cardiform teeth behind the outer row of pointed teeth. Molars arranged in 2 rows		DXII/XIII 10/12	AIII 9/10	C 17	P 15/16	VI 5	Brown-reddish colour on the back, instead the belly silver; Black spot above the pectoral fin base that extends to the top of the base of the fins; buccal cavity and gills of colour yellow-orange or red salmon; fins pink.
P. affinis (Pagello indiano) 				**D (first two very short; those following)**	**A**	**C**	**P**	**VI 5**	
P. bellotti (Pagello atlantico) 		Both jaws with pointed teeth anteriorly and molar-like ones posteriorly. An inner band of numerous slightly smaller, cardiform teeth behind the outer row of pointed teeth. Molars arranged in 2 rows.	Gill rakers on first arch 9 or 10 lower and 5 or 6 uppers.	DXII 9/11 ,	AIII 10	C	P	V I 5	Colour more less bright red with silvery reflections; often blue spots following scale rows on sides; interorbital space darker; a small, dark red spot at origin of lateral line and along upper margin of opercle; base of pectoral fin darker; fins pinkish yellow (in many specimens from the Gulf of Guinea) or greyish. Caudal fin often margined with red or orange; inside of mouth whitish. The red vertical bars described by authors may correspond to a fright pattern.
P. bogaraveo (Pagello pezzogna o Pezzogna) 	**Snout shorter than the eye diameter**	**Both jaws with pointed teeth anteriorly and molar-like ones posteriorly. An inner band of numerous slightly smaller, cardiform teeth behind the outer row of pointed teeth. Molars arranged in 2 rows**		**D XIII/XIII 11/14**	**AIII 11/13**	**C 17**	**P 16/17**	**VI 5**	**Colour: pink with silvery reflections; Black spot above the pectoral fin base. And another at the origin of the side line; buccal cavity and gills of colour yellow-orange or red salmon; fins pink.**
P. erythrinus (Pagello fragolino o fragolino) 	The eyes are smaller than those of *P. bogaraveo* and *P. acarne*.	Both jaws with pointed teeth anteriorly and molar-like ones posteriorly. An inner band of numerous slightly smaller, cardiform teeth behind the outer row of pointed teeth. Molars arranged in 2 -3 rows superiorly and 2 rows inferiorly.		DXII 10/11	AIII 8/9	C 17	P 15	VI 5	Colour is silver in colour with a pink tinge, upper the edge of operculum is dark red; a red spot particularly above the pectoral fin base; buccal cavity and gills of colour dark brown.

1.4.4.2 Substitution between species belonging to different genera

Another possible fraud can occur with the sale of species belonging to different genera. For example:

- Pagro sold for *D. angolensis, D. barnardi, D. canariensis* as they are both reddish colored;
- Pagro sold for Pagello fragolino (*P. erythrinus*);

The monospecific genera such as *Sparus*, *Oblada*, *Sarpa*, *Spondyliosoma*, do not pose doubts to their precise identification, thanks to their different dental conformation (Malandra & Renon, 1998).

1.4.4.3 Replacement among species belonging to different families

As for the substitution between species belonging to different families *Dentex, Pagellus or Pagrus*, being pink-red in color of the livery, they are liable to be fraudulently substituted with some species of *Lutianidae* (*Lutjanus bohar, L. sebae, L. malabaricus*) belonging to Indian and Pacific Oceans (Malandra & Renon, 1998). The substitution can also occur between sparids with silver-gray in color of the livery (snappers and porgy) and *Letrinidae* (*Lethrinus atlanticus*) (Malandra & Renon, 1998) (Figure 6). Also, in this case the analysis of the dental table is indispensable, since the families *Sparidae, Luthianidae* and *Lethrinidae* have in common the following characters: ventral fins in thoracic position, a single dorsal fin along the entire dorsal margin, ovoidal shape of the body, caudal bilobata fin (Carpenter, 1989) (Allen, 1985) (Figure 7 and 8).

Figure 6. Species of family Lethrinus and Lutjanus suitable to replace species belonging to Sparidae family

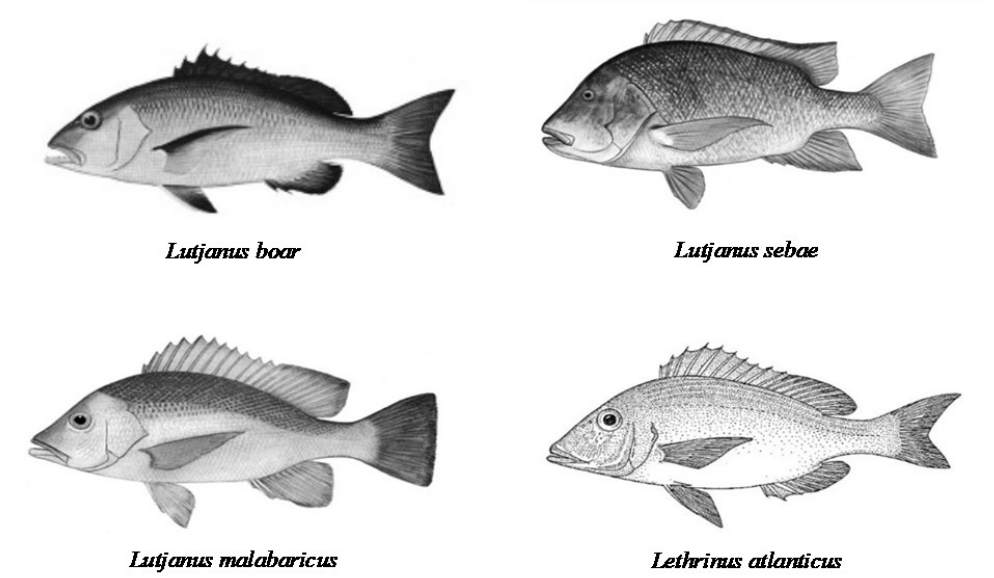


Figure 7. Morphological features of Lethrinidae and Sparidae. Source by Carpenter, 1989.

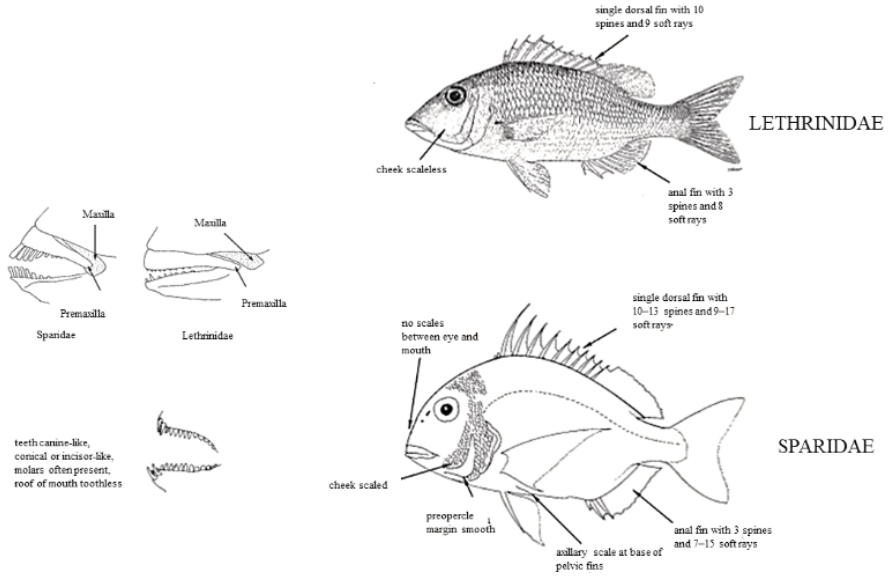
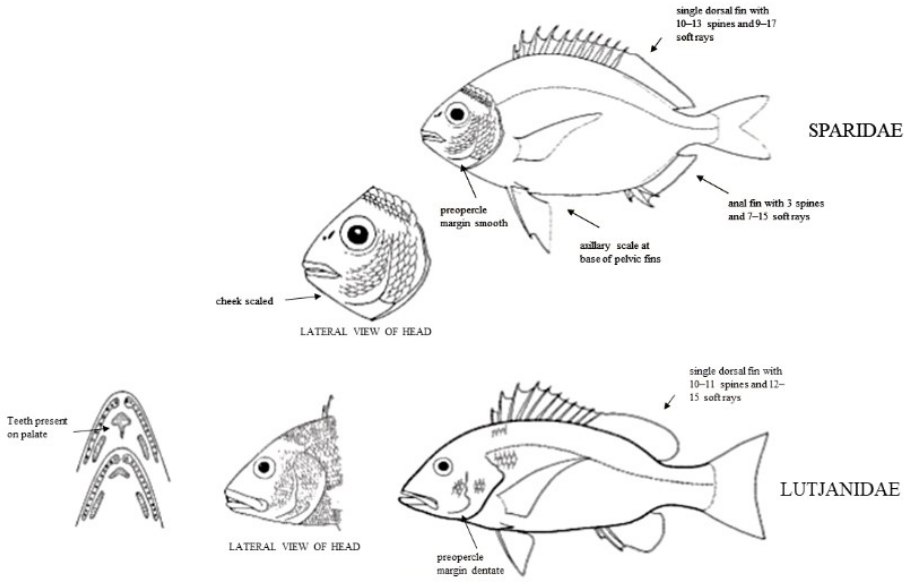


Figure 8. Morphological features of Lutjanidae and Sparidae. Source by Allen, 1985.



1.5. Seafood frauds: how to fight them?

In most countries of the world, there are different government agencies responsible for regulating seafood fraud and there is not a single food law directly addressing all aspects. The responsibility for controls is usually shared among official food monitoring authorities, border protection agencies, customs import authorities and specialist agencies within the national police force.
The primary driver of fish fraud is the profit of the perpetrators. Some of the most common cases of fish fraud are replacement of lower-value species with more expensive varieties and mislabelling of illegally caught fish to laundering through the legitimate fish marketing chain. Additional aspects that contribute to fish fraud are (FAO, 2018):
a) a deficiency of harmonisation between common fish names applied at national level in the different country (eg. some species have the same common names);
b) importation controls are not rigorous, where accredited methods of fish species identification using *DNA* barcoding are not regularly used;
c) poor organisation at national level between food control authorities;
d) the absence of an approved international regulation on "food fraud".
The global trade management of the fishery and aquaculture is regulated by a combined control system that covers throughout the supply chain to retail, with the dual objective of ensuring the sustainability of the sector and the traceability of the products. In EU after the implementation of legal provisions intended at preventing illegal fishing IUU (EC, 2016) and improving transparency and responsibility in the seafood supply chain (EU, 1169/2011), seafood fraud rates have apparently started to decrease since 2011 (Warner et al., 2016), primarily in those EU countries where the rules are obligatory.
The European Union (EU), within the renewal plan of the Common Fisheries Policy and the Common Market Organization, with the Cape IV of Regulation (EU) 1379/2013 (EU, 1379/2013) have introduced new requirements for the labelling of fisheries and aquaculture products. These, as well as providing consumers with more complete information, integrate the provisions of Regulation (EU) 1169/2011 and acts as a tool to prevent frauds. The dispositions introduced by Regulation (EU) 1379/2013 entered into force the 14st November 2014. This regulation, at the Art. 35 of the Cape IV, establishes the new mandatory labelling requirements for Fisheries and Aquaculture Products (FAPs):

a) commercial and scientific name of the species;
b) production method ("… caught …" or "… caught in freshwater …" or "… farmed …"); area where the product was caught or farmed using the FAO Major Fishing Areas (FAO, 2018b);
c) category of fishing gear used;
d) whether the product has been defrosted (with limited exceptions);
e) date of minimum durability – where appropriate.
In addition, prepacked fishery and aquaculture products marketed in the European Union (Member Organization) must also display all the relevant information specified in Articles 9 and 10 of Regulation (EU) 1169/2011 on the provision of food information to consumers.
The Art. 35 of the Regulation (EU) n. 1379/2013, refers to the Common Customs Tariff Combined Nomenclature (CN) (EEC, 2658/1987). According to Art. 35 of Cape IV, pre-packed and non-prepacked FAPs, registered in Annex I of the Regulation (EC) n. 1379/2013 under letters (a), (b), (c) and (e), may be "offered for sale to the final consumer or to a mass caterer", only if correctly labeled. The "mass caterers" of the Art. 35 includes restaurants, canteens, schools, hospitals and catering enterprises. In addition, with the Regulation (EU) n. 1169/2011 a detailed definition for mass caterers was provided: "*any establishment (including a vehicle or a fixed or mobile stall), such as restaurants, canteens, schools, hospitals and catering enterprises in which, in the course of a business, food is prepared to be ready for consumption by the final consumer*" .While in the Regulation (EU) n. 1169/2011 clearly reports that…"*shall apply to all foods intended for the final consumer, including foods delivered by mass caterers and foods intended for supply to mass caterers*". The FAPs cited to in points (a), (b) and (c) are all under the Chapter 3 to the Common Customs Tariff Combined Nomenclature (CN) (EEC, 2658/1987), "Fish and Crustaceans, Molluscs and other Aquatic Invertebrates" contain un-processed FAPs (fresh, chilled or frozen) as whole, filleted or minced and can be also found certain processed products, such as those dried, salted, in brine or smoked, the point (e) refers to the category "Seaweeds and other algae", under the Chapter 12 at of the Customs Tariff Code (Table 8). All the other processed FAPs (preserves, semi-preserves, canned, breaded, ecc), falling under Chapter 16 of the Customs Tariff Code, do not have to be labeled according to the Art. 35.

Table 8. Product categories of the Council Regulation (EEC) n. 2658/87 (and its following amendments), which fall under the Cape IV of the Regulation (EU) n. 1379/2013 as reported by D'Amico et al., 2016.

Point Annex I of Reg. (UE) 1379/2013	CN code	Description of the goods
a)	0301	Live fish
	0302	Fish, fresh or chilled, excluding fish fillets and other fish meat of heading 0304
	0303	Fish, frozen, excluding fish fillets and other fish meat of heading 0304
	0304	Fish fillets and other fish meat (whether or not minced), fresh, chilled or frozen
(b)	0305	Fish, dried, salted or in brine; smoked fish, whether or not cooked before or during the smoking process; flours, meals and pellets of fish, fit for human consumption consumption
(c)	0306	Crustaceans, whether in shell or not, live, fresh, chilled, frozen, dried, salted or in brine; smoked crustaceans, whether in shell or not, whether or not cooked before or during the smoking process; crustaceans, in shell, cooked by steaming or by boiling in water, whether or not chilled, frozen, dried, salted or in brine; flours, meals and pellets of crustaceans, fit for human
	0307	Molluscs, whether in shell or not, live, fresh, chilled, frozen, dried, salted or in brine; smoked molluscs, whether in shell or not, whether or not cooked before or during the smoking process; flours, meals and pellets of molluscs, fit for human consumption
(e)	1212 20 00	Seaweeds and other algae

Table 9. Product categories of the Council Regulation (EEC) n. 2658/87 (and its following amendments) regarding the species Dentex dentex and Pagellus spp., which fall under the Cape IV of the Regulation (EU) n. 1379/2013 as reported by D'Amico et al., 2016, modified by the autor.

Point Annex I of Reg. (UE) 1379/2013	CN code	Description of the goods
(a)	0301	**Live fish**
	0301 9985 80	- - - - Sea breams of the species *Dentex dentex* and *Pagellus spp.*
	0302	**Fish, fresh or chilled, excluding fish fillets and other fish meat of heading 0304**
	302	- -Snappers (*Sparidae*)
	0302 8510 00	- - -Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0303	**Fish, frozen, excluding fish fillets and other fish meat of heading 0304**
	303 0 00	- - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0304	**Fish fillets and other fish meat (whether or not minced), fresh, chilled or frozen**
	0304 41	- Fillets, fresh or chilled, from other fish
	0304 4990 60	- - - - - of Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0304 5300 _31	- - - - Intended for processing
	0304 5990 40	- - - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0304 81	- Frozen fillets of other fish
	0304 8990 30	- - - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0304 9923 10	- - - - - - Sides, weighing more than 80 g per piece, intended for processin
	0304 9999 20	- - - - - - Sea breams of the species *Dentex dentex* and *Pagellus spp.*
(b)	0305	**Fish, dried, salted or in brine; smoked fish, whether or not cooked before or during th smoking process; flours, meals and pellets of fish, fit for human consumption consumpti**
	0305 1000 30	- - Sea breams of the species *Dentex dentex* and *Pagellus spp.*
	0305 31	- Fish fillets, dried, salted or in brine, but not smoked
	0305 3990 70	- - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0305 3990 80	- - - - Basses (Dicentrarchus labrax)
	0305 3990 90	- - - - others
	305	- Smoked fish, including fillets, other than edible fish offal
	0305 4980 40	- - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*
	0305 51	- Dried fish, other than edible fish offal, whether or not salted but smoked
	0305 5985 65	- - - - Sea breams of the species *Dentex dentex* and *Pagellus spp*

Currently the Art. 35 is not applicable to prepared and processed FAPs and the dishes made of seafood can be indicated in the menu with general and incomplete names, such as “Dentice”, “Pagello”, “Pagro” with no further specifications. In this way, restaurants and other mass caterers are not obliged to put the mandatory information on their menus unless the Competent Authority requires so. They can do it voluntarily to advance the credibility of their business, as they are just obliged to keep such information and show the documents to the consumers if they require it. In these regards, it has been highlighted mislabeling of fish and seafood at restaurants and take aways are suggestively greater than that happening at retail level (Pardo et al., 2016). Frauds at restaurant level are much more frequent in the case of sales of local seafood often substituted with imported similar products, as well as the replacement of wild with farmed fish is frequent (Warner et al., 2016). Sparids is an excellent example of this problem; mislabeling of *Sparidae* species could be voluntary and aimed at charging higher prices on low commercial value species (Armani et al., 2015a).

This problem is strongly felt by the UE that, with Resolution of 12 May 2016 on traceability of fishery and aquaculture products in restaurants and retail (2016/2532 RSP), published on Official Journal of the European Union, C 76, 28 February 2018 (EU, 2016) at point 1) states “*Expresses its serious concern and discontent at the results of various studies showing significant levels of mislabelling on fish products sold on the EU market, including in the restaurants of the EU institutions; reaffirms that the intentional and fraudulent mislabelling of fish species is a breach of EU regulations, including the Common Fisheries Policy, and may constitute a criminal offence under national law”*, at point 5)“*Calls on the Member States to strengthen national controls, including on non-processed fish for restaurants and the catering sector, in an effort to tackle fraud and identify the stage in the supply chain where fish is mislabelled; is concerned by the substitution of high-quality species with lower-quality counterparts”*, at point 11) “*Notes that some commercial fish denominations in force in the Member States vary among them owing to national practices that could lead to a degree of confusion; welcomes the work undertaken by the Commission to launch a pilot project, as adopted by Parliament, aimed at introducing a public database which will provide information on commercial denominations in all the EU official languages”.*

As for the commercial fish denominations in force in the Member States mentioned in point 11 of the Resolution of the EU Parliament of 2016 and in point a) at the Art. 35 of the Cape IV of the Regulation n. 1379/2013, they are established in each Member State with specific regulatory device. In fact, the subsequent art. 37 point 1 of the Regulation n. 1379/2013, regulates further clarifications for the "commercial designation" of seafood product, asking the Member States to publish a list of the commercial designations accepted in their territory, together with their scientific names, to which the food business operator must comply for the correct preparation of labels and traceability systems. Specifically, the list shall indicate:

(a) the scientific name for each species, in accordance with the FishBase Information System or the ASFIS database of the Food and Agriculture Organization (FAO), where relevant;

(b) the commercial designation:

(i) the name of the species in the official language or languages of the Member State concerned;

(ii) where applicable, any other name or names that are accepted or permitted locally or regionally.

The provisions of art. 37, letter a) has generated considerable operational difficulties for administrations, food business operators and the competent authority for the verification of traceability.

In detail, it is highlighted that, for the identification of the scientific name, according to the characteristics of the product, the geographical distribution and other parameters, the Reg. (EU) 1379/2013 identifies the "FishBase information system" (ASFISS) which presents a large collection of "fish", about 33.700 species, but no information on crustaceans, bivalve molluscs, molluscs cephalopods, echinoderms, tunicates, etc.

With this in mind, the Ministry of Agricultural, Food and Forestry Policies, with the Ministerial Decree of 22 September 2017 concerning the "Attribution of denominations in Italian of fish species of commercial interest" (DM 19105 of 22 September 2017), intended to provide details for identifying the taxonomic characteristics of fish species of commercial interest (in the attached actually are present 1054 species) in the Art. 5 regulates the reference to the SeaLifeBase and Worms databases (World Register of Marine Species), exclusively for crustaceans, molluscs, echinoderms and tunicates, in addition to Fishbase and ASFIS, identified for fish. The decree leaves unchanged the reference to the ASFIS database,

prepared by the FAO, which, however, gives only indication of the scientific name.
In the meantime, at EU level, it was established "The EU Database on marketed fishery and aquaculture products" as already anticipated in the Resolution of European Parliament, where commercial denominations in all the EU official languages are present and available at the following link: https://mare.istc.cnr.it/fisheries/demo/index.xhtml?lang=en.
In conclusion as mentioned in the point 3 by the Resolution of European Parliament:
"Supports a strong traceability system, from landing to consumption, which would give confidence to consumers and, in turn, decrease commercial dependency on imported fishery and aquaculture products, thereby strengthening the EU market; calls on the Commission to exploit the potential of DNA barcoding, which could assist in the identification of species by DNA sequencing, in order to enhance traceability".

The European law aims to ensure a high level of protection of human life and health, establishing a comprehensive system of traceability with a "farm to fork" approach. Therefore, traceability must be assured at all production steps. Scientific and technical methods have always to support Community legislation. This research provides the basis for the molecular traceability of fish products.

Currently, there is a need for harmonization and standardization of analytical techniques and for universal access to a standard database. Nowadays, there is not an official and univocal method internationally used for the genetic identification of species. In Italy, official laboratories (Ministry of health, Zooprofilattici Institutes, etc.) do not use all the same protocol of analysis for species identification. A better cooperation between food control authorities and law enforcement agencies is required to prevent and fight food fraud activities. The introduction of new and performing analytical method for fish species identification means that food inspectors and consumers will be aware that what is declared on the label exactly corresponds to the real purchased food product.

1.6. Seafood species identification: *DNA* analysis

Label information frequently do not offer enough security about the true contents of a seafood product (See section 1.5). Consequently, it results highly necessary to identify and/or authenticate seafood components to

guard both consumers and producers from illegal substitutions. Fish species identification is usually based on the analysis of many morphological and meristic features (see section 1.4), mainly represented by body shape, colours, number and position of fins or teeth shape, which have also been incorporated and descripted in detail on identification sheets proposed by FAO (Carpenter & Niem, 2001). Another important tool to identify fish by morphologic features is represented by the online "FishBase information system" which presents a large collection of "fish", about 33,700 species (taxonomy, biology, trophic ecology, life history and uses). The SeaLifeBase and Worms databases (World Register of Marine Species) are useful to identify exclusively fcrustaceans, molluscs, echinoderms and tunicates.

During real inspection activity, however, this kind of approach frequently requires considerable expertise and can be used only when specimens are intact. On the opposing, species identification becomes extremely complicated in fillets and other prepared forms, and totally impossible in highly processed products as key morphological characteristics are absent. Consequently, to implement labelling regulations and prevent product substitution, there is a need for sensitive analytical methods that can be used to determine the species of a seafood product with no obvious external features. Methods for species recognition are principally based on the finding of polymorphism in proteins or *DNA* that are exclusive to each species, as a sort of “fingerprint”. The diagnostic techniques used to establish the unique fingerprint must be adjusted for the specific product under examination, besides being able to offer incontrovertible and repeatable results. Most important challenges are characterised by the fact that several species might have similar fingerprints or individuals from the same species display different fingerprints due to intra-species variation (Rasmussen & Morrissey, 2008). Furthermore, certain processing is known to denature proteins and partially degrade *DNA*, making analysis of processed seafood foods especially hard (Armani et al., 2012a). To conclude, some compounds existing in processed foods may also act as inhibitor during laboratory analysis (Teletchea et al., 2005). Numerous analytical methods have been developed and optimized for the difference of seafood species in a variety of product types. Traditional methods used in species identification are based essentially on the separation and characterization of specific proteins using electrophoretic techniques (Teletchea, 2009), such as:

-isoelectric focusing (IEF),

-capillary electrophoresis (CE),

-high-performance liquid chromatography (HPLC),

-immunoassay systems (Enzyme Linked Immuno Sorbent Assay - ELISA).

Isoelectric Focusing methodology (Lundstrom, 1980) was used by FDA as official method to afford species-specific biochemical patterns for some of the fish species reported in the Regulatory Fish Encyclopedia. Even if this method has been positively used in several cases, it cannot always be applied, specially for processed products. Most proteins are temperature labile, they become irreversibly denatured after strong heat processing or drying and consequently can no longer be examined by techniques suitable for their natural states (Teletchea, 2009). A protein-based technique that may demonstrate to be useful, even in heat-sterilized products, is Enzyme-Linked Immunosorbent Assay (ELISA). However, it can be ineffective at differentiating closely related species and results inclined to cross-reactivity with non-target proteins (Rasmussen & Morrissey, 2008). Additionally, seafood processing may occasionally induce antigen conformational change, producing altered epitopes that are not recognized by specific antibodies (Preedy, 2015). At the beginning of the years '90 an article titled ''*The label said Snapper, the lab said baloney*'' was published. Since that time, researchers have tried to develop new methods, based on *DNA* markers, for seafood species identification. Today, most of the methods for seafood species identification are based on *DNA* analysis (Griffith et al., 2014). The use of *DNA*-based approaches presents in fact a number of benefits over protein-based methods. First of all, molecular evolution and phylogenetic studies have revealed that, because of the degeneracy of the genetic code and the presence of many non-coding regions, *DNA* provides much more data than proteins do (Teletchea et al., 2005). Before, while proteins vary with tissue type, age, and status, DNA is mostly independent of these factors, is present in all cell types and therefore it can be recovered from any tissue allowing the collection and storage of samples in every stage of the fishery chain (Civera, 2003). Other key feature, even though it might be transformed by various processing (e.g. canning, heating), is that *DNA* is general more resistant and thermostable than proteins (Teletchea et al., 2005). For these reasons,

many *DNA*-based methods have been applied, and are currently being studied for seafood species identification. Essentially, in all the techniques the *DNA* is firstly extracted from the tissue and then the *DNA* fragment(s) of interest is amplified using Polymerase Chain Reaction (PCR). The resultant PCR amplicons are then analyzed (differently according to the technique) to reveal the characteristic polymorphisms under study. These approaches vary in their series of applications, difficulty and costs, and depending on the objective of the study, kind of samples analyzed and available funds, it is possible to choose between different methods based on their respective advantages and disadvantages (Teletchea, 2009).

1.6.1. Polymerase Chain Reaction (PCR).

Developed in 1983 by Kary Mullis and colleagues (Mullis et al., 1986), Polymerase Chain Reaction (PCR) is now routinely applied in labs to amplify part of the *DNA*. Amplification of genetic material with PCR necessitates a thermostable *DNA* polymerase, 2 oligonucleotide primers, 4 deoxynucleotide triphosphates (dNTPs), and magnesium ions. PCR involves numerous cycles of 3 reaction steps carried out at different temperatures: denaturation (approximately 95 ◦C), annealing (50 to 60 ◦C) and extension (approximately 72 ◦C). Through these 3 steps, the template *DNA* is initially divided into 2 single strands by heat denaturation, then the oligonucleotide primers anneal to complementary sequences on opposing ends of a specific fragment of the template *DNA*, and next a thermostable *DNA* polymerase uses the 4 dNTPs to produce copies of the target *DNA* fragment. *DNA* polymerases copy single-stranded *DNA* templates by adding deoxynucleotides to a growing chain (extension product). Chain elongation occurs at the 3' end of a primer and the extension product grows by the formation of a phosphodiester bridge between the 3'-hydroxyl group on the primer and the 5'-phosphate group of the incoming deoxynucleotide. Growth occurs in the 5' -> 3' direction. Generally, about 20 to 50 cycles of denaturation, annealing, and extension are done, and the *DNA* fragment is amplified into millions of copies. The amplified *DNA* fragment, called amplicon, is then present in enough amounts for analysis by a variety of post-PCR analysis. If standard PCR conditions do not yield the desired amplicon, PCR optimization is necessary to obtain better results. The stringency of a reaction may be modulated such that the specificity is adjusted by altering variables (e.g., reagent concentrations, cycling conditions) that affect the outcome of the amplicon profile (Lorenz, 2011). Occasionally, success simply may depend on on changing the

concentration of MgCl2, dNTPs, primers, template *DNA*, or *DNA* polymerase; the concentration of MgCl2 can have a profound effect and yield of amplification.

Stabilizer reagents possibly will yield results when all else fails. For example, addition of specific proteins such as Bovine serum albumin (BSA) is described to improve PCR performance (Farell & Alexandre, 2012). Modification of cycling condition may also represent a valid strategy, especially regards the annealing time and temperature (Lorenz, 2011).

Real-time PCR uses specialized thermocyclers that perceive the fluorescent signal in separately well. This signal is indicative of the quantity of double-stranded *DNA* within the reaction tube or well. This signal, in relative fluorescent units, is plotted by the thermocycler software versus cycle number.

1.6.2. Genetic markers for PCR-based analysis

A molecular marker may be definite as a *DNA* sequence that can be used to ‘‘mark’’ a specific position (locus) on a precise chromosome, i.e. marker gene (Bhattacharya et al., 2015). Marker genes are known to have an exact location or clear phenotypic expression that can be distinguished by simple analytical methods and which can then enable the study of inheritance of a trait or a gene (Hohenlohe et al., 2011). Molecular genetic markers are powerful tools to detect genetic variation between individuals, populations or species. The genetic variation is essentially represented by point mutations named also single nucleotide polymorphisms (*SNPs*), indels (insertions or deletions of nucleotide sequences), inversion of a segment of *DNA* within a locus, and rearrangement of *DNA* segments around a locus of interest (Dudu et al., 2015). To choose a good genetic marker for seafood species identification is essential to consider some aspects:

- **Inter-species variability**: All organisms are subject to mutations because of normal cellular operations or interactions with the environment, leading to genetic variation (polymorphism) that develops the ability of organisms to adapt to changing environment and is necessary for survival of the species. In combination with other evolutionary forces like selection and genetic drift, genetic variation arises among individuals leading to differentiation at the level of population, species and higher order taxonomic groups. A suitable

DNA marker for the identification at species level should be sufficiently variable between species (particularly between the closest ones) and display either low or no intra-specific variations across the geographic distribution area (Teletchea et al., 2005).

- **Marker sequence length**: Depending on the tipe of samples analysed, the quality of *DNA* varies greatly, the method chosen to isolate the marker.
  In un-processed seafood, *DNA* is generally unaltered and large *DNA* fragments (~1-2 kb) can potentially be amplified with classical PCR apporoach (Teletchea, 2009), other authors amplified much larger fragments such as ~8-9 kb and more with Long PCR approach (Mascolo et al., 2017; Miya et al., 2001). However, amplicons of >1000 bp may need extra time to be completed during the extension step and need to consider extra time to sequence that frangments with classical Sanger sequencing.
  In the case of amplification with Real-Time PCR the amplicon size plays an important role in amplification efficiency, it is generally recommending the use of relatively short amplicon lengths, in the range of 50 to 150 (bp) base pairs (Debode et al., 2017).
  The amplicon size, moreover, depends on the platform of sequencing method (Table 7).
  The quality and the tipe of samples analysed, influence the marker sequence length. In many processed seafood products, *DNA* may be degraded by treatments such as heat exposure, high pressure, low pH, irradiation, drying, nucleases that cause enzymatic degradation, depurination and hydrolysis, among others and, in this case, identification methods must be based on the analysis of very short *DNA* fragments (~50 and 250 bp) (Carvalho et al., 2015; Cawthorn et al., 2012).

Table 7. Amplicon size and platform of sequencing method.

	Platform type	Amplicon size (bp)
1st generation	Roche 454 GS FLXTM[1]	700-800
	Life technologies Ion Torrent[2]	100, 200, 400
	Illumina MiSeq 4000[3]	150
	Illumina MiSeq[3]	300
	Illumina MiniSeq[3]	150
2st generation	MinION[4] (Oxford Nanopore Technologies Ltd)	Up to 300

- **Available sequences**: appropriate genetic marker should be generally studied for many species to permit comparison of the nucleotide sequence from an unknown sample with reference sequences in a database (Giusti et al., 2016; Teletchea et al., 2005).

1.6.3. Nuclear DNA or mitochondrial DNA??

The identification of seafood species can be carried out using either nuclear *DNA* (n*DNA*) or mitochondrial *DNA* (*mtDNA*) (Martinez et al., 2005). In animals, *mt*DNA occurs as a single double-helical circular molecule containing 13 protein-coding genes, 2 ribosomal genes, a 2 non-coding region, and several t*RNA*s (Waugh, 2007). To time, most studies focused on *mtDNA* rather than n*DNA* due to several advantages. Firstly, it displays overall a higher copy number (about 1,000 X the copies of n*DNA*); in fact, since each mitochondrion contains several *DNA* circular molecules and in turn each cell has several mitochondria, exploitable source of *DNA* is alike relatively abundant even though sample tissue is limited (Waugh, 2007); *mtDNA* it is relatively simple and small compared to n*DNA* because it lacks features such as large non-coding sequences (introns), pseudogenes, repetitive *DNA* and transposable elements; it is maternally inherited, as in

fish the paternal mt*DNA* is degraded after entering the egg (Brown, 2008), so that individuals have only one allele and this avoids sequence ambiguities from heterozygous genotypes; moreover, it usually evolves much faster than nuclear *DNA* and its higher relative mutation rate results in the accumulation of enough sequence differences to enable discrimination of closely related species (Rasmussen & Morrissey, 2008); then, the complete sequence of *mtDNA* is known for several aquatic organisms, because studied for evolutionary studies (Palumbi et al., 2002); lastly, *mtDNA* is usually preferred due to the concept that its circular structure gives it greater resistance to heat-induced degradation and it can be consequently considered a better aspirant in processed seafood analysis (Civera, 2003). On the contrary, one disadvantage of using *mtDNA* is assumed by the fact that hybrids cannot be distinguished (Brown, 2008).

1.6.3.1. Most used nuclear genes

Infact, despite the numerous benefits of *mtDNA* in species identification research, a number of n*DNA* targets have also proven to be positive, particularly in identifying farmed seafood that are quite often represented by hybrids, as in case of tilapia, catfish and carp species (Rehbein, 2013), and also for wild freshwater fish, where hybridization between related fish species is not unusual (Wyatt et al., 2006). In several studies, different varieties of nuclear genes have been used for identification of fish species by PCR. The nuclear *5S* r*RNA* gene, which consists of a small 120 bp conserved region coding for *5S* r*RNA* and a variable region of non-coding *DNA*, termed the non-transcribed spacer (*NTS*), that have a species specific length and sequence has been used in the past to recognize mackerel, gadoids, salmonids, sharks, and others (Aranishi et al., 2005; Morán & Garcia-Vazquez, 2006; Pendas et al., 1995). Additional tested genes are represented by the *p53* gene (Carrera et al., 2000), the nuclear ribosomal internal transcribed spacer 2 (*ITS2*) locus, the *18S* r*RNA* gene, the internal transcribed spacer (*ITS 1*) located between the *18S* and *5.8S* r*RNA* genes (Herrero et al., 2011) protein encoding genes like rhodopsin (Rehbein, 2013; Sevilla et al., 2007), *α*-actin or *β*-actin (Lee & Gye, 2001), parvalbumin (Rehbein, 2013) and for *7S* ribosomal protein (Chow & Hazama, 1998), and a major histocompatibility complex (*MHC*) class II gene (Rasmussen & Morrissey, 2008). In gadoid species identification, partial *DNA* sequence encoding the integral membrane protein of cytoplasmatic microvesicules pantophysinI was sometime chosen (Hubalkova et al., 2008). n*DNA* also contains short tandemly repeated

segments of *DNA* that occur throughout the genome and exhibit a high degree of polymorphism, known as microsatellites. Thanks to their high levels of degeneracy and variability, microsatellites, also referred to as Variable Number of Tandem Repeats (*VNTR*), have demonstrated to be very valuable in studies on population genetics and have been used, for example, in distinguishing species of european sturgeons (Chassaing et al., 2011) as well as other commercial marine species, including rainbow trout, channel catfish (*Ictalurus punctatus*), sun-catfish (*Horabagrus brachysoma*), carp, salmonids and various more (Rasmussen & Morrissey, 2008).

1.6.3.2. Most used mitochondrial genes

The cytochrome b (*cytb*) certainly was the most appreciated gene of the first decade of 2000, used in more than half of the phylogenetic studies published in those years, mostly thanks to the fact that it contains both slowly and rapidly evolving codon positions, as well as more conservative and more variable regions or domains overall, which makes it appropriate for the differentiation of even closely related species through the project of universal primers (Pereira et al., 2008). It has in fact been used, nowadays as in the past, to identify flatfish, gadoids, anchovies, eels, scombroids, anglerfish and many others (Castigliego et al., 2015; Armani et al, 2011; Chow et al., 2003; Lin & Hwang, 2007; Pepe et al., 2007; Santaclara et al., 2006; Akasaki et al., 2006; Pepe et al., 2005; Calo-Mata et al., 2003; Sotelo et al., 2001). Other common *mtDNA* targets in species identification research are the small *12S* r*RNA* gene (819 to 975 bp in vertebrates) and the larger *16S* r*RNA* gene (1571 to 1640 bp in vertebrates). The r*RNA* genes are between the most conserved nucleotide sequences existing in a cell. Percentages of the *16s* r*RNA* sequence from distantly related organisms have been found strangely similar. That means that sequences from remotely related organisms can be exactly aligned with each other to measure true differences between them (Bhattacharya et al., 2015). For this motive, it have been widely used to determine taxonomy, phylogeny (evolutionary relationships), and to estimate rates of species divergence among numerous commercial species, such as flatfish, eel, cardinalfish, mackerel, icefish, hairtail species, crab, cephalopods (Armani et al., 2015c; Lakra et al., 2011; Chakraborty et al., 2005; Imai et al., 2004; Itoi et al., 2005; Chapela et al., 2002). With the expansion and the success of the *DNA*-Barcoding technique proposed by Hebert et al. (2003) (Hebert et al., 2003), Cytochrome oxidase subunit I (*COI)* gene has been

recognized by many authors as the marker of choice for species discrimination. It has in fact a better potential for phylogenetic studies than any other mitochondrial gene and there are a broad range of very robust universal primers applicable across a wide range of taxa. In addition to the *cytb*, *COI* and r*RNA* sequences, there are several additional *mt*DNA targets. In general, non-coding segments like the *D-loop* exhibit elevated levels of variation relative to coding sequences such as the *cytb*, presumably due to reduced functional limits and relaxed selection pressure. Thanks to the hyper variable *D-loop* or control region sequence it has recently been used as target in commercial seafood identification studies (Nagpure et al., 2015; Natacha et al., 2016; Pedrosa-Gerasmio et al., 2012). Additional less used marker is represented by the gene coding for cytochrome oxidase subunit III (*COIIII*), which has been used to differentiate rainbow trout (*Oncorhynchus mykiss*) and Atlantic salmon (*Salmo salar*); and the flanking region between *COIIII* and the NAD dehydrogenase 5 (*NAD5*) used to differentiate various species of tunas (Mya et al., 2015).

1.6.4. DNA-Barcoding.

Growths in *DNA*-sequencing technology has permits the opportunity of using variations in short sequences of *DNA* as labels for species in a technique that has become known as *DNA*-Barcoding (Waugh, 2007). The intent of *DNA* barcoding is to use one or a few reference genes to assign newly sampled individuals to species, or to identify cases in which no name yet occurs for a specified species (Hanner & Gregory, 2007). A *DNA* barcode is in fact a short sequence of nucleotides reserved from an appropriate part of an organism's genome that is used to identify it at species level. Intraspecific variation in this fragment is an order of magnitude less than that observed interspecifically, and this offers the means by which species are differentiated (Waugh, 2007). Essentially, the amplified target fragment is sequenced and this sequence, the ''barcode'', is matched with existing barcodes or material from voucher specimens deposited on official database "Barcode of Life Database"(http://v3.boldsystems.org/) The Kimura 2-parameter (K2P) genetic distance correction is used to quantify sequence divergences among individuals since it is the most effective model when distances are low (Hebert et al., 2003). Hebert et al. (2003) proposed in fact to standardize the various approaches used in species identification through

the establishment of a *DNA*-Barcoding system, similar in practice to a supermarket barcode, for all living organisms, based on a single sequence, represented by a ~650-bp portion of the *COI* gene. *COI* has so far prospered in providing species level resolution across diverse groups of insects, birds, fishes and primates and also confirmed the ability to differentiate species in other compartments of life, including protists and fungi (Hanner & Gregory, 2007). Several studies have shown the applicability of *DNA* barcoding for accurate identification of a wide range of fish species (Armani et al., 2015a; Carvalho et al., 2011; Carvalho et al., 2015; Cawthorn et al., 2012).

Therefore, the use of the gene *COI* had suffered success, contemporary research shows that mitochondrial DNA markers perform well for certain species but may be less discriminating for others (Trotta *et al.*, 2005). Commonly used *COI* markers may evolve too slowly and have low nucleotide sequence divergence in certain taxonomic groups, thus impeding discrimination of closely related fish species (*e.g.*, *Tuna*) (Rubinoff *et al*., 2006; Ward *et al*. 2009; Terio *et al*., 2010) when the gene does not contain effective regions for barcoding applications (Deagle *et al*., 2014). Consequently, the *COI* barcode sequence cannot be used extensively for the identification of all fish species. This situation has led to the formulation of the proposal to study and analyze the complete *mtDNA* sequence with the aim to identify mitochondrial markers or multiple marker approaches (Miya *et al*., 2015) with higher and more specific discrimination capacity.

1.6.5. Pcr primer pairs: importance of their selection in analytical techniques

PCR primer pairs are responsible for binding exact regions of target *DNA* to define the PCR fragment to be amplified and their variety is probably the most significant factor to consider for the efficacious identification of seafood species (Rasmussen & Morrissey, 2008). Primer design is pointed at obtaining a equilibrium between two goals: specificity and efficiency of amplification. Specificity is definite as the frequency with which a mispriming incident occurs. Primers with poor specificity be disposed to generate PCR products with undesirable amplicons. Efficiency is defined as how close a primer pair can amplify a product to the theoretic optimum of a twofold intensification of product for each PCR cycle (Dieffenbach et al., 2008). Together specificity and efficiency of amplification mainly depend on some primer design variables. The specificity is generally due

by the length of the primer, the annealing temperature of the PCR reaction. Also, the placement of the 3' end of the primer is critical for a successful PCR reaction: perfect base-pairing between the 3' end of the primer and the template is best for obtaining good results (Kwok et al., 1990). Additional factors influencing primers efficiency are the Guanine (G) and Cytosine (C) content and the primer Tm. Inside a primer pair, the GC content and Tm should be well matched, as poorly matched primer pairs can be less efficient and specific because loss of specificity arises with a lower Tm and the primer with the higher Tm has a greater chance of mispriming under these conditions.

1.6.5.1. Universal primers

Universal primers are designed to anneal to regions of *DNA* that are usually conserved across species groups and amplify a *DNA* fragment that displays interspecies variation (Rasmussen & Morrissey, 2008). Although universal primers can bind to a wide variety of *DNA* templates, they cannot assure *DNA* amplification of all types of organisms, due to the presence of mutations which cause primer sequence mismatches (Armani et al., 2012a). Consequently, even yet the Taq polymerase is tolerant to mismatches, these primers are usually degenerated at variable nucleotide sites to increase the possibility of amplification, with a single PCR reaction, of the same DNA fragment from a wide range of taxa (Lang & Orgogozo, 2011). Due to the widespread use of *mt*DNA in genetic research, many universal primers have been designed on this molecule and, particularly for seafood, most of the available primers were projected on the *cytb*, *COI*, *16S, 12S* genes (Armani et al., 2012a). Between these primers, those targeting the *COI* gene are often degenerated while the high degree of conservation of *16S* gene does not need this modification (Cawthorn et al., 2012). The possibility to amplify *DNA* from a wide range of species has also been tested through the utilization of cocktail of primers targeting different molecular genes (Ivanova et al., 2007). Universal primers are useful for the amplification of a *DNA* fragment that will be analyzed through post-PCR methods, such as PCR coupled with Restriction Fragment Length Polymorphism (PCR-RFLP) and PCR-sequencing.

1.6.5.2. Specie-specific primers.

Species-specific primers are designed based on single nucleotide polymorphisms (*SNPs*) to anneal only to *DNA* from a given species or

category (Lockley & Bardsley, 2000). Though this method requires detailed knowledge of the *DNA* sequences from target species, this data is becoming available with the access to public genetic databases. High-quality primers can be designed through specifically dedicated available software (Beacon designer, Primer3, DNASTAR etc.) as well as directed manually. In this latter case, three key parameters should be considered (Giusti et al., 2016):
a) The specificity of the primers toward the target sequence, which is the capability of the primers to amplify only the premeditated target (and obviously not to amplify any unplanned ones). The selection of the conserved or polymorphic regions for designing common or genera - specific primers can be carried out by aligning the sequences with software like BioEdit (Hall, 1999);

b) The amplicon length, which must be different between target *DNA* of at least 100 bp to easily discriminate them on an electrophoresis gel, should also be selected by taking into consideration the level of *DNA* degradation of the samples analyzed;

c) The annealing score can be assigned by specific software, such as Primer Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) to the based on the primers' overall characteristic (melting temperature, GC content, molecular weight, extinction coefficient, and tendency to form dimers). Once the size of the product has been predicted, identification is confirmed if the expected sized amplicon is seen on a gel. By pairing species-specific primers with a nonselective (as e.g. family-specific) counterpart, it is possible to test for the presence of more than one species simultaneously, as in the case of multiplex PCR approach. The non-selective primer tends to be based on a sequence that is common to all species under study in a given system; its precise location in the gene can be used to dictate the size of the amplification products that will be generated (Lockley & Bardsley, 2000). The utilization of species-specific primers, allowing species detection by the presence (or absence) of the PCR amplicon on an agarose gel, do not need for traditional analytical procedures such as PCR-RFLP or PCR-sequencing.

1.6.6. Post-PCR analytical methods

Succeeding PCR amplification, the resulting DNA fragments must be properly analyzed with techniques that might be considered by a relatively low procedure cost (for routinely application in labs) and, obviously, should be reproducible, quick, and dependable. The most common traditional method for fish species identification are PCR coupled with Restriction Fragment Length Polymorphism (PCR-RFLP), most used in the past and a little put aside in the last years, and PCR-sequencing, which nowadays represents the most applied one.

1.6.6.1. Restriction Fragment Length Polymorphism (PCR-RFLP).

PCR-RFLP is a technique in which an amplified fragment is cut by endonucleases identifying specific restriction sites, resulting in few smaller fragments of different sizes. The different fragments are then separated by agarose gel electrophoresis. When no sequences are available for precise groups of species, the target DNA fragment must be amplified initially by PCR and then sequenced to find polymorphisms among the species of interest and choose the appropriate restriction enzymes (Teletchea, 2009). When the sequence of the fragment has been established, the initial sequencing step is no longer necessary, because the PCR amplicon of interest is simply digested with the preselected restriction enzymes and then its restriction pattern is compared with reference samples for species identification. A main difficulty of PCR-RFLP is the possibility for intra-species variation, in which individuals from the similar species exhibit different restriction patterns due to degeneracy in the *DNA* fragment being analyzed (Lockley & Bardsley, 2000). For this reason, it has been suggested that species identification with PCR-RFLP be carried out with caution if there is no considerable information available concerning sequence polymorphisms within and between species groups (Sotelo et al., 2001).

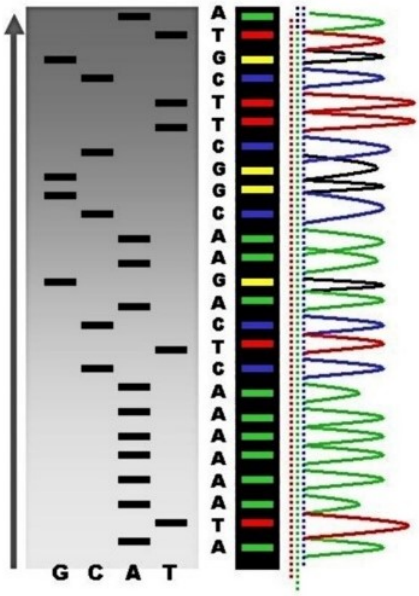
1.6.6.2. Classical PCR-Sequencing techniques.

Although PCR-sequencing was originally time consuming, technically demanding and rather expensive, both prices and time have today significantly been reduced, and this method has practically become the gold standard in species identification, since it results the method that produces the largest amount of information (Ceruso et al., 2019; Teletchea, 2009). The *DNA* sequencing technique developed by Fred Sanger (Sanger

et al., 1977) forms the basis of automated "cycle" sequencing reactions today. This type of sequencing, also known as the chain termination method, works as a classical PCR on the purified PCR fragment using the same primers together with a *DNA* polymerase, normal deoxy-nucleoside triphosphates (dNTPs) and, contrarily from the classical PCR, modified nucleotides (ddNTPs) that terminate *DNA* strand elongation. These ddNTPs lack a 3′-OH group that is required for the formation of a phosphodiester bond between two nucleotides, causing the extension of the *DNA* strand to stop when a ddNTP is added. The *DNA* sample is divided into four separate sequencing reactions, containing all four of the standard dNTPs (dATP, dGTP, dCTP, and dTTP), the *DNA* polymerase, and only one of the four ddNTPs (ddATP, ddGTP, ddCTP, or ddTTP) for each reaction. After rounds of template *DNA* extension, the *DNA* fragments that are formed are denatured and separated by size using gel electrophoresis with each of the four reactions in one of four separated lanes. The *DNA* bands can then be visualized by UV light or autoradiography, and the *DNA* sequence can be directly read off the gel image or the X-ray film. The ddNTPs may also be radioactively or fluorescently labeled for detection in automated sequencing machines. The four reactions can be incorporated into one reaction run, and the *DNA* sequence can be read from radioactive or fluorescent labels. A camera captures an image of this fluorescence, assumed that each of the four ddNTPs has its own unique fluorescent spectrum and thus there are four possible fluorescent images. The images captured by the camera are converted to a readable form called electropherogram (Figure 10), from which the sequence of the *DNA* of interest can then be determined. Consequently, the electropherogram is a graphical image of data received from a sequencing machine and is also known as a trace.

Noteworthy, universal primers are frequently tailed with non-degenerated non-homologous sequences at their 5′ ends (tails), to improve sequencing production (Binladen et al., 2007; Messing & Vieira, 1982; Steffens et al., 1993). M13 universal tails are the greatest used to date and are characterized by short sequences belonging to the genome of the wild-type M13 phage, a filamentous bacteriophage with a genome of single-stranded circular *DNA*. These tails have in fact been proved to be utilized as cloning site in Sanger sequencing technique (Sanger et al., 1977) which requires single-stranded *DNA* as a template and M13 can be easily obtained in this form.

Figure 10. Sequence ladder by radioactive sequencing compared to fluorescent peaks. Source by Abizar at English Wikipedia, CC BY-SA 3.0, (https://commons.wikimedia.org/w/index.php?curid=3800855).



1.6.6.3. Forensically Informative Nucleotide Sequencing (FINS).

Forensically informative nucleotide sequencing (FINS) is a method that combines DNA sequencing and phylogenetic analysis, used to identify samples based on informative nucleotide sequences. The sequence with the lowest genetic distance, or number of nucleotide substitutions, from the target fragment represents the species group to which the original sample belongs (Bartlett & Davidson, 1992). A combination of 2 mathematical modeling systems is generally employed to carry out the phylogenetic analysis: the Tamura-Nei method, to calculate the genetic distances among sequences (Tamura & Nei, 1993), and the Neighbor-Joining method, to construct a phylogenetic tree based on these genetic differences (Saitou & Nei, 1987). Clearly, since FINS is based on nucleotide sequence substitutions, it is significant to choose a fragment that displays high inter-species variability but low intra-species variability to avoid ambiguities in the determination of species (Rasmussen & Morrissey, 2008). FINS represent a reliable chance to evaluate the nature of seafood products and to verify the information reported on the label. To date, numerous studies, selecting different genetic markers, have applied

this technique to fish species identification (Armani et al., 2015c; Blanco et al., 2008; Espiñeira et al., 2009; Espiñeira et al., 2008; Lago et al., 2011).

1.6.7. Methods not requiring post-PCR analysis

Since for many species complete sequence information has today become available, species may be directly identified through single-phase PCR techniques that exploit the detection capability of species-specific primers explicitly designed. Specific primers, under suitable stringent reaction conditions, generate a fragment only in the presence of *DNA* from a specified species (Ceruso et al., 2019). This possibility has the potential to produce considerable savings of time and effort within the laboratory. In fact, once amplified the *DNA*, species authentication can be simply carried out through visualization by agarose gel electrophoresis or by melting analysis in real time-PCR.

1.6.8. Multiplex PCR.

Multiplex PCR permits the simultaneous amplification of many targets in the same reaction and is considered one of the preferred "no-sequencing" techniques, due to its rapidity and simplicity of execution, considered an alternative method to be particularly used with screening purposes to minimize expenses and save time (Trotta et al., 2005). Target species *DNA* can in fact be analyzed in a single run by using a combination of species-specific primers resulting in *DNA* fragment lengths that vary with species. The length of the fragments can be predicted if the complete sequence is known, and a specified species can be identified by the appearance of an amplicon of appropriate size on an agarose gel. The main challenge in setting up a multiplex PCR proficient of specifically distinguishing between very closely-related species from a phylogenetic point of view is to design suitable primer (Castigliego et al. 2015). Initially, they obviously should be characterized, as also declared above, by a good level of specificity. The number of regions of a certain gene that diverges sufficiently among all the species to be able to distinguish between them is however limited. So, the possibility of alternatives is also limited, so that the higher the number of species, the lower the potential number of these alternatives. Therefore, the number of species included in the assay undoubtedly influences its efficiency, also because this technique is based on a delicate equilibrium among the species specific primers, and the presence of many pairs in the reaction increases the chance of obtaining

unspecific primer annealing on the sample *DNA*, spurious amplification products and may increase the possibility to obtain the formation of primer dimers (Giusti et al., 2016). Therefore, selecting primers with a low capability of hetero-dimerization is required and, if not possible, reducing the number of primers in the reaction mix. One option is to choose a common forward or reverse primer. Moreover, the utilization of primers with very similar optimum annealing temperatures is recommended (Castigliego et al., 2015). Additional fundamental step for obtaining a specific amplification, stringently connected with primers concentration, is the selection of an adequate *DNA* template concentration. In fact, if the primer-template ratio is too low, specific products will not accumulate exponentially while primer dimers may be amplified more efficiently than the desired target. Moreover, Taq polymerase, dNTPs and MgCl2 concentration should be opportunely evaluated, as well as the cycling condition (Giusti et al., 2016). Suitable controls should also be included to prevent the possibility of false positive or negative results being obtained (the lack of amplified fragment on the gel may be due to technical problems relatively than due to the absence of the target *DNA)* (Teletchea, 2009).

1.6.9. Real-time PCR (RT-PCR).

Real-time PCR (also known as quantitative PCR, Real-time quantitative PCR, or qPCR) is a method of simultaneous *DNA* amplification and detection (Teletchea, 2009). It is in fact an automated process, where no post-PCR processing is mandatory to analyze the amplification output. In this way, the probabilities of post-PCR contact contamination decrease, as it is possible to observe and analyze RT-PCR products without removing them from the instrument. This is attributable to the technique ability to detect, at every cycle of the PCR, the amount of PCR product (amplicon) using fluorescence (Salihah et al., 2016). A fluorescent reporter molecule is in fact comprised in the assay mix and monitored with an optical thermocycler that provides fluorescent excitation and quantification of the fluorescent emission. The fluorophores may be covalently linked to an oligonucleotide to form a labelled primer or probe or may be free molecules that bind to double stranded *DNA*. Many different designs are possible, the common feature being that they must display a change in fluorescence during PCR so that product accumulation can be monitored. A RT-PCR read-out is given as the number of PCR cycles ("cycle threshold" Ct) necessary to achieve a given level of fluorescence. The most

popular real-time PCR assay, so-called "TaqMan" approach, used in several seafood species identification studies (Taboada et al., 2017; Bajzik et al., 2012; Lopez & Pardo, 2005; Velasco et al., 2013) is based on the hybridization of a dual labelled probe to the PCR product, and the development of a signal by loss of fluorescence quenching as PCR degrades the probe (Ponchel et al., 2003). Another common approach is founded upon the binding of the fluorescent dye SYBRGreen I into the PCR product (PE Applied Biosystems, Warrington, UK) (Trotta et al., 2005; Chuang et al., 2012).

1.6.10. Next generation sequencing (NGS) technologies

Next-generation sequencing (NGS), also identified as high-throughput sequencing, is the catch-all term used to describe a few different modern sequencing technologies including:

- Illumina (Solexa) sequencing,
- Roche 454 sequencing,
- Ion torrent: Proton / PGM sequencing,
- SOLiD sequencing.

These recent technologies permit to sequence *DNA* and *RNA* much more quickly and cheaply than the earlier used Sanger sequencing, and as such have revolutionised the study of genomics and molecular biology. The biggest advantages are:

- ✓ offers a high-resolution, base-by-base view of the genome,
- ✓ captures together large and small variants that might otherwise be missed
- ✓ recognises potential causative variants for additional follow-on studies of gene expression and regulation mechanisms,
- ✓ carries large volumes of data in a short amount of time to sustenance assembly of novel genomes.

Present methods, which mostly count on on PCR amplification, are nowadays exploited for the analysis of an extremely wide range of seafood, from fresh to processed (marinated, salted, smoked, canned, frozen, etc.), are based on *DNA* Sanger sequencing. These methods, while reliable for the identification of single species, have been demonstrated poorly effective for the detection of species within mixed foods (Le Fresne et al., 2011), since Sanger method has been designed to produce a single sequence, generally from a single amplicon. On the other hand, even

though other methods not requiring sequencing step, as multiplex PCR or the more appreciated RT-PCR, allow species recognition in complex food matrices, the number of detectable species is often too exiguous. Furthermore, these techniques necessarily need the use of specific suitable primers, restraining the usefulness of the analysis if applied to products containing unknown species. This deficiency could represent a limit for food inspection activities aimed at maximizing consumer safeguard. In fact, the evermore-complex global seafood chain may act as a boost for fraudulent activities. Highly processed products composed by several species are those most at risk for mislabelling and may compromise consumer's safety if involving toxic or allergic species. For these reasons, Next Generation Sequencing (NGS) technologies could effectively become a turning point in the food inspection field, overcoming the limits of the standard analytical methods in the detection of multispecies matrices. NGS technologies (also known as "second generation sequencing technologies" or "high throughput sequencing technology") represent a fundamental shift away from the application of automated Sanger sequencing for *DNA* analysis (Metzker, 2010; Ceruso et al., 2019). In fact, massively parallel and clonal sequencing have increased the ability to increase sequence information even from a single molecule within a complex or degraded DNA source. NGS is already becoming a standard approach in many studies in many different fields, particularly for sequencing of genomes (Green et al., 2006; Miller et al., 2008; Mascolo et al., 2019; Ceruso et al., 2019) and metagenomic studies (Brandon-Mong et al., 2015; Hajibabaei et al., 2011; Yu et al., 2012). With NGS it is possible to identify species on a large scale, using a so-called metagenomics (also referred to as metabarcoding) approach (Staats et al., 2016). NGS power is increasingly favorably observed also for species identification in food, because it is possible to conduct a food metagenomic approach to detect species in highly processed products, without previous knowledge of which species are expected to be found. Despite the evident benefit that it could provide to food inspection, only few studies with this purpose have been conducted. A recent one, directed by Prosser and Hebert (2017), applied metabarcoding to the identification of the botanical and entomological sources of honey, since they indicate a substantial difference in the market value of the various product types. Preliminary studies were performed on artificial meat species mixture samples to verify the method's ability to detect all species within a mixture (Bertolini et al., 2015; Tillmar et al., 2013). The method was practically applied to

commercial products by the work of Muñoz-Colmenero et al. (2017), that detected the animal species contained in candies, selected as a model of highly processed foods. The research background results somewhat weak even for the seafood products analysis. Based on our knowledges, only one study applied NGS techniques on species identification within fish cake (Park et al., 2012). The first complete study comparing the different sequencing methods applied on fish *mt*DNA to identify a marker region suitable as barcoding in sparids, is the study conducted by our research group (Mascolo et al., 2019).

1.6.10.1. Analytical workflow of the most common NGS technologies (Illumina).

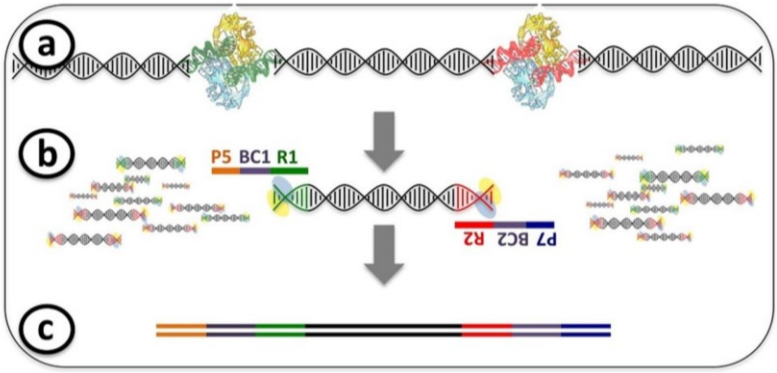
The definition "NGS technologies", although coupled with the idea of a system improving sequencing phase, is directed to define a more complex and integrated system, essentially characterized by three consecutive very efficient analytical steps, each of which must suitably work in order to allow an optimal global reaction output. These analytical steps are represented by the (1) library preparation, the (2) massive parallel *DNA* clonal amplification and the final (3) sequencing by synthesis. These three basic analytical steps are common to all the current NGS platform type, which are distinguishable from each other based on the technique they used to execute them. Workflow description of Illumina platform is reported below.

1.6.10.1a. Library preparation phase.

NGS firstly requires the conversion of the source genome into standard libraries suitable for loading onto a sequencing instrument. The techniques used for preparing libraries partly vary according to the type of the target (whole genome, different parts of genome simultaneously, single sequence fragment) but generally involves in casually breaking genomic *DNA* (chemically, mechanically or using primers amplifying a specific fragment) into smaller sizes from which either fragment templates or mate-pair templates are created. Then, randomly broken fragments are end-repaired and fused on both 3' and 5' end with adapters, which are chemically synthetized short oligonucleotides with known sequence that contain the essential elements for the immobilization of the template on a solid surface and sequencing (Van Dijk et al, 2014). Each platform has characteristic adaptors. Illumina adaptors are called P5 and P7. P5 comprehend a specific site able to subsequently hybridize with a

complementary one located on oligonucleotides covalently linked to the solid surface (Figue 1.11).

Figure 11 Illumina (Nextera) workflow display how two tagmentations are required to produce PCR amplifiable library molecules prepared for sequencing. (a) Transposomes fit in into genomic DNA. (b) Tagmentation produces amplifiable and non-amplifiable library molecules until transposomes run out. (c) The library is cleaned to remove Tn5 protiens bound to the ends of DNA fragments, then PCR-amplified to add flowcell compatible adapters and dual-indexes for multiplexed sequencing. Source by James Hadfield: Too good to be true?! What can Nextera do for you?



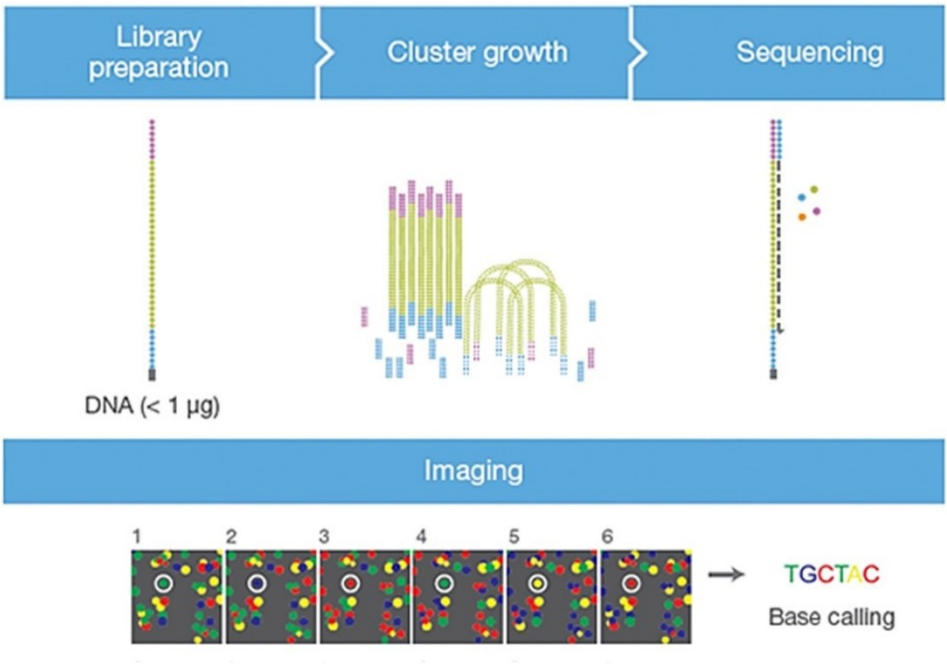
1.6.10.1b. Massive parallel DNA clonal amplification phase.

Before the sequencing step it is essential to product clusters, which are grouped of identical molecules spatially divided from each other, through a clonal PCR. The immobilization of spatially separated template positions permits thousands to billions of sequencing reactions to be performed concomitantly. The cluster generation methods, utilized in Illumina technology, is solid-phase amplification. It is also named "bridge amplification" and the clonal group of template *DNA* is bound to the surface of a flow cell. Each cluster is seeded by a single template *DNA* strand and clonally amplified through bridge amplification until the cluster has roughly 1.000 copies. Respectively cluster of the flow cell products a single sequencing read; for example, 10.000 clusters on the flow cell would produce 10.000 single reads and 20.000 paired ends reads (Figure 12).

1.6.10.1c. Sequencing by synthesis phase.

With NGS methods is possible to produce *DNA* in vitro through the utilization of *DNA* polymerase, dNTPs and suitable trigger primers for copying the *DNA* template. It is the possibly to detect the bases incorporated in the sequence during the amplification phase (real time detection) relatively than at the end such as occurred in Sanger method. Illumina technology is characterized by the so-called "cyclic reversible termination" that, as the name suggests, uses reversible terminators in a cyclic method that comprises nucleotide incorporation, fluorescence imaging and cleavage. In the first step, a *DNA* polymerase, bound to the primed template, adds or incorporates just one fluorescently modified dNTP (noteworthy, differently from the Sanger technique, the dNTPs are already fluorescent, permitting their real time detection by the system), which embodies the complement of the template base. After the addiction of the fluorescent complementary dNTP, a chemical reversible stop of *DNA* synthesis takes place, and imaging is then performed to determine the identity of the incorporated dNTP. Following incorporation, the remaining unincorporated nucleotides are washed away. This is followed by a cleavage step, which removes the terminating/inhibiting group and the fluorescent dye (Metzker, 2010). Additional washing is performed before starting the next incorporation step. The technological evolution of Illumina system is thus represented by the availability of extremely accurate optical techniques capable to quickly detect the fluorescent emission emanated from each base and consequently permit to exactly reconstruct the target sequence (Figure 12).

Figure. 12. The cut ends are repaired and adapters, indices, primer binding sites, and terminal sites are added to each strand of the DNA. The DNA attaches to the flow cell via complementary sequences. The strand bends over and attaches to a second oligo forming a bridge. A polymerase synthesizes the reverse strand. The two strands release and straighten. Each forms a new bridge (bridge amplification). The result is a cluster of DNAs forward and reverse strands clones.
Tagged nucleotides are added to the DNA strand. Each of the four nucleotides have a recognizing label that can be excited to emit a characteristic wavelength. A computer records all the emissions, and from this data, base calls are made. Source by Illumina (https://emea.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html?langsel=/it/).



References

Accum, F. (1820). A Treatise On Adulterations Of Food and Culinary Poisons, and Methods Of Detecting Them. *Evergreen Review*, Inc. New York.

Akazaki, M. (1962) Studies on Sparform Fishes-Anatomy, Phylogeny, Ecology and Taxonomy, p.8. Kosugi Co. Ltd, Osaka, Japan.

Allen, G. R. (1985). FAO species catalogue. v. 6: Snappers of the world. An annotated and illustrated catalogue of Lutjanid species known to date.

Antonucci, F., Costa, C., Aguzzi, J., & Cataudella, S. (2009). Ecomorphology of morpho-functional relationships in the family of sparidae: A quantitative statistic approach. *Journal of Morphology*, 270(7), 843-855.

Armani, A., Guardone, L., Castigliego, L., D'Amico, P., Messina, A., Malandra, R., …& Guidi, A. (2015a). DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. *Food Control*, 50, 589–596. 17)

Armani, A.; Castigliego, L.; & Guidi, A. (2012a). Fish frauds: the DNA challenge. *CAB Reviews* 7 (071), 112

Bajzik, P., Golian, J., Židek, R., Krall, M., Walczycka, M., & Tkaczewska, J. (2012). Identification of the Common Carp (Cyprinus carpio) Species Using Real-Time PCR Methods. *Żywność: nauka-technologia-jakość*, (5 (84)), 166-176.

Bargelloni, L, J A Alarcon, M C Alvarez, E Penzo, A Magoulas, C Reis, and T Patarnello. "Discord in the Family Sparidae (Teleostei): Divergent Phylogeographical Patterns across the Atlantic–Mediterranean Divide." *Journal of Evolutionary Biology* 16, no. 6 (2003): 1149–58.

Bartlett, S E, and W S Davidson. "FINS (Forensically Informative Nucleotide Sequencing): A Procedure for Identifying the Animal Origin of Biological Specimens." *Biotechniques* 12, no. 3 (1992): 408–11.

Bauchot, M. L., & Hureau, J. C. (1990). 1990. Sparidae. Check List of the Fishes of the Eastern Tropical Atlantic, *Clofeta* II, 790-812.

Bénard-Capelle, J., Guillonneau, V., Nouvian, C., Fournier, N., Le Loët, K., & Dettai, A. (2015). Fish mislabelling in France: substitution rates and retail types. *PeerJ*, 2, e714.

Bhattacharya, M., Sharma, A. R., Patra, B. C., Sharma, G., Seo, E.-M., Nam, J.-S., ..& Lee, S.-S. (2015). DNA barcoding to fishes: current status and future directions. *Mitochondrial DNA Part A*, 27(4), 2744-2752

Binladen, Jonas, M Thomas, P Gilbert, Paula F Campos, and Eske Willerslev. "5′-Tailed Sequencing Primers Improve Sequencing Quality of PCR Products." *Biotechniques 42*, no. 2 (2007): 174–76.

Blanco, M, Ricardo Isaac Pérez-Martín, and C G Sotelo. "Identification of Shark Species in Seafood Products by Forensically Informative Nucleotide Sequencing (FINS)." *Journal of Agricultural and Food Chemistry* 56, no. 21 (2008): 9868–74.

Buxton, C. D., & Garratt, P. A. (1990). Alternative reproductive styles in seabrearns (Pisces: Sparidae). In Alternative life-history styles of fishes (pp. 113-124). *Springer*, Dordrecht.

Buxton, C.D. & Garrat, P.A. (1990). Alternative reproductive styles in seabreams (Pisces: Sapridae). *Environmental Biology of Fishes*, 28, 113-124.

Carpenter, K. E. (1989). FAO species catalogue (Vol. 9). *Food & Agriculture Org*.

Carpenter, K. E., & Johnson, G. D. (2002). A phylogeny of sparoid fishes (Perciformes, Percoidei) based on morphology. *Ichthyological Research*, 49(2), 114-127.

Carvalho, D. C., Neto, D. A. P., Brasil, B. S. A. F., & Oliveira, D. A. A. (2011). DNA barcoding unveils a high rate of mislabeling in a commercial freshwater catfish from Brazil. *Mitochondrial DNA*, 22(1), 97–105.

Carvalho, D. C., Palhares, R. M., Drummond, M. G., & Frigo, T. B. (2015). DNA Barcoding identification of commercialized seafood in South Brazil: A governmental regulatory forensic program. *Food Control*, 50, 784–788.

Castigliego, L., Armani, A., Tinacci, L., Gianfaldoni, D., & Guidi, A. (2015). Two alternative multiplex PCRs for the identification of the seven species of anglerfish (Lophius spp.) using an end-point or a melting curve analysis real-time protocol. *Food Chemistry*, 166, 1–9.

Cawthorn, D. M., Steinman, H. A., & Witthuhn, R. C. (2012). DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International*, 46(1), 30–40.

Chapela, M. J., Sotelo, C. G., & Pérez-Martín, R. I. (2003). Molecular identification of cephalopod species by FINS and PCR-RFLP of a cytochrome b gene fragment. *European Food Research and Technology*, 217(6), 524-529.

Chiba, S.N., Iwatsuki, Y., Yoshino, T., & Hanzawa, N., (2009). Comprehensive phylogeeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes and Genetic Systems*, 84, 153-170.

Choat J.H. & Robertson D.R. (1975) Protogynous hermaphroditism in fishes of the family Scaridae. In: Intersexuality in the Animal Kingdom (ed. R. Reinboth), pp 263-283. Springer-Verlag, New York.

Chuang, P. S., Chen, M. I., & Shiao, J. C. (2012). Identification of tuna species by a real-time polymerase chain reaction technique. *Food Chemistry*, 133(3), 1055–1061.

Civera, T. (2003). Species Identification and Safety of Fish Products. *Veterinary Research Communication*, 27 (Suppl.1), 481-489.

Commission Regulation (EC) No 2065/2001 of 22 October 2001 laying down detailed rules for the application of Council Regulation (EC) No 104/2000 as regards informing consumers about fishery and aquaculture products. OJ L 278, 23.10.2001, p 6–8.

Council regulation (EC) No. 104/2000 of 17 December 1999 on the common organisation of the markets in fishery and aquaculture products. OJ L 17, 21.01.2000, p. 22-52.

Council Regulation (EEC) No 2658/87 of 23 July 1987 on the tariff and statistical nomenclature and on the Common Customs Tariff. OJ L 256, 7.9.1987, p. 1–675. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A31987R2658.

D'Amico, P., Armani, A., Gianfaldoni, D., & Guidi, A. (2016). New provisions for the labelling of fishery and aquaculture products: Difficulties in the implementation of Regulation (EU) n. 1379/2013. Marine Policy, 71, 147-156.

Day, J. J. (2000). Comparative morphology and evolutionary relationships of the Sparidae (Teleostei: Percoidei) (Doctoral dissertation, University College London (University of London).

Day, J. J. (2002). Phylogenetic relationships of the Sparidae (Teleostei: Percoidei) and implications for convergent trophic evolution. *Biological Journal of the Linnean Society*, 76(2), 269-301.

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters*, 10(9), 20140562.

Debode, F., Marien, A., Janssen, É., Bragard, C., & Berben, G. (2017). Influence of the amplicon length on real-time PCR results. *Biotechnologie, Agronomie, Société et Environnement*, 21(1), 3-11.

Decreto n. 19105, 22 settembre 2017. Attribuzione delle denominazioni in lingua italiana delle specie ittiche di interesse commerciale. GU Serie Generale n.266 del 14-11-2017.

Dieffenbach, C. W., Lowe, T. M. J., & Dveksler, G. S. (2008). General concepts for PCR primer design. General Concepts for PCR Primer Design. *PCR Methods and Applications*, 3(3),

Dudu, A., Georgescu, S. E., & Costache, M. (2015). Evaluation of Genetic Diversity in Fish Using Molecular Markers. *Molecular Approaches to Genetic Diversity*. https://doi.org/10.5772/60423

Duftner, N., Sefc, K. M., Koblmüller, S., Salzburger, W., Taborsky, M., & Sturmbauer, C. (2007). Parallel evolution of facial stripe patterns in the Neolamprologus brichardi/pulcher species complex endemic to Lake Tanganyika. *Molecular phylogenetics and evolution*, 45(2), 706-715.

Espiñeira, M., Gonzalez-Lavín, N., Vieites, J. M., & Santaclara, F. J. (2009). Development of a method for the identification of scombroid and common substitute species in seafood products by FINS. *Food Chemistry*, 117(4), 698-704.

Espiñeira, Montserrat, Nerea González-Lavín, Juan M Vieites, and Francisco J Santaclara. "Development of a Method for the Genetic Identification of Flatfish Species on the Basis of Mitochondrial DNA Sequences." *Journal of Agricultural and Food Chemistry* 56, no. 19 (2008): 8954–61.

EUMOFA (2018). The EU fish market 2018 edition. https://www.eumofa.eu/documents/20178/108446/The+EU+fish+market+2017.pdf/80acad95-907f-4b90-b2a7-1086964df3d9

European Commission (EC) 2016. Illegal fishing (IUU). In: Fisheries - European Commission [online]. https://ec.europa.eu/fisheries/cfp/illegal_fishing_en

European Parliament (EC). Resolution of 12 May 2016 on traceability of fishery and aquaculture products in restaurants and retail (2016/2532(RSP). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016IP0222

FAO & WHO. (2007). Protein and amino acid requirements in human nutrition: Report of a joint FAO/WHO/UNU expert consultation. WHO Technical Report Series 935.

FAO (2018). The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome. http://www.FAO.org/3/I9540EN/i9540en.pdf

FAO (2018a) ASFIS List of Species for Fishery Statistics Purposes. http://www.FAO.org/fishery/collection/asfis/en

FAO (2018b). FAO Major Fishing Areas [online]. http://www.FAO.org/fishery/area/search/en

FAO. (1981). Eastern central Atlantic: fishing area 34 and part of 47. Species identification sheet for fishery purposes. Vol I. pag 22.

FAO. (2007). The State of food and agriculture. www.FAO.org/3/a-a1200e.pdf

FAO. (2011-2019). Fisheries and aquaculture software. FishStatJ - software for fishery statistical time series. In: FAO Fisheries and Aquaculture Department [online]. Rome. Updated 21 July 2016.

Farell, E. M., & Alexandre, G. (2012). Bovine serum albumin further enhances the effects of organic solvents on increased yield of polymerase chain reaction of GC-rich templates. *BMC Research Notes*, 5, 1.

Garcia-Vazquez, E., Machado-Schiaffino, G., Campo, D., & Juanes, F. (2012). Species misidentification in mixed hake fisheries may lead to overexploitation and population bottlenecks. *Fisheries Research*, 114, 52–55.

Ghiselin M.T. (1969). The evolution of hermaphroditism among animals. Quarterly Review of Biology, 44, 189-208.

Giusti, A., Castigliego, L., Rubino, R., Gianfaldoni, D., Guidi, A., & Armani, A. (2016). A Conventional Multiplex PCR Assay for the Detection of Toxic Gemfish Species (Ruvettus pretiosus and Lepidocybium flavobrunneum): A Simple Method To Combat Health Frauds. *Journal of Agricultural and Food Chemistry*, 64(4), 960–968.

Gloerfelt-Tarp, T., & Kailola, P. J. (1984). Trawled Fishes of Southern Indonesia and Northwestern Australia. xvi+ 406 pp. Directorate General of Fisheries, (Jakarta) Indonesia and German Agency for Technical Assistance.

Griffiths, A. M., Sotelo, C. G., Mendes, R., Pérez-Martín, R. I., Schröder, U., Shorten, M., ... & Mariani, S. (2014). Current methods

for seafood authenticity testing in Europe: Is there a need for harmonisation?. *Food Control*, 45, 95-100.

Hall, Tom A. "BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT." In Nucleic Acids Symposium Series, 41:95–98. [London]: Information Retrieval Ltd., c1979-c2000., 1999.

Hanel, R., & Sturmbauer, C. (2000). Multiple recurrent evolution of trophic types in northeastern Atlantic and Mediterranean seabreams (Sparidae, Percoidei). *Journal of Molecular Evolution*, 50(3), 276-283.

Hanner, R. H., & Gregory, T. R. (2007). Genomic Diversity Research and the Role of Biorepositories. *Cell Preservation Technology*, 5(2), 93–103.

Hohenlohe, P. A., Amish, S. J., Catchen, J. M., Allendorf, F. W., & Luikart, G. (2011). Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, 11(1), 117–122.

Hudson, E. (1998, May). HELFMAN, GS, COLLETTE, BB FACEY, DE 1997. The diversity of fishes. 528 pp. Blackwell Science: Oxford. ISBN 0 86542 256 7 (hard cover). *In Animal Conservation forum* (Vol. 1, No. 2, pp. 151-151). Cambridge University Press.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 7(4), 544–548.

Johnson, D. G., & Patterson, C. (1993). Percomorph phylogeny: a survey of acanthomorphs and a new proposal. *Bulletin of Marine Science*, 52(1), 554-626.

Johnson, R. (2014). Food Fraud and Economically Motivated Adulteration of Food and Food Ingredients. Available at: http://www.fredsakademiet.dk/ORDBOG/lord/food_fraud.pdf

Jordan, D. S., & Fesler, B. (1893). A Review of the Sparoïd Fishes of America and Europe.

Kwok, S., Kellogg, D., McKinney, N., & Spasic, D. (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, 18(4), 999–1005. S30-S37.

Lago, F. C., Herrero, B., Vieites, J. M., & Espineira, M. (2011). Genetic identification of horse mackerel and related species in seafood

products by means of forensically informative nucleotide sequencing methodology. *Journal of agricultural and food chemistry*, 59(6), 2223-2228.

Lang, M., & Orgogozo, V. (2011). Identification of homologous gene sequences by PCR with degenerate primers. *Molecular Methods for Evolutionary Genetics*, 245-256.

Le Fresne, S., Popova, M., Le Vacon, F., & Carton, T. (2011). Application of Denaturing HighPerformance Liquid Chromatography (DHPLC) for the Identification of Fish: A New Way To Determine the Composition of Processed Food Containing Multiple Species. *Journal of Agricultural and Food Chemistry*, 59(23), 12302–12

Le Gall, J. (1929). Note sur la sexualitd de la dorade (Pagellus centrodontus). *Revue Trav. Off. Pech. Marit*, 2, 31-32.

Lockley, A K, and R G Bardsley. "DNA-Based Methods for Food Authentication." *Trends in Food Science & Technology* 11, no. 2 (2000): 67–77.

Lopez, I., & Pardo, M. A. (2005). Application of Relative Quantification TaqMan Real-Time Polymerase Chain Reaction Technology for the Identification and Quantification of Thunnus alalunga and Thunnus albacares. *Journal of Agricultural and Food Chemistry*, 53(11), 4554–4560.

Lorenz, T. C. (2011). Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *Journal of visualized experiments*: JoVE, (63), e3998-e3998. on Quantitative Biology, 51(1), 263–273

Lundstrom, R. C. (1980). Fish species identification by thin layer polyacrylamide gel isoelectric focusing: collaborative study. *Journal-Association of Official Analytical Chemists*, 63(1), 69-73.

Malandra, R., & Renon, P. (1998). Le principali frodi dei prodotti della pesca. VLU-Libreria Universitaria Multimediale.

Manzoni, P. (1993). Enciclopedia illustrata delle specie ittiche marine di interesse commerciale aventi denominazione stabilita dalla normativa italiana.

Martinez, I., James, D., & Loréal, H. (2005). Application of modern analytical techniques to ensure seafood safety and authenticity. FAO. Fisheries Technical Paper. Available at: https://doi.org/10.1007/s00217-003-0743-y

Mascolo, C., Ceruso, M., Sordino, P., Palma G., Anastasio, A., & Pepe, T. (2017). Development of a method to extract and amplify the complete mitogenome of some Sparidae species. *Italian Journal of Food Safety*, 6, 6154.

Messing, Joachim, and Jeffrey Vieira. "A New Pair of M13 Vectors for Selecting Either DNA Strand of Double-Digest Restriction Fragments." *Gene* 19, no. 3 (1982): 269–76.

Metzker, M. L. (2010). Sequencing technologies - the next generation. Nature Reviews. *Genetics*, 11(1), 31–46. 308.

Miya, M., Kawaguchi, A., Nishida, M. (2001). Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, 18, 1993-2009. https://doi.org/10.1093/oxfordjournals.molbev.a003741

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Kondoh, M. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society open science*, 2(7), 150088.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and Application of a Database of Food Ingredient Fraud and Economically Motivated Adulteration from 1980 to 2010. *Journal of Food Science*, 77(4), R118-R126.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. *Cold Spring Harbor Symposia*.

Munro, I.S.R. (1949) Revision of Australian silver breams Mylio and Rhabdosargus. *Memories of the Queensland museum*, 12 (4), 182-223.

NSF International. (2013). The Growing Global Epidemic of Food Fraud. https://d6prv7be4nrvy.cloudfront.net/wp-content/uploads/NSF-Whitepaper.pdf

Pardo, M., Jiménez, E., & Pérez-Villarreal, B. (2016). Misdescription incidents in seafood sector. *Food Control*, 62, 277–283.

Pepe T, Trotta, M, Di Marco I, Cennamo P, Anastasio A, Cortesi ML, 2005. Mitochondrial cytochrome b DNA sequenze variations: An approach to fish species identification in processed fish products. *J Food Protect* 68:421-5.

Preedy, V. R. (Ed.). (2014). Processing and impact on active components in food. Academic Press

Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-based methods for the identification of commercial fish and seafood species. *Comprehensive Reviews in Food Science and Food Safety,* 7(3), 280–295. 254)

Rasmussen, R. S., & Morrissey, M. T. (2009). "DNA-based Methods for the Identification of Commercial Fish and Seafood Species." *Comprehensive Reviews in Food Science and Food Safety* 7, no. 3 (2008): 280–95.

Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 october 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004. OJ L 304, 22.11.2011, p. 18-63. 260)

Regulation (EU) No 1379/2013 of the European Parliament and of the Council of 11 December 2013 on the common organisation of the markets in fishery and aquaculture products, amending Council Regulations (EC) No 1184/2006 and (EC) No 1224/2009 and repealing Council Regulation (EC) No 104/2000. OJ L 354, 28.12.2013, p. 1–21.

Rubinoff, D., Cameron, S. & Will, K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity*, 97:581-94. https://doi.org/10.1093/jhered/esl036

Saitou, Naruya, and Masatoshi Nei. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution* 4, no. 4 (1987): 406–25.

Salihah, N. T., Hossain, M. M., Lubis, H., & Ahmed, M. U. (2016). Trends and advances in food analysis by real-time polymerase chain reaction. *Journal of food science and technology*, 53(5), 2196-2209.

Sanger, Frederick, Steven Nicklen, and Alan R Coulson. "DNA Sequencing with Chain-Terminating Inhibitors." *Proceedings of the National Academy of Sciences* 74, no. 12 (1977): 5463–67.

Shames, L. (2009). Seafood fraud: FDA program changes and better collaboration among key federal agencies could improve detection and prevention. DIANE Publishing.

Shapiro, D.Y. (1987). Differentiation and evolution of sex change in fishes. *Bioscienc*e, 37, 490-496.

Sotelo, Carmen G, Pilar Calo-Mata, Ma Jose Chapela, Ricardo I Pérez-Martín, Harmut Rehbein, Georgina L Hold, Valerie J Russell, Susan Pryde, Javier Quinteiro, and Mónica Izquierdo. "Identification of Flatfish (Pleuronectiforme) Species Using DNA-Based Techniques." *Journal of Agricultural and Food Chemistry* 49, no. 10 (2001): 4562–69..

Spink, J. (2007). Global counterfeit food and beverage packaging: Impacts on food safety. In Association of Food and Drug Officials (AFDO), Annual Conference. Available from: http://www. afdo. org/afdo/Conferences/upload/070619-Food-1600.

Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of Food Science*, 76(9), R157-R163.

Spink, J., Moyer, D. C., Park, H., & Heinonen, J. A. (2013). Defining the types of counterfeiters, counterfeiting, and offender organizations. *Crime Science*, 2(1), 8.

Spink, J., Moyer, D. C., Park, H., & Heinonen, J. A. (2014). Development of a product-counterfeiting incident cluster tool. *Crime Science*, 3(1), 3.

Steffens, D L, S L Sutter, and S C Roemer. "An Alternate Universal Forward Primer for Improved Automated DNA Sequencing of M13." *Biotechniques* 15, no. 4 (1993): 580–82.

Taboada, L., Sánchez, A., & Sotelo, C. G. (2017). A new real-time PCR method for rapid and specific detection of ling (Molva molva). *Food chemistry*, 228, 469-475.

Tacon, A. G. J., & Metian, M. (2013). Fish Matters: Importance of Aquatic Foods in Human Nutrition and Global Food Supply. *Reviews in Fisheries Science*, 21(1), 22–38.

Tamura, Koichiro, and Masatoshi Nei. "Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees." *Molecular Biology and Evolution* 10, no. 3 (1993): 512–26.

Teletchea F, Maudet C, Hanni C. (2005). Food and forensic molecular identification: update and challenges. Trends Biotechnol 23 (7), 359–66.

Teletchea F., (2009). Molecular identification methods of fish species: reassessment and possible applications. *Fish Biology and Fisheries*;19 (3), 265– 93.

Terio, V.; Di Pinto, P.; Decaro, N.; Parisi, A.; Desario, C.; Martella, V.; Buonavoglia, C. & Tantillo, M.G. (2010). Identification of tuna species in commercial cans by minor groove binder probe real-time polymerase chain reaction analysis of mitochondrial DNA sequences. *Molecular and Cellular Probes*, 24, 352-6.

Tortonese, E. (1975). Fauna d'Italia,'Osteichtyes': pesci ossei. Calderini Bologna, 11, 133-139.

Trotta, M., Schönhuth, S., Pepe, T., Cortesi, M. L., Puyet, A., Bautista, J. M. (2005). Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (Epinephelus and Mycteroperca species) and common substitute species. *Journal of Agricultural and Food Chemistry*, 53, 2039-45.

Upton, H. F. (2015). Seafood Fraud. Available at: http://nationalaglawcenter.org/wpcontent/uploads/assets/crs/RL34124.pdf.

Velasco, A., Sánchez, A., Martínez, I., Santaclara, F. J., Pérez-Martín, R. I., & Sotelo, C. G. (2013). Development of a real-time PCR method for the identification of Atlantic mackerel (Scomber scombrus). *Food Chemistry*, 141(3), 2006–10.

Ward, R. D., Hanner, R., Hebert, P. D. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, 74:329-56.

Warner, K., Lowell, B., Geren, S., & Talmage, S. (2016). Deceptive Dishes : Seafood Swaps Found Worldwide, Available at: http://usa.oceana.org/publications/reports/deceptive-dishes-seafoodswaps-found-worldwide.

Warner, K., Timme, W., Lowell, B., & Hirshfield, M. (2013). Oceana study reveals seafood fraud nationwide. *Oceana*, 11, 1–69.

Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, 29(2), 188-197.

Williams, G. C. (2018). Adaptation and natural selection: A critique of some current evolutionary thought (Vol. 61). Princeton university press.

Williamson, H. C. (1911). Report on the Reproductive Organs of Sparus Centrodontus, Delaroche; Sparus Cantharus L., Sebastes Marinus (L.), and Sebastes Dactylopterus (Delaroche); and on the Ripe Eggs

and Larvae of Sparus Centrodontus and Sebastes Marinus (No. 1). *HM Stationery Office*.

Xiong, X., Guardone, L., Cornax, M. J., Tinacci, L., Guidi, A., Gianfaldoni, D., & Armani, A. (2016b). DNA barcoding reveals substitution of Sablefish (Anoplopoma fimbria) with Patagonian and Antarctic Toothfish (Dissostichus eleginoides and Dissostichus mawsoni) in online market in China: How mislabeling opens door to IUU fishing. *Food Control*, 70, 380–

Research aims

The present work aims to:

- improve the techniques currently used for *mt*DNA isolation from fish muscle and develop a rapid, simple, and efficient method to extract and sequence it;
- apply the selected method for *mt*DNA isolation and sequencing from fish muscle to obtain the complete *mt*DNA sequence of some species of the genus *Dentex*, *Pagellus* and *Pagrus*;
- study and analyze the complete *mt*DNA sequence with the aim to identify mitochondrial markers or multiple marker approaches with higher and more specie-specific discrimination capacity;
- identify a *mt*DNA gene with variable sequence useful for sparids species discrimination, providing a rapid and simple tool that could be useful for fish industry.

In particular, the projects have been divided in nine chapters, on the basis of the research steps addressed:

Chapter 1 Development of a method to extract and amplify the complete mithogenome of some *Sparidae* species. Mascolo, C., Ceruso, M., Sordino, P., Palma, G., Anastasio, A., Pepe, T. (2017). *Italian Journal of Food Safety*. 6:6154 pag.75-78. Doi: 10.4081/ijfs.2017.6154

Chapter 2 Comparison of mitochondrial DNA enrichment and sequencing methods from fish tissue. Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2019). Comparison of mitochondrial DNA enrichment and sequencing methods from fish tissue. *Food Chemistry,* accepted for pubblication.

Chapter 3 The complete mitochondrial genome of the Pink dentex Dentex gibbosus (Perciformes: *Sparid*ae). Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., & Sordino, P. (2018). The complete mitochondrial genome of the Pink dentex *Dentex gibbosus* (Perciformes: *Sparidae*). *Mitochondrial DNA Part B*, *3*, 525–526. Doi: 10.1080/23802359.2018.1467230

Chapter 4 The complete mitochondrial genome of the common pandora *Pagellus erythrinus* (Perciformes: *Sparidae*). Ceruso, M., Mascolo, C., Lowe, E. K., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2018b). The complete mitochondrial genome of the Common Pandora *Pagellus erythrinus* (Perciformes: *Sparidae*), Mitochondrial DNA Part B, 3:2, 624-625. https://doi.org/10.1080/23802359.2018.1467235

Chapter 5 The complete mitochondrial genome of the Common dentex, *Dentex dentex* (Perciformes: *Sparidae*). Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. (2018). *Mitochondrial DNA Part B*. 3: 391-392. Doi: 10.1080/23802359.2018.1450675

Chapter 6 The complete mitochondrial genome of the Axillary seabream, *Pagellus acarne* (Perciformes: *Sparidae*). Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Sordino, P., Pepe, T. (2018 a). *Mitochondrial DNA Part B*, 3, 434-435. Doi: 10.1080/23802359.2018.1450674

Chapter 7 The complete mitochondrial genome of the Angolan dentex *Dentex angolensis* (Perciformes: Sparidae). Mascolo C., Ceruso M., Chirollo C., Palma G., Anastasio A., Sordino P. and Pepe T. (2019). *Mitochondrial DNA Part B*. Doi: 10.1080/23802359.2019.1591248

Chapter 8 The complete mitochondrial genome of the Red porgy *Pagrus pagrus* (Perciformes: *Sparidae*), preliminary results.

Chapter 9 Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers. Ceruso, M., Mascolo C., Anastasio, A., Pepe, T., Sordino, P. *Food Control.* In press.

Chapter 1

Development of a method to extract and amplify the complete mithogenome of some *Sparidae* species.

Mascolo, C., Ceruso, M., Sordino, P., Palma, G., Anastasio, A., Pepe, T. (2017). *Italian Journal of Food Safety*. 6:6154 pag.75-78. Doi: 10.4081/ijfs.2017.6154

Development of a method to extract and amplify the complete mithogenome of some *Sparidae* species.

1.1.Abstract

Previous studies showed that fish mitochondrial DNA (*mt*DNA) is set up by a closed circular molecule of 16-17 kilobases (kb), comprising 2 ribosomal RNA genes (rRNA), 22 transfer RNA genes (tRNA), 13 protein-coding genes and 2 non-coding regions. The analysis of single *mt*DNA genes, such as *Cytb, COI, 16S* and *12S,* or short segment of them, has been widely used against species substitution in both fresh and processed fish products. The analysis of the complete mitochondrial genome of fishery products allows to better study and characterise fish species. The aim of this research was to extract and amplify the complete *mt*DNA of some fish species of commercial interest belonging to the *Sparidae* family. The studied species were *Dentex dentex*, *Dentex gibbosus*, *Dentex nufar*, *Pagellus acarne* and *Pagellus erythrinus*. The entire mitogenome was obtained by gene amplification using long polymerase chain reactions. The analysis of the complete mitochondrial sequences will allow to gain further insights on these species and to find polymorphic sites that assess the degree of genetic variability of the species belonging to the family *Sparidae*.

1.2. Introduction

The molecular characterisation of the entire mitochondrial genome provides complete genetic information for phylogenetic analysis of organisms. The complete nucleotide sequence of fish mitochondrial genomes was determined from a growing number of species (Miya et al., 2003). Fish mitochondrial DNA (*mt*DNA) is a circular molecule of 16-17 kilobases (kb) kb in length, normally consisting of 2 ribosomal RNA genes (rRNA), 22 transfer RNA genes (tRNA), 13 protein-coding genes and and 2 non-coding regions (Shi et al., 2012). The study of *mt*DNA has become a very useful approach in population genetics and evolutionary studies (Manchado et al., 2004) and is used as marker to detect fraudulent substitutions in prepared and transformed fish products (Pepe et al., 2005, 2007); the nucleotide sequences that are fragments belonging to the genes *cytochrome b* (*cyt b*) genes, *ribosomal 16S* and *12S subunits*, and *cytochrome c oxidase subunit 1* (*COI*) (Espiñeira et al., 2008; Hubalkova et al., 2008; Zhang and Hanner, 2012; Chin et al., 2016). However, the use of short segments of the *mt*DNA (~100-700 base pairs) may give ambiguous results, because the fragments are too short to contain sufficient genetic information, and the variations among species are represented by few polymorphisms expressed by point

mutations (Bottero et al., 2007). It appears that more mitochondrial genomic information is needed to highlight the presence of more variable regions within the species. The *Sparidae* family is one of the most valuable and popular fish resources in the world and comprises about 41 species of different commercial value (D.M. MIPAAF, 31 January 2008; Italian Republic, 2008). Species substitution is very common in prepared and processed fish products, due to the profits resulting from the placing on the market of less expensive species. At the present time, nine *Sparidae* complete mitochondrial genome sequences are available in GenBank (*Acanthopagrus latus, Acanthopagrus schlegelii, Dentex tumifrons, Pagellus bogaraveo, Pagrus auriga, Pagrus major, Parargyrops edita, Rhabdosargus sarba, Sparus aurata*). Increase the mitochondrial genomic data on the others *Sparidae* species appear of great interest. The extraction and the amplification of the mitochondrial genome is a primary key to correctly continue the study of mitochondrial DNA through the comparison and analysis of properly obtained sequences. The aim of this research was to find a useful method for the extraction and amplification of the complete *mt*DNA of five fish species of commercial interest belonging to the *Sparidae* family, with the future aim to analyze and compare them, increasing our knowledge with regard to *Sparidae* mitogenomics. The *mt*DNA was isolated using long-polymerase chain reaction (PCR). The long PCR method was selected because represents a major advance for the high-yield purification of *mt*DNA (Yamauchi et al., 2004) and is one of the most efficient ways to isolate and successfully sequence the ntire mitogenome of fishes (Miya and Nishida, 1999).

1.3. Materials and Methods

A total of five different *Sparidae* species were tested. Here, the species *Dentex dentex, Dentex gibbosus, Dentex nufar, Pagellus acarne* and *Pagellus erythrinus* were analyzed. The whole specimens were identified, according to their anatomical and morphological features, as belonging to *D. dentex, D. gibbosus, D. nufar, P. acarne* and *P. erythrinus* species at the Department of Veterinary Medicine and Animal Production, University Federico II, Naples. *D. dentex* specimen was fished in Adriatic Sea (near Vieste). *D. nufar, P. acarne* and *P. erythrinus* specimens were supplied at Salerno fish market. *D. gibbosus* specimen was collected at Pozzuoli fish market. Fish were frozen on board at -20°C and shipped in insulated boxes to the laboratory. The tissues sampled from each specimen were: tongue

muscle, dorsal fin, skeletal muscle, caudal fin and liver. Total DNA was extracted from each sampled tissue, using the following methods: DNeasy tissue kit (Qiagen, Hilden, Germany) and NucleoSpin tissue kit (Macherey-Nagel, Düren, Germany), both according to the manufacturer's instructions, CTAB based method (Doyle and Doyle, 1987), salting-out method (Martínez et al., 1998). Extracted DNA was quantified using NanoDrop ND-2000C (Thermo Scientific, Waltham, MA, USA). DNA concentration was expected to be between ~35 and 200 ng/mL and the purity of DNA in the range of 1.8-2.0 ratio of absorbance wavelength A260/A280. Electrophoretic analysis was also done using 1% agarose gel to examine the degree of degradation or the extracted DNA. For the long PCR (Miya et al., 2001), the primers that were used are shown in Table 1.1. These primers were chosen after multiple alignment (Figure 1.1) using BioEdit Sequence Alignment Editor (Hall, 1999) of *Sparidae* complete mitochondrial genome sequences available in GenBank: *Acanthopagrus latus* (NC_010977.1, Xia et al., 2008), *Acanthopagrus schlegelii* (JQ746035.1, Shi et al., 2012), *Dentex tumifrons* (NC_029479.1, Zeng et al., unpublished), *Pagellus bogaraveo* (NC_009502.1, Ponce et al., 2008), *Pagrus major* (NC003196.1, Miya et al., 2001), *Pagrus auriga* (NC005146.1, Ponce et al., unpublished), *Parargyrops edita* (EF107158.1, Xia et al., 2007), *Rhabdosargus sarba* (KM272585.1,Li et al., 2016), *Sparus aurata* (LK022698.1, Dray et al., 2014). The primers were used to amplify the complete mitochondrial genomes in two long PCRs (two-step strategy) (Figure 1.2). Long PCRs (final volume=50 µL) were performed in a PTC-100 thermal cycler (MJ Research) and standardized as follows: 29.75 µL sterile distilled H2O, 10 µL Q5 PCR buffer (NEB), 5 µL dNTP (2 mM), 0.75 µL forward primer (50 pmol/µL), 0,5 µL reverse primer, 1 µL of 2000 U/mL Q5 High-Fidelity Taq polymerase (NEB), and 3 µL (100 ng) of DNA template. The thermal cycle profile is that of *shuttle* PCR: denaturation at 98°C for 10 seconds, with annealing and extension combined at the same temperature (72°C) for 9 minutes and 20 seconds. Double-stranded PCR products were purified using High Pure PCR Product Purification Kit (Roche, Basilea, Switzerland).

Table 1.1. Sequences of selected long polymerase chain reaction primers**.**

N°	Primer	5' ⟶ 3'	Tm
FW 6	S-LA-16SH R	GATGTTGGATCAGGACATCCYAATGGTGCA	70,2 °C
FW 5	L12321-LEU	GGTCTTAGGAACCAAAAACTCTTGGTGCAA	72,8 °C
REV 2	S-LA-16SH	TGCACCATTRGGATGTCCTGATCCAACATC	70,2 °C
REV 8	L12321-LEU R	TTGCACCAAGAGTTTTTGGTTCCTAAGACC	72,8 °C

Figure 1.1. Alignment of forward primer sequences with their respective annealing sites on the mitochondrial genome of 9 *Sparidae* species. The primer order was 5 −3. Dots indicate identical sites, dashes indicate gaps and divergent sites are indicated by the corresponding nucleotide. Reverse primers (S-LA-16SH, annealing site at 3790 bp; L12321LEU R, annealing site at 13282 bp) were obtained from the reversed and complement sequences of forward primers.

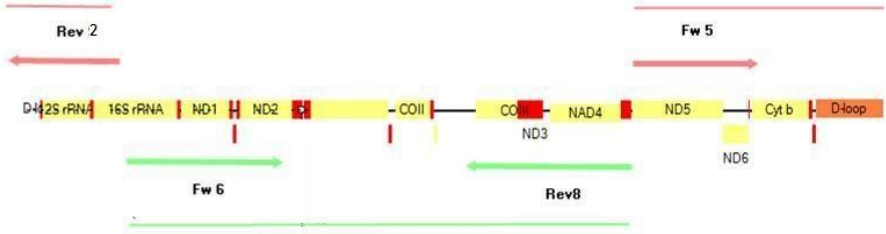
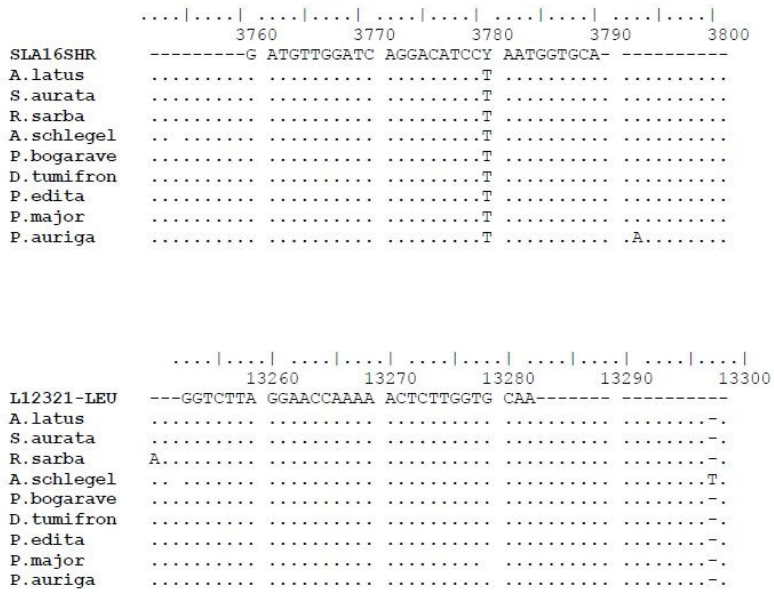


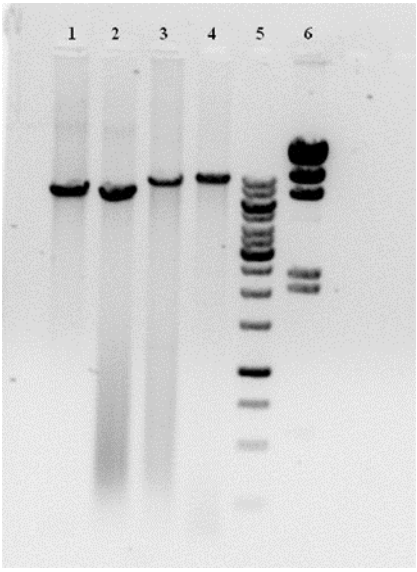
Figure 1.2. Schematic model of the structural organisation of the *Sparidae* (e.g. Pagrus major NC003196.1) mitochondrial genome. Primer annealing sites are indicated by green arrows (~9.5 kb amplicon) and pink arrows (~8 kb amplicon).



1.4. Results

DNA was correctly extracted from all examined samples. The best quality and quantity of DNA was obtained from the dorsal fin using DNeasy tissue kit (Qiagen). The long PCR approach allowed to obtain the complete mitochondrial genome of *D. dentex, D. gibbosus, D. nufar, P. acarne* and *P. erythrinus* in two well-resolved amplicons ranging of ~8 and ~9.5 kb (amplicon A: L12321-Leu/ S-LA-16SH, and amplicon B: S-LA-16SH R/ L12321-LeuR, respectively). These two fragments were complementary to each other. High yields of specific long-PCR amplicons were obtained for the tested species *D. gibbosus* and *P. erythrinus*, without any further optimization of the long PCR conditions (Figure 1.3). Weakness in the reproducibility of amplification was found in the other three species, in particular for the longer (~ 9.5 kb) amplicon.

Figure 1.3. Long polymerase chain reaction products. Lane 1 - ~ 8 kb amplicon of Dentex gibbosus; lane 2 - ~ 8 kb amplicon of Pagellus erythrinus; lane 3 - ~ 9.5 kb amplicon of Dentex gibbosus; lane 4 - ~ 9.5 kb amplicon of Pagellus erythrinus; lane 5 - GeneRuler 1 kb DNA Ladder (ThermoFisher Scientific, Waltham, MA, USA); lane 6 - Lambda DNA/HindIII marker (ThermoFisher Scientific).



1.5. Discussion

Preliminary results of this study allowed suggesting a useful method for the extraction and amplification of the complete mitogenome of five species belonging to the *Sparidae* family. The best results were obtained for *Dentex gibbosus* and *Pagellus erythrinus*. The amplification was also possible for the other three species *Dentex dentex, Dentex nufar* and *Pagellus acarne* with a still lower degree of reproducibility. For these three species, we are testing different primer set amplifying for smaller fragments designed to facilitate the amplification and the sequencing steps and sets in order to increase the level of reproducibility. The mitogenome amplification is a primary key to correctly continue the study of mitochondrial DNA. The next step of this study will include the sequencing and analysis of the genetic variability of the sequenced mitogenomes. The complete *mt*DNA is able to provide important information that may highlight the presence of any

variable regions within the species, with the aim to design primer sets that are able to amplify species-specific fragments.

1.6. Conclusions

This technique will allow a rapid species identification using a single PCR reaction, providing the basis for the molecular traceability of the fish products, in agreement with the provisions of Regulation (EU) 1379/2013 (European Commission, 2013).

References

D.M. MIPAAF, 31 Gennaio 2008

Bottero, M.T., Dalmasso, A., Cappelletti, M., Secchi, C., Civera, T. (2007). Differentiation of five tuna species by a multiplex primer-extension assay. J Biotechnol. 129(3):575-80.

Chin T.C., Adibah A.B., Danial Hariz Z.A., Siti Azizah M.N. (2016). Detection of mislabelled seafood products in Malaysia by DNA barcoding: improving transparency in food market. Food Control. 64:247-56.

Dray, L.M., Neuhof, M., Diamant, A., Huchon, D. (2016). The complete mitochondrial genome of the gilthead seabream Sparus aurata L. (Sparidae). Mitochondrial DNA. 27(1):781-2.

Espiñeira, M., González-Lavín, N., Vieites, J.M., Santaclara, F.J. (2008). Authentication of Anglerfish Species (Lophius spp) by Means of Polymerase Chain ReactionRestriction Fragment Length Polymorphism (PCR-RFLP) and Forensically Informative Nucleotide Sequencing (FINS) Methodologies. J Agric Food Chem. 56(22):10594-9.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symposium Series. 41:95-98.

Hubalkova, Z., Kralik, P., Kasalova, J., Rencova, E. (2008). Identification of gadoid species in fish meat by polymerase chain reaction (PCR) on genomic DNA. J Agric Food Chem. 56(10):3454-9.

Li, J., Yang, H., Xie, Z., Yang, X., Xiao, L., Wang, X., Li, S., Chen, M., Zhao, H., Zhang, Y. (2016). The complete mitochondrial genome of the Rhabdosargus sarba (Perciformes: Sparidae). Mitochondrial DNA A DNA MappSeq Anal. 27(3):1606-7.

Manchado M, Catanese G, Infante C. (2004). Complete mitochondrial DNA sequence of the Atlantic bluefin tuna Thunnus thynnus. Fish Sci. 70:68-73.

Miya, M., Nishida, M. (1999). Organization of the Mitochondrial Genome of a Deep-Sea Fish, Gonostoma gracile (Teleostei: Stomiiformes): First Example of Transfer RNA Gene Rearrangements in Bony Fishes. Mar Biotechnol (NY). 1(5):416-26.

Miya, M., Kawaguchi, A., Nishida, M. (2001). Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol Biol Evol. 18(11):1993-2009.

Miya, M., Takeshima, H., Endo, H., Ishiguro, N.B., Inoue, J.G., Mukai, T., Satoh, T.P., Yamaguchi, M., Kawaguchi, A., Mabuchi, K., Shirai, S.M.,

Nishida, M. (2003). Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. Mol Phylogenet Evol. 26:121-138.

Pepe, T., Trotta, M., di Marco, I., Cennamo, P., Anastasio, A., Cortesi, M. L. (2005). Mitochondrial cytochrome b DNA sequenze variations: An approach to fish species identification in processed fish products. J Food Prot. 68:421-425.

Pepe, T., Trotta, M., Di Marco, I., Anastasio, A., Bautista, J.M., Cortesi, M. L. (2007). Fish

species identification in surimi-based products. J Agric Food Chem. 55(9):3681-5.

Ponce, M., Infante, C., Jimenez-Cantizano, R.M., Perez, L., Manchado, M. (2008). Complete mitochondrial genome of the blackspot seabream, Pagellus bogaraveo (Perciformes: Sparidae), with high levels of length heteroplasmy in the WANCY region

Gene. 409(1-2):44-52.

Ponce, M., Infante, C., Catanese, G., Cardenas, S., Manchado, M. Complete mitochondrial DNA of redbanded seabream (Pagrus auriga).

Unpublished.

Shi X, Su Y, Wang J, Ding S, Mao Y. (2012). Complete mitochondrial genome of black porgy Acanthopagrus schlegelii (Perciformes, Sparidae). Mitochondrial DNA. 23(4):310-2..

Xia, J., Xia, K., Gong, J., Jiang, S. (2007). Complete mitochondrial DNA sequence, gene organization and genetic variation of control regions in Parargyrops edita. Fish Sci. 73(5):1042-1049.

Xia, J., Xia, K., Jiang, S. (2008). Complete mitochondrial DNA sequence of the yellowfin seabream Acanthopagrus latus and a genomic comparison among closely related sparid species. DNA Seq. 19(4):385-393.

Yamauchi, M.M., Miya, M.U., Machida, R.J., Nishida, M. (2004). PCR-based approach for sequencing mitochondrial genomes of decapod crustaceans, with a practical example from kuruma prawn (Marsupenaeus japonicus). Mar Biotechnol (NY). 6(5):419-29.

Zhang, J., Hanner, R. (2012). Molecular approach to the identification of fish in the South China Sea. PLoS One. 7(2):e30621.

Zeng, L., Jiang, L., Wu, C. Complete mitochondrial genome of yellowback seabream (Evynnis tumifrons). Unpublished

Chapter 2

Comparison of mitochondrial *DNA* enrichment and sequencing methods from fish tissue

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2019). Comparison of mitochondrial DNA enrichment and sequencing methods from fish tissue. *Food Chemistry* accepted for pubblication.

2.1.Abstract

Sparid fish species have different commercial value related to their organoleptic features. The mitochondrial (*mt*) DNA provides potential tools to distinguish at the species level. The enrichment of high-quality *mt*DNA from total genomic DNA is critical to obtain entire *mt*DNA sequences. Conventional *mt*DNA isolation is relatively low-cost and proficient. However, high numbers of PCR cycles can lead to artefacts (10^{-6} mutations/bp). We describe a rapid protocol for *mt*DNA extraction and enrichment from fish tissues, based on conventional miniprep columns, without the need to employ PCR amplification, and paramagnetic bead-based purification. This newly described method generates a substrate for next-generation sequencing (NGS) analysis and is likely to have wider applications for mitochondrial studies on other fish families and help ensure traceability and differentiation of fish with high commercial value.

2.2. Introduction

Among research studies performed on fish mitochondrial DNA (*mt*DNA or mitogenome), those related to species identification are of great importance for food authentication and safety. Species substitution is very common in fish products, due to potential profits resulting from the misuse of less expensive species to replace more expensive products (Marko et al., 2004). Currently, nucleotide sequences belonging to mitochondrial genes such as *Cytb, COI1, 16S* and *12S,* are used widely as markers to identify fish species and possible fraudulent substitutions (http://www.fishbol.org/) (Pepe et al., 2007; Espiñeira et al., 2008; Cawthorn et al., 2012a; Cawthorn et al., 2012b; Shokralla et al., 2015). However, a short nucleotide segment can contain insufficient genetic characteristics to generate a signal differentiating species accurately, especially where they are closely related (Paracchini et al., 2017). Also, some mitochondrial target regions perform well for certain species, but are less discriminating for others and this phenomenon varies greatly among taxa and study designs. Analysis of mitogenome sequences is a better approach for resolving these ambiguities (Arnason et al., 2002; Miya, & Nishida, 2000) and for unequivocal species characterization, allowing exploration of alternative specie-specific mitochondrial barcoding regions. In vertebrates, the mitochondrial genome is a closed circular molecule of

about 15-22 kilobases (Kb) length, which encodes a set of 37 genes (2 rRNAs, 13 protein-coding genes and 22 tRNAs) (Broughton et al., 2001; Gissi et al., 2008). *Mt*DNA accounts for only ~1-2% of total cellular DNA by mass, and the nuclear genome contains multiple mitochondrial pseudogenes that, if included in a sequencing analysis, might be erroneously interpreted as *mt*DNA variants (Hazkani-Covo et al., 2010; Just et al., 2015). Therefore, enrichment of high-quality *mt*DNA from total DNA extract represents a critical initial step for a correct analysis. Different methods for *mt*DNA enrichment from animal tissue have been reported (Kasamatsu et al., 1971; Burgener et al., 1998). They may rely on special equipment (*e.g.*, ultracentrifugation in CsCl density gradient), expensive kits, or PCR amplification from total cellular DNA. The latter is the most used method, relatively low-cost and proficient, yet leading to artefacts because of high number of PCR amplification cycles that are often needed for sufficient enrichment (Quispe-Tintaya et al., 2015), and because high-fidelity DNA polymerases show error frequencies typically in the range of 10^{-6}mutations/bp (Cline et al., 1996). This can lead to misinterpretation of results and ultimately, incorrect conclusions. The most common way to sequence whole mitochondrial genome is based on a combination of long and short PCR reactions with a number of versatile primers followed by Sanger sequencing method (Miya, & Nishida, 1999). This approach reduces PCR bias and the inconsistent coverage but does not decrease PCR-generated "variants" or nuclear pseudogenes (Duan et al., 2018).

More recently, next-generation sequencing (NGS) and advances in bioinformatics tools have enabled the analysis of mitogenomes to be extended to non-model organisms on an unprecedented scale (Mardis, 2008). High-throughput analysis of whole *mt*DNA genomes across multiple samples improves coverage breadth with high resolution of very low-level variants. This technology revolutionized biodiversity studies by generating exceptional amounts of *DNA* sequence data for comparative genomics studies. The assembly of millions or billions of short reads allows covering entire *mt*DNA sequences repeatedly, and can be directly applied in comparative analyses. Standard methods for mitochondrial *DNA* extraction do not provide sufficient amount for direct NGS and must be followed by long-range PCR amplification, which can bias the interpretation of sequencing results (Green et al., 2006; Miller et al., 2008; Metzker, 2010).

With the aim to find the best techniques in terms of time, costs and efficiency to obtain the complete mitogenome sequence, we compared combined shotgun Illumina MiSeq and 'traditional' Sanger sequencing of long (7.3-9.3 Kb) and short amplicons (1.4-3.2 Kb), respectively, with Illumina MiSeq approach to miniprep kit-enriched *mt*DNA. The methods were tested on muscle tissue of the family *Sparidae* (Perciformes: Teleostei) because it is one of the most valuable and popular fish resources in the world, for which species substitution is very common (Armani et al., 2015). Therefore, the production of new sequences can enhance the authentication of seafood products. In fact, even though the family comprises about 38 genera and 159 species (https://www.fishbase.de/Summary/FamilySummary.php?ID=330; accessed 14.12.18), only for nine *Sparidae* species the complete mitochondrial genome sequences are available in GenBank.

In particular, we performed and compared the results of the following procedures: 1) long PCR followed by NGS, 2) primer walking method followed by Sanger sequencing, 3) *mt*DNA isolation by miniprep columns followed by direct NGS (shotgun sequencing), as reference method having small, estimated sequence inaccuracies relative to PCR-based methods. After comparing the three combined protocols for *mt*DNA enrichment and sequencing, a rapid, simple, and efficient technique for *mt*DNA isolation from fish muscle that can be used for NGS-based analyses was selected. The method allowed obtaining the complete mitogenome sequences from *Dentex gibbosus* (MG653593) and *Pagellus erythrinus* (MG653592), (D.M. MIPAAF, 22 Settembre 2017). This procedure maximizes efficiency in separation and sequencing of the *mt*DNA from fish tissue also allowing for balanced time, costs and sequence quality.

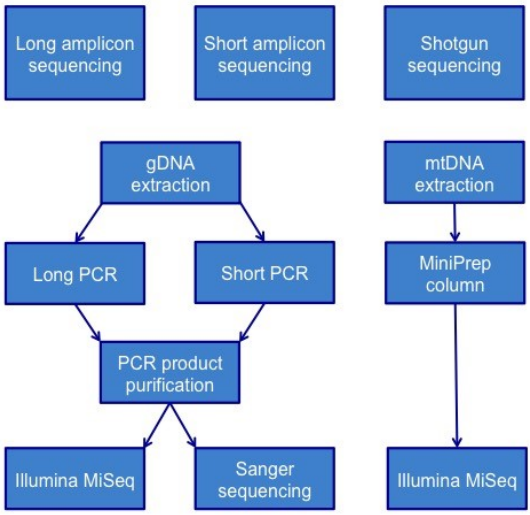
2.3. Materials and methods

2.3.1. *Fish samples*. One entire specimen of each species was caught in the Mediterranean Sea, and supplied by fish markets in Pozzuoli (Naples) (*Dentex gibbosus*), and Salerno (*Pagellus erythrinus*), Italy. Whole fish were frozen on board (–20° C) and shipped in insulated boxes to the laboratory. Specimens were identified, according to their anatomical and morphological features, as belonging to *D. gibbosus* and *P. erythrinus* species at the Department of Veterinary Medicine and Animal Production,

University "Federico II" (Naples, Italy). Muscle samples were taken and stored at -80 ºC for further analysis.

2.3.2 ***Total genomic DNA extraction***. Total genomic *DNA* (gDNA) was extracted from the dorsal fin of both specimens using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Using Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA), gDNA concentration was 40 ng/µl, and purity was in the range of 1.8-2.0 ratio at A260/A280. As an important step in the preparation of NGS libraries, quantity and integrity of extracted gDNA were assessed by electrophoretic analysis in 1% agarose gel. A sketched representation of the three sample processing workflows is reported in Figure 2.1.

Figure 2.1. Sample processing workflows for mtDNA enrichment and sequencing.



2.3.3. ***Mitogenome enrichment and sequencing.***

2.3.3.1 ***Long product amplification and NGS.*** The *mt*DNA genomes of *D. gibbosus* and *P. erythrinus* were amplified almost in their entirety as

reported in Mascolo *et al.*, 2017. Amplicons were purified with the High Pure PCR Product Purification Kit (Roche, Basel, Switzerland). *Mt*DNA complete sequence was determined with Illumina MiSeq sequencing (Illumina, San Diego, CA, USA) at Genomix4life S.R.L. (Baronissi, Salerno, Italy). Indexed libraries were prepared from 1 ug of each purified amplicon using Nextera *DNA* Library Preparation Kit protocol (Illumina, San Diego, CA, USA), according to the manufacturer's instructions. A sequence index is useful to tag each sample in unique manner, so after pooling it is possible to identify each of them. Libraries were quantified using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) and pooled such that each index-tagged sample was present in equimolar amounts, with a final concentration of pooled samples of 2 nM. Pooled samples were then subjected to cluster generation and sequenced using Illumina MiSeq 2500 System with v3 chemistry (Illumina, San Diego, CA, USA) in a 2x300 paired-end format at a final concentration of 8 pmol.

High-quality reads of *D. gibbosus* were processed for *de novo* assembly using Tadpole (BBTool, DOE Joint Genome Institute, Walnut Creek, USA). To remove possible contaminants from the assembled contigs, sequences were blasted against the NCBI NT database (minimum e-value set to 0.1 and -max_target_seqs 1). The taxonomic classification of each BLAST hit was obtained, and all sequences with a best hit whose taxa were not included in the *Sparidae* family, were discarded from the assembly. Multiple alignments of ten *D. gibbosus* contigs and three *P. erythrinus* contigs was performed with ClustalW 2.0 (Larkin et al., 2007). The known mitochondrial genome of *Parargyrops edita* was used as reference (EF107158.1).

2.3.3.2 *Short product amplification and Sanger sequencing*. *D. gibbosus* and *P. erythrinus* mitogenomes were amplified by using the primer set reported in Table 2.6 (supplementary material). Primers were designed by eye after multiple alignment of the *Sparidae* complete mitochondrial genome sequences available in GenBank using BioEdit Sequence Alignment Editor (Hall, 1999) (Figure 2.2 - supplementary material). The *mt*DNA sequences used for primers projecting were: *Acanthopagrus latus* (NC_010977.1, Xia *et al*., 2008), *Acanthopagrus schlegelii* (JQ746035.1, Shi et al., 2012), *Dentex tumifrons* (NC_029479.1), *Pagellus bogaraveo* (NC_009502.1, Ponce et al., 2008), *Pagrus major* (NC003196.1, Miya et al., 2001), *Pagrus auriga* (NC005146.1), *Parargyrops edita* (EF107158.1, Xia et al., 2007), *Rhabdosargus sarba* (KM272585.1, Li et al., 2016),

Sparus aurata (LK022698.1, Dray et al., 2016). We used Multiple Primer Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) to verify Tm, screen for secondary structure, self-annealing and inter primer binding. Amplicon size ranged from 1,443 bp to 3,293 bp. PCR reactions were carried out in a 25 µl reaction volume using 2.5 µl of dNTP mix (2 mM), 2.5 µl of $MgCl_2$ (10x), 0.5 µl of each primer (50 pMol/µl), 0.25 µl of 5 U/µl Taq (Roche, Basel, Switzerland), 16.75 µl of sterile distilled water, and 2 µl of *DNA* template containing approximately 80 ng of *DNA*. The thermal cycling profile was: an initial denaturation step of 95 °C for 5 minutes (min) followed by 35 cycles of denaturation at 95 °C for 1 min, annealing at 50-60 °C for 1 min, and extension at 72 °C for 3 min. PCR products were electrophoresed on a 1% agarose gel, visualized via ultraviolet transillumination, and purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). The amplicons were cloned using the TOPO TA PCR cloning kit (Invitrogen, Waltham, MA, USA). Plasmids were purified with QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and sequenced with the Automated Capillary Electrophoresis Sequencer 3730 *DNA* Analyzer (Applied Biosystems, Foster City, California, USA) by the Molecular Biology Service at the Stazione Zoologica Anton Dohrn. Mitochondrial *DNA* sequences were analyzed using the BioEdit Sequence Alignment Editor (Hall, 1999). After sequence quality analysis of the cloned products, new pairs of primers were designed to close *mt*DNA sequence gaps by performing further PCR reactions step-by-step (primer walking).

2.3.3.3. *Miniprep column enrichment and NGS (shotgun sequencing*). We assumed, as have others (Defontaine et al., 1991; Peloquin et al., 1993), that since *mt*DNA properties are similar to those of bacterial *DNA* (*i.e.*, it is supercoiled and its size is in the range of conventional plasmids), a common miniprep kit can be used for extraction and enrichment of *mt*DNA from fish tissues. In detail, tissue samples (25 mg of dorsal muscle) were collected into microcentrifuge tubes (2 ml), with 80 µl of cold 1% PBS and with 1 of 5-mm stainless steel beads (Qiagen, Valencia, CA, USA). Two samples for each species (*D. gibbosus:* DG1 and DG2; *P. erythrinus* PE8 and PE9) were grinded in Qiagen Retsch TissueLyser Mixer Mill Grinder *DNA*-RNA Homogenization 85210 MM301 (Qiagen, Valencia, CA, USA) at – 18 °C (TissueLyser Adapter Sets were cooled at – 18 °C for 12 h) with oscillation frequencies of 30 Hz for 30, 15 and 5 min.

Next, lysate tissues were pelleted by centrifugation at 8000 × *g* for 8 min at room temperature. The *Mt*DNA was extracted in duplicate by sample using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. The *DNA* was eluted into 50 µl of elution buffer. *DNA* quality and quantity were determined using Qubit™ 4 Fluorometer (Thermo Scientific Inc., Waltham, MA, USA) and Agilent 4200 TapeStation system (Agilent Technologies, Santa Clara, CA, USA) respectively, at Genomix4life s.r.l. (Baronissi, Salerno, Italy). *Mt*DNA NGS was performed at Bio-Fab Research s.r.l. laboratories (Rome, Italy). Libraries were prepared from 1 µg of each purified *DNA* sample, using XT kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions, and quantified using an Agilent Bioanalyzer 2100 chip HS *DNA* (Agilent Technologies, Santa Clara, CA, USA) and Qubit™ 4 Fluorometer (Thermo Scientific Inc., Waltham, MA, USA). Then, libraries were pooled in equimolar amounts at a final concentration of 4 nM. Pooled samples were subjected to cluster generation and sequenced using Illumina MiSeq (Illumina, San Diego, CA, USA) in a 2x300 (*D. gibbosus*) and in a 2x250 (*P. erythrinus*) paired-end format at a final concentration of 16 pM.

The quality of raw sequencing reads was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by removal of Illumina adapters and trimming using Trimmomatic (v0.33) (Bolger et al., 2014). Trimming was performed at both the 5'- and 3'-end of reads, eliminating bases with a Phred score of +33. Paired reads were processed simultaneously, and the orphan reads were removed. High-quality reads were used as input for *de novo* assembly of mitogenomes with a hybrid assembler method using Velvet (Zerbino et al., 2008) and MIRA (Chevreux et al., 1999). Scaffolds were obtained with Scaffolding Pre-Assemblies After Contig Extension (SSPACE) (Boetzer et al., 2011), and their annotation was made with Infernal: INFERence of RNA ALignments, and with BLAST/BLASTX for coding sequences (CDS) using RefSeq and MitoFish (Iwasaki et al., 2013).

2.3.4 Data analysis

To study the *DNA* polymerase error frequencies and verify any sequencing errors, we analysed the consensus sequences obtained with long and short PCR clones in each species for variants. With minimum sequence and assembly mistakes, sequences obtained by shotgun sequencing of

Miniprep column-enriched *mt*DNA were considered as reference. Sequences were compared calculating the average nucleotide identity (ANI) using QUAST 4.6.1 (Gurevich et al., 2013) and MUMmer (Delcher et al., 2003). ANI is the appropriate *in silico* alternative for *DNA-DNA* hybridisation (DDH) using blast algorithm. ANI method, in this case, followed the basic algorithm. All *Sparidae mt*DNA sequences were compared with that of a very distant arctic fish, *Pagothenia borchgrevinki*. Percentage of nucleotide identity of the matching regions was calculated as average for all matching regions. The percentage threshold for species boundary is 95% ANI (Richter, & Rosselló-Móra, 2009).

2.4.Results

2.4.1. *Long product amplification and MiSeq NGS.* *DNA* extracted from *D. gibbosus* and *P. erythrinus* muscle tissue was of good quality and quantity (40/80 ng/µl). The long PCR approach allowed amplifying the complete mitochondrial genome of these two sparid fish species. For each species, two well-resolved amplicons of 7336 and 9348 Kb were obtained, complementary to each other. NGS allowed to obtain high quality reads of *D. gibbosus* (7336 kb amplicon: 4034628 reads; 9348 kb amplicon: 4969506 reads) and *P. erythinus* (7336 kb amplicon: 6626288 reads; 9348 kb amplicon: 3264850 reads). Reads were assembled with a *de novo* method. Quality control was performed on the raw sequencing data to remove low-quality portions while preserving the longest high-quality reads. After normalization, high-quality reads were assembled for each sample separately. *De novo* assembly and alignment with *Parargyrops edita* reference mitogenome revealed the presence of five sequence gaps for *D. gibbosus* and one for *P. erythrinus* (Table 2.5 - supplementary material) with a total number of unsequenced bases of 5248 bp (17.96%) in *D. gibbosus* and 3025 bp (31.29%) in *P. erythrinus mt*DNA sequences**.**

2.4.2. *Short product amplification and Sanger sequencing.* The *DNA* samples obtained from muscle cells of the dorsal fin of *D. gibbosus* and *P. erythrinus* were of good quality and quantity (40/80 ng/µl). The PCR primers designed for short product PCR amplification (Table 2.6 - supplementary material) by alignment with the complete *mt*DNA genomes of *Sparidae* available in GenBank (Figure 2.2 - supplementary material), showed good efficiency. The consensus sequences obtained after amplicon assembly were aligned with the complete sequence of the *P. edita*

mitochondrial genome (EF107158.1). The coverage was of 9867 bp for *D. gibbosus* (58.83%) and 9130 bp for *P. erythrinus* (54.25%).

2.4.3. *Miniprep column enrichment and NGS.* The mitochondrial *DNA* fractions were isolated using 5 min fragmentation time with QIAprep Spin Miniprep Kit (Figures 2.3, 2.4 - supplementary material). The *P. erythrinus* and *D. gibbosus* samples selected for MiSeq sequencing were PE9 and DG1, respectively, because, despite lower concentration of *g*DNA, they were richer in *mt*DNA (Figures 2.3, 2.4 - supplementary material). NGS produced 2300388 reads for *D. gibbosus* and 3659022 reads for *P. erythrinus*. The quality control was performed on the raw sequencing data to remove low-quality portions of reads while preserving the longest high-quality ones. After normalization, high-quality reads were assembled for each sample separately (Table 2.7 - supplementary material). Coverage was calculated using *de novo* scaffolds (Table 2.8 - supplementary material). For *D. gibbosus*, the longest and over-represented scaffold was the complete mitogenome of 16771 bp (Figure 2.5 - supplementary material) (Mascolo et al., 2018). For *P. erythrinus*, instead, two different scaffolds were selected that matched the full mitochondrial characteristics: scaffold_1 length: 12091 bp and scaffold_3 length: 4543 bp. At first, both scaffolds were annotated separately. Finally, the full sequence of 16828 bp of *P. erythrinus* complete mitogenome was obtained for annotation (Ceruso et al., 2018).

2.4.4. *Data analysis.* Results of the ANI analysis were reported in the Tables 2.1, 2.2. The analysed genome fraction of *D. gibbosus* was 66.23% for long PCR amplicons and 59.13% for Sanger sequencing of short amplicons; the analysed genome fraction of *P. erythrinus* was: 61.99% for long amplicons and 58,53% for short ones. We did not find differences in the number of mismatches between sequences obtained with long PCR and the sequences obtained using shotgun sequencing. Instead, significant sequence variation was observed when shotgun sequences were compared with Sanger-based sequences of short amplicons (Tables 2.3, 2.4).

Table 2.1. D. gibbosus average nucleotide identity (ANI). Percentage of identity (1=100%) between shotgun sequences (mtDNA_extraction_gibbosus), long (Long_amplicon) and short (Sanger) PCR products, and P. borchgrevinki (Pagothenia).

	*mt*DNA extraction *gibbosus*	Long_amplicon	*Pagothenia*	Sanger
*mt*DNAextraction	1	0,99843	0,7216384615	0,9755

gibbosus				
Long_amplicon	0,99843	1	0,6996428571	0,9603
Pagothenia	0,7352428571	0,7115571429	1	0,7107
Sanger	0,9754444444	0,9603333333	0,7136625	1

Table 2.2. P. erythrinus average nucleotide identity (ANI). Percentage of identity (1=100%) between shotgun sequences (Complete_ery), long (Long_amplicon) and short (Sanger) PCR products, and P. borchgrevinki (Pagothenia).

	Long_amplicon	Complete_ery	*Pagothenia*	Sanger
Long_amplicon	1	0,99961	0,7065625	0,96715
Complete_ery	0,9993909091	1	0,7345785714	0,9921777778
Pagothenia	0,7107	0,7318928571	1	0,7516666667
Sanger	0,96745	0,98025	0,7490333333	1

Table 2.3. Statistics of D. gibbosus long (Long_amplicon) and short (Sanger) PCR product alignment with shotgun sequence.

Genome statistics	Long_amplicon	Sanger
Genome fraction (%)	66.233	59.132
Total aligned length	11108	9933
NG50	17005	16800

NGA50	11108	9933
Unaligned		
# fully unaligned contigs→	0	0
Fully unaligned length❖	0	0
# partially unaligned contigs	1	1
Partially unaligned length	5897	6867
Mismatches		
# mismatches	0	12
# indels	0	14
Indel length	0	16
# mismatches per 100 kb†	0	121
# indels per 100 kb♦	0	141.17
# indels (<= 5 bp)	0	14
# indels (> 5 bp)	0	0
# N's	5913	7066
# N's per 100 kb□	34.772	42.060
GC (%)⬧	45.29	45.96

□ Percentage of aligned bases in the reference genome; ▪ total number of aligned bases in the assembly; NG50 is the length for which the collection of all contigs of that length or longer covers at least half the reference genome; NGA50 ("A" stands for "aligned") similar to the corresponding metrics without "A", but in this case aligned blocks instead of contigs are considered. Aligned blocks are obtained by breaking contigs at mis-assembly events and removing all unaligned bases. →number of contigs that have no alignment to the reference sequence; ❖ total length of all unaligned regions in the assembly (sum of lengths of fully unaligned contigs and unaligned parts of partially unaligned ones); † average number of mismatches per 100000 aligned bases. True SNPs and sequencing errors are not distinguished and are counted equally; ♦ average number of indels per 100000 aligned bases. Several consecutive single nucleotide indels are counted as one indel; □ average number of uncalled bases (N's) per 100000 assembly bases; ⬧ total number of G and C nucleotides in the assembly, divided by the total length of the assembly.

Table 2.4. Statistics of P. erythrinus long (Long_amplicon) and short (Sanger) PCR product alignment with shotgun sequence.

Genome statistics	Long_amplicon	Sanger
Genome fraction (%)◘	61.992	54.528
Total aligned length▪	10436	9176
NG50	10437	16900
NGA50	10436	9176
Unaligned		
# fully unaligned contigs→	0	0
Fully unaligned length❖	0	0
# partially unaligned contigs	0	1
Partially unaligned length	0	7724
Mismatches		
# mismatches	0	1
# indels	4	2
Indel length	4	6
# mismatches per 100 kb†	0	10.9
# indels per 100 kb♦	38.34	21.8
# indels (<= 5 bp)	4	2
# indels (> 5 bp)	0	0
# N's	5	7793
# N's per 100 kb□	47.91	46112
GC (%)⬧	44.19	44.83

◘ Percentage of aligned bases in the reference genome; ▪ total number of aligned bases in the assembly; NG50 is the length for which the collection of all contigs of that length or longer covers at least half the reference genome; NGA50 ("A" stands for "aligned") similar to the corresponding metrics without "A", but in this case aligned blocks instead of contigs are considered. Aligned blocks are obtained by breaking contigs at mis-assembly events and removing all unaligned bases.

→ number of contigs that have no alignment to the reference sequence; ❖ total length of all unaligned regions in the assembly (sum of lengths of fully unaligned contigs and unaligned parts of partially unaligned ones); † average number of mismatches per 100000 aligned bases. True SNPs and sequencing errors are not distinguished and are counted equally; ♦ average number of indels per 100000 aligned bases. Several consecutive single nucleotide indels are counted as one indel; □ average number of uncalled bases (N's) per 100000 assembly bases; ♦ total number of G and C nucleotides in the assembly, divided by the total length of the assembly.

2.4.5. *Comparing speed, costs and accuracy*. Next-generation sequencing has become a very useful way in large-scale *DNA* sequence analysis, opening the possibility to enable cost-effective high-throughput approaches. In order to evaluate the characteristics associated to each methodological approach that we used to extract, enrich and sequence the entire mitogenome of Sparid fish species (Fig. 1), we present a side-by-side empirical comparison based on data gathered by direct or hands-on experience of the authors, from companies' websites and literature. Our analysis indicated that, whereas the shotgun approach correlates with NGS platforms, it allows faster, less expensive, more accurate and consistent analysis of *mt*DNA genome sequences (Table 5 – supplementary material).

2.5. Conclusion

The size of the mitogenomes sequenced during this study was similar to that determined for fish within the *Sparidae* family, as well as for teleosts belonging to other families (Broughton et al., 2001; Inoue et al., 2001; Miya et al., 2001). Illumina MiSeq sequences were confirmed by Sanger method (short amplicons), and blasted against the NCBI NT database. The taxonomic classification of each BLAST indicates unambiguous taxonomic membership to the teleost fish family *Sparidae*. The consensus sequences generated by long and short amplicons mapped to nucleotide sequences obtained by shotgun sequencing. Each of these mapped sequences was used to perform the ANI analysis. We did not find a significant difference in the number of nucleotide variations between consensus sequences obtained with long PCR, but significant variation was detected in consensus sequences obtained with short amplicons followed by Sanger sequencing. Probably, this is due to different error rates among Taq *DNA* polymerases. Taq error rate was measured at about 1 in 9000 nucleotides (Tindall et al., 1988) but fidelity values differ among research groups and assay methods (McInerney et al., 2014). However, it is reported that Q5 High-Fidelity *DNA* Polymerase (NEB, Ipswich, MA, USA) error rate is $<0.44 \times 10^{-6}$ (https://www.neb.com/tools-and-resources/selection

charts/dna-polymerase-selection-chart) and that Taq *DNA* polymerase (Roche, Basel, Switzerland) error rate is 1.3 x 10^{-5} with a mutation frequency of 5.1% (http://netdocs.roche.com/ddm/effective/pdf_0900b2fc80b90ba4.pdf).

Our analysis of staff time, labor costs and performance characteristics of three different enrichments and sequencing procedures allowed to determine that the best methodology for complete *mt*DNA sequencing is the shotgun sequencing, based on *mt*DNA enrichment with miniprep columns followed by MiSeq NGS. This rapid, simple, and efficient procedure can be used for routine molecular biology studies. Preparation of enriched *mt*DNA without PCR amplification simplifies the analysis and delivers an unbiased sequence. The analysis of fish species based upon appropriate and careful analysis of shotgun sequencing of total *mt*DNA provides sequences that have wide application in fish identification, phylogenetics and biogeography. Continuous improvements in sequencing platforms and analysis tools will make this method more reliable and cost effective. The significant advances in these competing sequencing platforms have been foreshadowed by the manufacture promise in further cost reductions in the near future. This would make shotgun sequencing of the whole *mt*DNA the preferred option for fish identification in many research and industrial applications. A comparison of the platforms at the same level of coverage is useful to provide a guide to likely comparative performance since the number of reads on these platforms will increase in the future (Brozynska et al., 2014). Costs will change for each platform as read length and volume are increased. However, cost is not the only advantage of this approach. This technique provides the basis for the molecular traceability of fish products, in agreement with the provisions of Regulation (EU) 1379/2013 (European Commission, 2013), in order to guarantee consumer safety and as support for official control activities.

Supplementary material

Table 2.5. Gaps in Dentex gibbosus and Pagellus erythrinus mtDNA sequences after de novo assembly.

	Gap	Length (nt)	Position (nt)*
D. gibbosus	1	2876	16561 - 3156
	2	149	4238 - 4559
	3	1255	5784 - 7143
	4	951	10933 - 12087
	5	17	14655 - 15030
P. erythrinus	1	1743	2563 - 4306

* Gap position after alignment with the complete mitochondrial genome sequence of *Parargyrops edita* (EF107158.1).

Table 2.6. Primer pairs used for generating short amplicons.

N°	Name	5'→ 3'	Position (nt)*	Amplicon size (nt)▪	Tm °C
1	**Fw1497**	GCCGGTAAAACTCGTGCCAGC	1497	2412	50.0
	Rev3908	CGGTCCTTTCGTACTACT	3908		
2	**Fw3760**	GATGTTGGATCAGGACATCCYAATGGTGCA	3760	1443	56.0
	Rev5203	CTAGGAAGTGGTGTAGTGGAAGCAC	5203		
3	**Fw5067**	CTGTGCCTGAAGTAAAGGGCCAC	5067	3050	67.6
	Rev8117	GGGTAGTCRGAGTATCGACG	8117		
4	**Fw7657**	CACATGTTTYACAGTYGG	7657	2652	50.0
	Rev10309	CCYTGRAATGTYCCTTCTCG	10309		
5	**Fw10156**	TACAGGKGCAGTTGCAGTTGCCGCCCT	10156	3124	72.0
	Rev13280	TTGCACCAAGAGTTTTTGGTTCCTAGACC	13280		
6	**Fw13118**	CACCGAGAKAGGCTYGCTA	13118	2592	56.0
	Rev15710	GAGTDACAACGGTGGTTTTTCAAGCC	15710		
7	**Fw15422**	GAGAATCHGCAGCAAGAGC	15422	1283	50.0
	Rev16993	AGAACGYTAGCTTTGGGAG	16993		
8	**Rev1716**	CATAGTGGGGTATCTAATCCCAGTTTG	1715	3098	68.0
	Fw15422	GAGAATCHGCAGCAAGAGC	15710		

* Primer position after alignment with mitochondrial genomes of *Sparidae* available in GenBank (NC_010977.1, JQ746035.1, NC_029479.1, NC_009502.1, NC003196.1, NC005146.1, EF107158.1, KM272585.1, LK022698.1); ▪ Amplicon size is the effective size of amplified products.

Table 2.7. Summary statistics for de novo assembly.

	Longest scaffold	Raw reads	EC reads	% reads passing EC	Raw nt
D. gibbosus	16771	2300388	2247342	97.69	547583072
P. erythrinus	12091	3659022	2414609	65.99	918414522

Longest scaffold: length of the longest scaffold; Raw reads: number of raw reads used as input; EC reads: number of reads used for assembly, after quality control and error correction; % reads passing EC: percentage of the raw reads that the EC reads represent; Raw nt: number of nucleotides used as input.

Table 2.8. Summary statistics for coverage.

	Raw cov	EC cov	Median cov	10th percentile cov	bases>= Q40 % GC	% GC
D. gibbosus	242.26	169.14	19	3	1851830	43.2
P. erythrinus	771.69	266.04	17	3	951847	42.2

Raw cov: average depth of sequence coverage provided by the raw data; EC cov: coverage after quality control and error correction; Median cov: median actual depth of coverage in the assembly; 10th percentile cov: 10th percentile depth of coverage -- 90% of sites have greater coverage; bases >= Q40 % GC: number of bases that have a PHRED-scale quality greater or equal to Q40; % GC: GC content.

Table 2.9. Analysis of time, costs and sequence quality for each analysed methodology.

Methods for *mt*DNA sequencing		Time		Costs		Base Call Accuracy[2]
		Step by step	Total[1]	Step by step	Total[1]	
	1- Short amplicon sequencing					
Steps	a- Total genomic *DNA* extraction	3 h	320 h	7 €	960 €	
	b- Short PCR and amplicon purification	6 h		5 €		
	c- Cloning	6 h		40 €		
	d- Sanger sequencing	2 h		8 €		~99.4%[3]
	e- Data analysis and primer design	3 h				
	2- Long amplicon sequencing					
Steps	a- Total genomic *DNA* extraction	3 h	130 h	7 €	768 €	
	b- Long PCR and amplicon purification	10 h		7 €		
	c- Illumina MiSeq	48 h		370 €		99.9%[3]
	e- Data analysis and primer design	4 h				
	3- Shotgun sequencing					
Steps	a- *Mt*DNA extraction	1 h	49 h	2 €	372 €	
	b- Illumina MiSeq	48 h		370 €		99.9%[3]

[1] Total time of the first method based on 16 "primer walking" cycles (steps a-d), at 1 kb newly generated *mt*DNA sequence per cycle. For the second method, time and costs are referred to a minimum of two long amplicons. For the third method, time and costs are referred to the entire mitogenome. [2] Base calling accuracy, measured by the Phred quality score (Q score) is the most common metric used to calculate the accuracy of a sequencing platform. It indicates the probability that the sequencer calls a given base erroneously. Low Q scores can increase false-positive variant calls, which can result in imprecise conclusions (Ewing et al, 1998; Ewing and Green, 1998). [3] https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf.

Figure 2.2. Position of the short PCR amplicons. Alignment with Sparidae mitogenomes available in GenBank (NC_010977.1, JQ746035.1, NC_029479.1, NC_009502.1, NC003196.1, NC005146.1, EF107158.1, KM272585.1, LK022698.1). Primer pairs 1-8 are reported in Table 2.

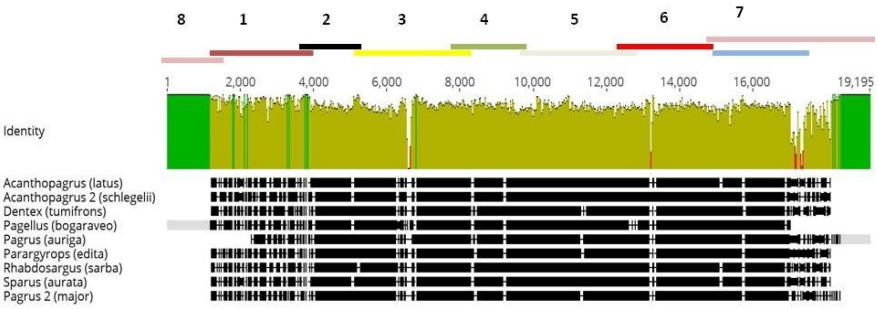


Figure 2.3. Quality control of Pagellus erythrinus mtDNA extraction and enrichment. Two large peaks were observed corresponding to a lower marker (average size 100 bp, from 65 to 151) and the mtDNA fraction (13843 bp, from 12588 to 18683) (sample PE9).

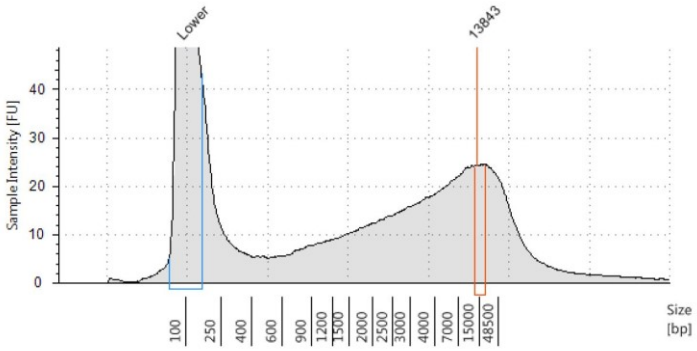
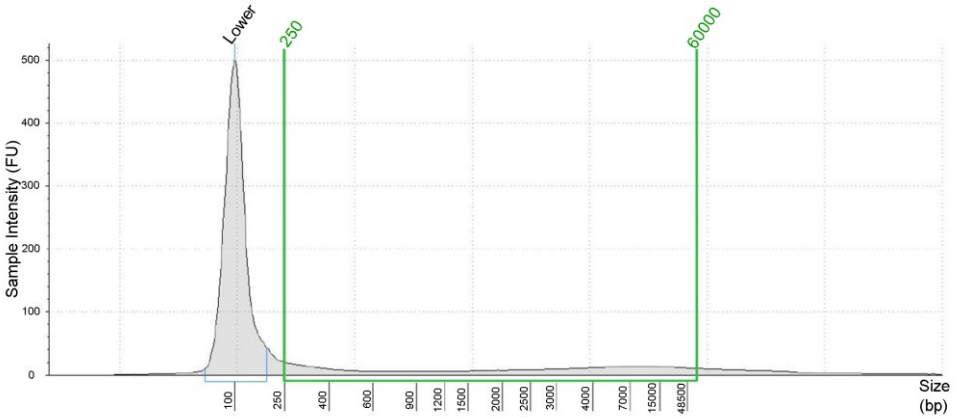


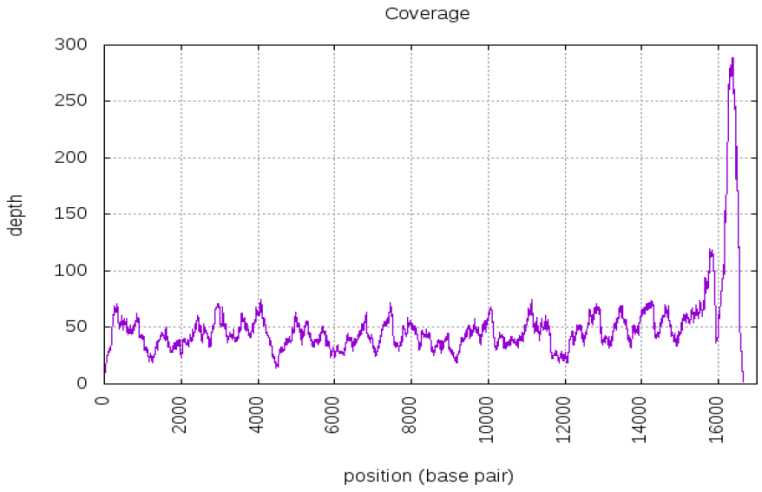
Figure 2.4. Quality control of Dentex gibbosus mtDNA extraction and enrichment. Two peaks were observed, a major one corresponding to a lower marker (average 100 bp, from 58 to 180) and a minor fraction (6264, from 250 to 60000) (sample DG1).



Region Table

From [bp]	To [bp]	Average Size [bp]	Conc. [ng/µl]	Region Molarity [nmol/l]	% of Total	Region Comment	Color
250	60000	6264	2.54	4.26	70.73		

Figure 2.5. Scaffold coverage of the complete Dentex gibbosus mtDNA sequence.



References

Armani, A., Guardone, L., Castigliego, L., D'Amico, P., Messina, A., Malandra, R., Gianfaldoni, D., & Guidi, A. (2015). DNA and Mini-DNA barcoding for the identification of Porgies species (family Sparidae) of commercial interest on the international market. *Food Control*, *50*, 589–596.

Arnason, U., Adegoke, J. A., Bodin, K., Born, E. W., Esa, Y. B., Gullberg, A., Nilsson, M., Short, R. V., Xu, X., & Janke, A. (2002). Mammalian mitogenomic relationships and the root of the eutherian tree. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 8151–8156.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*, 578–579.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Broughton, R. E., Milam, J. E., & Roe, B. A. (2001). The complete sequence of the zebrafish (*Danio rerio*) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. *Genome Research*, 11, 1958–1967.

Brozynska, M., Furtado A., & Henry R. J. (2014). Direct chloroplast sequencing: comparison of sequencing platforms and analysis tools for whole chloroplast barcoding. *PLoS ONE*, *27*, 6008–6013.

Burgener, M., & Hübner, P. (1998). Mitochondrial. DNA enrichment for species identification and evolutionary analysis. *Zeitschrift für Lebensmitteluntersuchung und Forschung A*, *207*, 261–263.

Cawthorn, D.-M., Steinman, H. A., & Witthuhn, R. C. (2012a). DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. *Food Research International*, *46*, 30–40.

Cawthorn, D.-M., Steinman, H. A., & Witthuhn, R. C. (2012b). Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene*, *491*, 40–48.

Ceruso, M., Mascolo, C., Lowe, E. K., Palma, G., Anastasio, A., Sordino, P., & Pepe, T. (2018). The complete mitochondrial genome of the common pandora *Pagellus erythrinus* (Perciformes: Sparidae). *Mitochondrial DNA Part B*, *3*, 624–625.

Chevreux, B., Wetter, T., & Suhai, S. (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, *99*, 45–56.

Cline, J., Braman, J. C., & Hogrefe H. H. (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, *24*, 3546–3551.

Defontaine, A., Lecocq, F. M., & Hallet, J. N. (1991). A rapid miniprep method for the preparation of yeast mitochondrial DNA. *Nucleic Acids Research*, *19*, 185.

Delcher, A. L., Salzberg, S. L., & Phillippy, A. M. (2003). Using MUMmer to identify similar regions in large sequence sets. *Current Protocols in Bioinformatics*, *10.3*.

D.M. MIPAAF (2017). Attribuzione delle denominazioni in lingua italiana delle specie ittiche di interesse commerciale. *Gazzetta Ufficiale*, *266*, 6–33.

Dray, L., Neuhof, M., Diamant, & A. Huchon, D. (2016). The complete mitochondrial genome of the gilthead seabream *Sparus aurata* L. (Sparidae). *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, *27*, 781–782.

Duan, M., Tu, J., & Lu, Z. (2018) Recent advances in detecting mitochondrial DNA heteroplasmic variations. *Molecules*, *23*, 324.

Espiñeira, M., González-Lavín, N., Vieites, & J. M. Santaclara, F. J. (2008). Authentication of anglerfish species (*Lophius* spp) by means of polymerase chain reaction restriction fragment length polymorphism (PCR-RFLP) and forensically informative nucleotide sequencing (FINS) methodologies. *Journal of Agricultural and Food Chemistry*, *56*, 10594–10599.

Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1988). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, *8*, 175–185.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, *8*, 186–194.

Gissi, C., Iannelli, F., & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, *101*, 301–320.

Green, R. E., Krause, J., Ptak, S. E., Briggs, A.W., Ronan, M. T., Simons, J. F., Pääbo, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature*, *444*, 330–336.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*, 1072–1075.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Inoue, J. G., Miya, M., Tsukamoto, K., & Nishida, M. (2001). Complete mitochondrial DNA sequence of the Japanese anchovy *Engraulis japonicus*. *Fisheries Science*, *67*, 828–835.

Hazkani-Covo, E., Zeller, R.M., Martin, W. (2010) Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genetics*, *6*, e1000834.

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., & Nishida, M. (2013). MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, *30*, 2531–2540.

Just, R.S., Irwin, J.A., Parson, W. (2015) Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International: Genetics*, *18*, 131–139.

Kasamatsu, H., Robberson D. L., & Vinograda, J. (1971). Novel closed-circular mitochondrial DNA with properties of a replicating intermediate. *Proceedings of the National Academy of Sciences of the United States of America*. *68*, 2252–2257.

Larkin, M. A., Blackshields, G. Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*, 2947–2948.

Li, J., Yang, H., Xie, Z., Yang, X., Xiao, L., Wang, X., Li, S., Chen, M., Zhao, H., & Zhang, Y. (2016). The complete mitochondrial genome of the *Rhabdosargus sarba* (Perciformes: Sparidae). *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, 27, 1606–1607.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*, 133–141.

Marko, P. B., Lee, S. C. Rice, A. M., Gramling, J. M., Fitzhenry, T. M., McAlister, J. S., Harper, G. R., &Moran, A. L. (2004). Fisheries: mislabelling of a depleted reef fish. *Nature*, *430*, 309–310.

Mascolo, C., Ceruso, M., Sordino, P., Palma G., Anastasio, A., & Pepe, T. (2017). Development of a method to extract and amplify the complete mitogenome of some Sparidae species. *Italian Journal of Food Safety*, *6*, 6154.

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., & Sordino, P. (2018). The complete mitochondrial genome of the Pink dentex *Dentex gibbosus* (Perciformes: Sparidae). *Mitochondrial DNA Part B*, *3*, 525–526.

McInerney, P., Adams, P., & Hadi, M. Z. (2014). Error rate comparison during polymerase chain reaction by DNA polymerase. *Molecular Biology International*, *287430*.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews Genetics*, *11*, 31–46.

Miller, W., Drautz, D. I., Ratan, A., Pusey, B., Qi, J., Lesk, A.M., … Schuster, S. C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*, *456*, 387–390.

Miya, M., & Nishida, M. (1999). Organization of the mitochondrial genome of a deep-sea fish, *Gonostoma gracile* (Teleostei: Stomiiformes): first example of transfer RNA gene rearrangements in bony fishes. *Marine Biotechnology*, *1*, 416–426

Miya, M., & Nishida, M. (2000). Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Molecular Phylogenetics and Evolution*, *17*, 437–455.

Miya, M., Kawaguchi, A., & Nishida, M. (2001). Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences *Molecular Biology and Evolution*, *18*, 1993–2009.

Paracchini, V., Petrillo, M., Lievens, A., Puertas Gallardo, A., Martinsohn, J. T., Hofherr, J., … Angers-Loustau, A. (2017). Novel nuclear barcode regions for the identification of flatfish species. *Food Control*, *79*, 297–308.

Peloquin, J. J., Bird, D. M., & Platzer E. G. (1993). Rapid miniprep isolation of mitochondrial DNA from metacestodes, and free-living and parasitic nematodes. *Journal of Parasitology*, *79*, 964–967.

Pepe, T., Trotta, M., Di Marco, I., Anastasio, A., Bautista, J. M., & Cortesi, M. L. (2007). Fish species identification in surimi-based products. *Journal of Agricultural and Food Chemistry*, *55*, 3681–3685.

Ponce, M., Infante, C., Jimenez-Cantizano, R. M., Perez, L., & Manchado, M. (2008). Complete mitochondrial genome of the blackspot seabream, *Pagellus bogaraveo* (Perciformes: Sparidae), with high levels of length heteroplasmy in the WANCY region. *Gene*, *409*, 44–52.

Quispe-Tintaya, W., White, R. R., Popov, V. N., Vijg, J., & Maslov, A. Y. (2015). Rapid mitochondrial DNA isolation method for direct sequencing. *Methods in Molecular Biol*ogy, *1264*, 89–95.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 19126–19131.

Shi, X., Su, Y., Wang, J., Ding, S., & Mao, Y. (2012). Complete mitochondrial genome of black porgy *Acanthopagrus schlegelii* (Perciformes, Sparidae). *Mitochondrial DNA*, *23*, 310–312.

Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., & Hajibabaei, M. (2015). A DNA Mini-barcoding system for authentication of processed fish products. *Scientific Reports*, *5*, 15894.

Tindall, K. R., & Kunkel, T. A. (1988). Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry*, *27*, 6008–6013.

Xia, J., Xia, K., & Jiang, S. (2008). Complete mitochondrial DNA sequence of the yellowfin seabream *Acanthopagrus latus* and a genomic comparison among closely related sparid species. *DNA Sequence*, *19*, 385–393.

Xia, J., Xia, K., Gong, J., & Jiang, S. (2007). Complete mitochondrial DNA sequence, gene organization and genetic variation of control regions in *Parargyrops edita*. *Fisheries Science*, *73*, 1042–1049.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs, *Genome Research, 18*, 821–829.

Comparison of mitochondrial *DNA* enrichment and sequencing methods from fish tissue

Chapter 3

The complete mitochondrial genome of the Pink dentex *Dentex gibbosus* (Perciformes: *Sparidae*)

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., & Sordino, P. (2018). The complete mitochondrial genome of the Pink dentex *Dentex gibbosus* (Perciformes: *Sparidae*). *Mitochondrial DNA Part B*, *3*, 525–526. Doi: 10.1080/23802359.2018.1467230

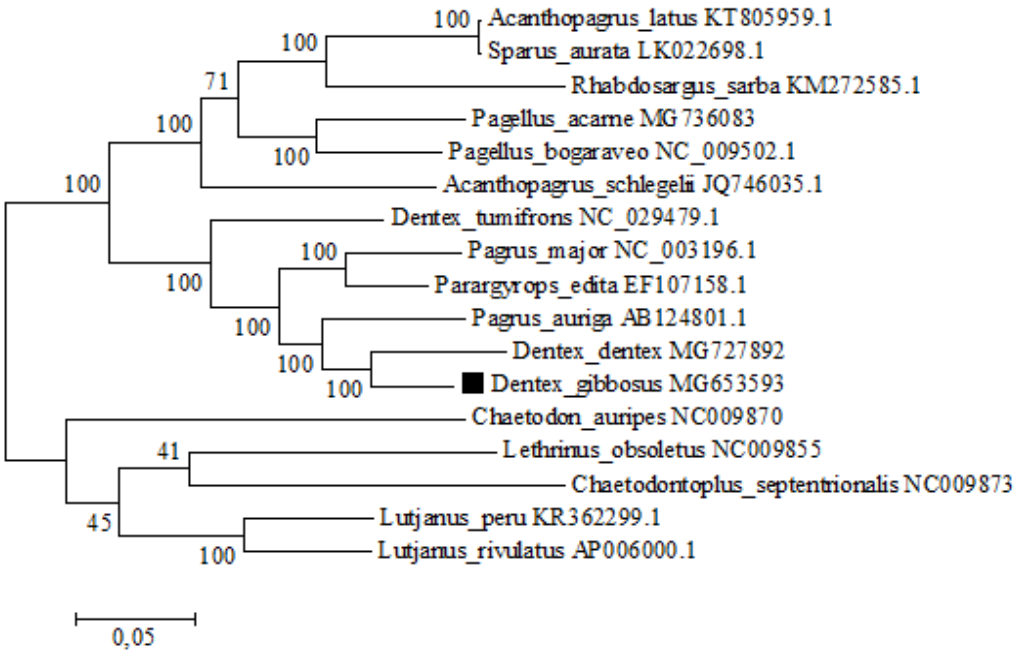
3. Abstract

The Pink dentex (*Dentex gibbosus,* Rafinesque 1810) is one of the most commercially important *Sparidae* species and it is often subject to fraud. Here, we report the complete mitochondrial genome of *Dentex gibbosus*. The mitogenome is 16.771 bp in length and contained 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and 2 non-coding regions. The overall base composition of *D. gibbosus mt*DNA is: 27,8 % for A, 28,60 % for C, 16,5 % for G, 27,05 % for T.

The Pink dentex (*Dentex gibbosus*) is a commercially important *Sparidae* species inhabiting the West African coast from Portugal to Angola, including Madeira, Canary and Sao Tome-Principe archipelagos (Wirtz *et al*. 2008). This sparid species is also present in the Mediterranean Sea except for the north-western coast and the northern Adriatic Sea (Dooley *et al.* 1985). *D. gibbosus* is now classified as “Least Concern” in the Red List of Threatened Species in the Mediterranean Sea (Russell *et al.*, 2014). Species substitution is very common in processed fishery products belonging to *Sparidae* family. In particular, *Dentex dentex* and *Pagrus pagrus*, are often fraudulently replaced with *D. gibbosus* (Katavic et al., 2000). Here, we describe the complete mitochondrial genome (mitogenome) of *D. gibbosus* (GenBank MG653593). A specimen caught in the Mediterranean Sea (N 38°25'15.1" E 3°53'09.4") was identified as *D. gibbosus* based on morphological features. *DNA* was extracted and is currently stored at Department of Veterinary Medicine and Animal Production, University "Federico II", Naples, Italy. The mitogenome of *D. gibbosus* has been obtained from high-throughput sequencing on complete mitochondrial *DNA* with Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA). The complete sequence is 16.771 bp long, containing 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). Mitochondrial arrangement and gene distribution are in agreement with the classic vertebrate mitogenomes (Wang *et al*. 2008). The majority of mitochondrial genes are encoded on the heavy strand, while the NAD dehydrogenase subunit 6 (*ND6*) and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu, Pro] are encoded on the light strand. Base composition is similar to other *Sparidae* mitochondrial

genomes, with 27,8 % for A, 28,60 % for C, 16,5 % for G, and 27,05 % for T (Ceruso *et al*., 2018). All protein-coding genes started with an ATG start codon but *COI* and *ND4*, which started with GTG. Stop codons were of 4 types, *i.e.* TAA (*ND1*, *ATP8*, *ATP6*, *ND4L*, *ND5, ND6*), AGG (*COI*), T (*COII*, *ND3, ND4*, *CYTB*) and TA (*ND2, COIII*). The 12S and 16S rRNA genes were located between the $tRNA^{Phe}$ (GAA) and $tRNA^{Leu}$ (TAA) genes and were separated by the $tRNA^{Val}$ gene as in other vertebrates (Li et al. 2016). The 22 tRNA genes vary from 66 to 74 bp in length. The 1091 bp long control region is located between $tRNA^{Pro}$ (TGG) and $tRNA^{Phe}$ (GAA). The non-coding region (L-strand origin of replication) is 40 bp long and is located between $tRNA^{Asn}$ (GTT) and $tRNA^{Cys}$ (GCA). To validate the phylogenetic position of *D. gibbosus*, we construct a phylogenetic tree using MEGA6 software (Tamura *et al*. 2013) (Figure 3.1). The resultant phylogeny shows that *D. gibbosus* is closely related to *D. dentex*, in agreement with Chiba *et al*. (2009). Results of this study provided useful genetic information for further studies on phylogeny, species identification and population genetics in *Sparidae* species.

Figure 3.1. Phylogenetic analysis of D. gibbosus based on the entire mtDNA genome sequences of 11 sparid fishes available in GenBank. Five outgroup species (Lutjanus peru, Lutjanus rivulatus, Lethrinus obsoletus, Chaetodontoplus septentrionalis and Chaetodon auripes) were selected and the maximum likelihood method was used. Numbers above the nodes indicate 1000 bootstrap values. Accession numbers are shown behind species names.



References

Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. 2018. The complete mitochondrial genome of the common dentex, *Dentex dentex* (Perciformes: Sparidae). *Mitochondrial DNA Part B*. 3: 391-392.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N. 2009. Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes Genet Syst*. 84:153–170.

Dooley JK, Van Tassell JL, Brito A. 1985. An Annotated Checklist of the Shorefishes of the Canary Islands. *American Museum Novitates* 2824: 1-49.

Li J, Yang H, Xie Z, Yang X, Xiao L, Wang X, Li S, Chen M, Zhao H, Zhang Y. 2016. The complete mitochondrial genome of the *Rhabdosargus sarba* (Perciformes: Sparidae), *Mitochondrial DNA Part A*. 27:1606–1607.

Russell B, Carpenter KE, Pollard D. 2014. *Dentex gibbosus*. The IUCN Red List of Threatened Species 2014:e.T170186A1289197. http://dx.doi.org/10.2305/IUCN.UK.20143.RLTS.T170186A1289197.en

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. *Gene*. 424:96–101.

Wirtz P, Fricke R, Biscoito MJ. 2008. The coastal fishes of Madeira Island-new records and an annotated check-list. *Zootaxa* 1715: 1-26.

Chapter 4

The complete mitochondrial genome of the common pandora *Pagellus erythrinus* (Perciformes: *Sparidae*)

Ceruso, M., Mascolo, C., Lowe, E. K., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2018b). The complete mitochondrial genome of the common Pandora Pagellus erythrinus (Perciformes: *Sparidae*), Mitochondrial DNA Part B, 3:2, 624-625. https://doi.org/10.1080/23802359.2018.1467235

4. Abstract

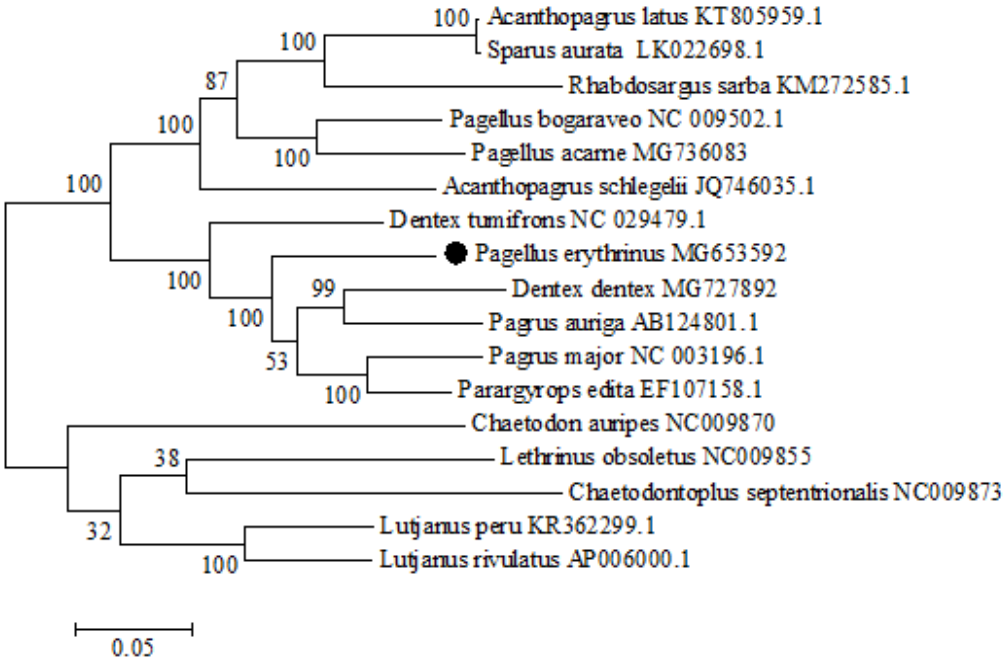
The common pandora (*Pagellus erythrinus,* Linnaeus 1758) is one of the most commercial caught fish species in the Mediterranean Sea and its potential as a significant aquaculture species is recognized. Here, we sequenced *Pagellus erythrinus* complete mitochondrial genome. The sequence is comprised of 16,828 bp and consists of 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and a two non-coding regions (D-loop and L-origin). The overall nucleotide composition is: 27.5% of A, 28.2% of C, 27.5% of T and 16.8% of G.

The common pandora (*Pagellus erythrinus*) is a benthopelagic sparid, distributed in the eastern Atlantic, from Scandinavia to Guinea-Bissau, including Cape Verde, Madeira and the Canary Islands (Sanches, 1991), and in the entire Mediterranean Sea. It is also rarely recorded in the Black Sea (Bauchot and Hureau, 1986). *P. erythrinus* is one of the sparid fishes which are harvested commercially in the greatest amounts in the Mediterranean Sea, and its potential as a significant aquaculture species is recognized (Basurco *et al*. 2011). However, genetic information about this species has yet to be adequately addressed. Here, we report the complete mitochondrial genome (mitogenome) of *P. erythrinus* (GenBank MG653592). A specimen caught in the Mediterranean Sea (N 40°22'54.0" E 14°35'44.0") was identified on anatomical and morphological features. *DNA* extracted from dorsal fin tissue is currently stored at Department of Veterinary Medicine and Animal Production, University "Federico II", Naples, Italy. The complete mitogenome of *P. erythrinus* was determined by using a combination of long and short PCR, followed by Sanger and Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA) sequencing methods. The complete sequence is 16,828 bp long, containing 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). Mitochondrial structure and gene organization are in agreement with the typical vertebrate mitogenome (Pereira 2000). The majority of mitochondrial genes were encoded on the heavy strand, with the NAD dehydrogenase subunit 6 (*ND6*) and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu, Pro] being encoded on the light strand. Base composition is 27.5% A, 28.2% C, 27.5% T and 16.8% G, similar to other

Sparidae mitochondrial genomes (Ceruso *et al*., 2018). All protein-coding genes started with an ATG start codon with the exception of *COI* and *ND4*, which started with GTG. Four types of stop codons were detected, *i.e.* TAA (*ND1*, *ND2*, *ATP8*, *ATP6*, *COIII*, *ND4L*, *ND6*), AGG (*COI*), T (*COII*, *ND4*, *CYTB*) and TAG (*ND3, ND5*). The 12S and 16S rRNA genes were located between the $tRNA^{Phe}$ (GAA) and $tRNA^{Leu}$ (TAA) genes, and were separated by the $tRNA^{Val}$ gene as in other vertebrates (Mascolo *et al.*, 2018). The 22 tRNA genes vary from 66 to 74 bp in length. The 1154 bp long control region is located between $tRNA^{Pro}$ (TGG) and $tRNA^{Phe}$ (GAA). The non-coding region (L-strand origin of replication) is 31 bp long and is located between $tRNA^{Asn}$ (GTT) and $tRNA^{Cys}$ (GCA).

The phylogenetic position of *P. erythrinus* was validated using MEGA6 software (Tamura *et al*. 2013), with the entire sparid *mt*DNA sequences available in GenBank. The species *Lutjanus peru*, *Lutjanus rivulatus*, *Lethrinus obsoletus*, *Chaetodontoplus septentrionalis* and *Chaetodon auripes* were used as outgroup for tree rooting (Figure 4.1). The resultant phylogeny shows that *P. erythrinus* is closely related to *D. tumifrons* and *Pagrus* spp., in agreement with the work of Chiba et al. (2009). *P. erythrinus* complete mitochondrial genome would be helpful in understanding the phylogenetics, evolution and species characterization of sparids.

Figure 4.1. Phylogenetic analysis of P. erythrinus based on the entire mtDNA genome sequences of 11 sparids and 5 outgroup species by maximum likelihood method. Numbers above the nodes indicate 1000 bootstrap values. Accession numbers are shown behind species names.



References

Basurco B, Lovatelli A, Garcia B. 2011. Current status of Sparidae aquaculture In M. Pavlidis, and C.C. Mylonas, Wiley-Blackwell editors. Sparidae: Biology and Aquaculture of Gilthead Sea Bream and Other Species. Oxford. p. 1–50

Bauchot ML, Hureau JC. 1986. Fishes of the north-eastern Atlantic and the Mediterranean. In P.J.P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen and E. Tortonese, editors. Sparidae. Paris. p. 883–907.

Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. 2018. The complete mitochondrial genome of the common dentex, *Dentex dentex* (Perciformes: Sparidae). *Mitochondrial DNA Part B*. 3: 391-392.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N. 2009. Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. Genes Genet Syst 84:153–170.

Dray L, Neuhof M, Diamant A, Huchon D. 2016. The complete mitochondrial genome of the gilthead seabream *Sparus aurata* L. (Sparidae). Mitochondrial DNA A DNA Mapp Seq Anal 27:781–782.

Li J, Yang H, Xie Z, Yang X, Xiao L, Wang X, Li S, Chen M, Zhao H, Zhang Y. 2016. The complete mitochondrial genome of the *Rhabdosargus sarba* (Perciformes: Sparidae), Mitochondrial DNA Part A. 27:1606–1607.

Sanches, J.G., 1991. Catálogo dos principais peixes marinhos da República de Guiné-Bissau. Publ. Avuls. Inst. Nac. Invest. Pescas 16:429 p.

Shi X, Su Y, Wang J, Ding S, Mao Y. 2012. Complete mitochondrial genome of black porgy *Acanthopagrus schlegelii* (Perciformes, Sparidae). Mitochondrial DNA. 23:310–312.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. Gene. 424:96–101.

Chapter 5

The complete mitochondrial genome of the Common dentex, *Dentex dentex* (Perciformes: *Sparidae*)

Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. 2018. The complete mitochondrial genome of the common dentex, *Dentex dentex* (Perciformes: *Sparidae*). *Mitochondrial DNA Part B*. 3: 391-392. Doi: 10.1080/23802359.2018.1450675

5. Abstract

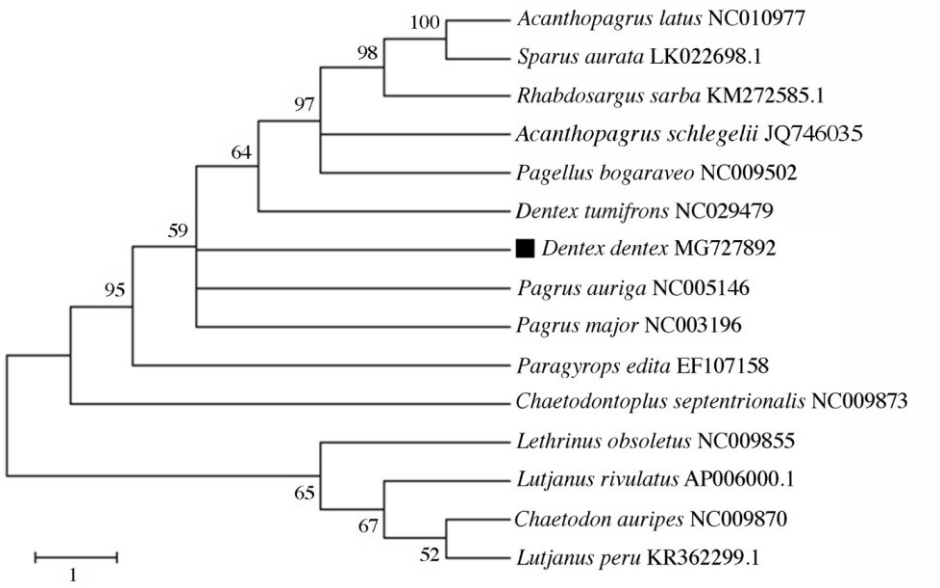
The common dentex (*Dentex dentex,* Linnaeus 1758) has a significant economic importance and is a highly valued table fish in the Mediterranean region. The paucity of genetic information relating to sparids, despite their growing economic value, provides the impetus for exploring the mitogenomics of this fish group. Here, we sequenced *Dentex dentex* complete mitochondrial genome. The sequence is comprised of 16,907 bp and consists of 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and a two non-coding regions (D-loop and L-origin). The overall nucleotide composition is: 27.5% of A, 28.7% of C, 26.9% of T and 16.9% of G.

The Common dentex (*Dentex dentex,* Linnaeus 1758) is one of the most commercially caught fish species in the Mediterranean Sea, very appreciated in European markets. It is a littoral and benthopelagic sparid distributed in the eastern Atlantic Ocean and in the entire Mediterranean Sea (Bauchot and Hureau 1986). Despite its significant economic importance, genetic information regarding this species is limited. We report the complete mitochondrial genome of *D. dentex* (GenBank MG727892). A specimen was caught in the Mediterranean Sea (N 41°46'58.3", E 16°26'06.7") and identified based on morphological features. Total *DNA* was extracted from dorsal fin tissue using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions. The complete mitogenome of *D. dentex* has been obtained from high-throughput sequencing on whole mitochondrial *DNA* with Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA). The complete sequence is 16,907 bp long, containing 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). Mitochondrial structure and gene organization are in agreement with the typical vertebrate mitogenome (Wang et al. 2008). The majority of mitochondrial genes were encoded on the heavy strand, with the NAD dehydrogenase subunit 6 (*ND6*) and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu, Pro] being encoded on the light strand. Base composition is 27.5% A, 28.7% C, 26.9% T and 16.9% G, similar to other *Sparidae* mitochondrial genomes (Shi *et al*., 2012; Dray et al. 2016). All protein-coding genes started with an ATG start codon but COI, which started with GTG. Stop codons were of 4 types, *i.e.*

TAA (*ND1*, *ND2*, *ATP8*, *ATP6*, *COIII*, *ND4L*, *ND5*, *ND6*), AGG (*COI*), T (*COII*, *ND4*, *CYTB*) and TAG (*ND3*). The 12S and 16S rRNA genes were located between the $tRNA^{Phe}$ (GAA) and $tRNA^{Leu}$ (TAA) genes, and were separated by the $tRNA^{Val}$ gene as in other vertebrates (Li et al. 2016). The 22 tRNA genes vary from 66 to 74 bp in length. The 971 bp long control region is located between $tRNA^{Pro}$ (TGG) and $tRNA^{Phe}$ (GAA). The non-coding region (L-strand origin of replication) is 39 bp long and is located between $tRNA^{Asn}$ (GTT) and $tRNA^{Cys}$ (GCA).

To validate the phylogenetic position of *D. dentex*, we used MEGA6 software (Tamura et al. 2013) with the entire *mt*DNA sequence of the sparids *Acanthopagrus latus*, *Acanthopagrus schlegelii*, *Dentex tumifrons*, *Pagellus bogaraveo*, *Pagrus major*, *Pagrus auriga*, *Parargyrops edita*, *Rhabdosargus sarba* and *Sparus aurata*. The species *Lutjanus peru*, *Lutjanus rivulatus*, *Lethrinus obsoletus*, *Chaetodontoplus septentrionalis* and *Chaetodon auripes* were used as outgroup for tree rooting (Figure 5.1). The resultant phylogeny shows that *D. dentex* is closely related to *D. tumifrons* and *Pagrus* spp., in agreement with the work of Chiba et al. (2009). The study of the mitochondrial genome of *D. dentex* and closely related species may reveal novel barcoding regions and inform on lineage diversification patterns in sparids.

Figure 5.1. Phylogenetic analysis of Dentex dentex based on the entire mtDNA genome sequences of 9 sparids and 5 outgroup species by maximum likelihood method. Numbers above the nodes indicate 1000 bootstrap values. Accession numbers are shown behind species names.



References

Bauchot ML, Hureau JC. 1986. Fishes of the north-eastern Atlantic and the Mediterranean. In P.J.P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen and E. Tortonese, editors. Sparidae. Paris. p. 883–907.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N. 2009. Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. Genes Genet Syst 84:153–170.

Dray L, Neuhof M, Diamant A, Huchon D. 2016. The complete mitochondrial genome of the gilthead seabream *Sparus aurata* L. (Sparidae). Mitochondrial DNA A DNA Mapp Seq Anal 27:781–782.

Li J, Yang H, Xie Z, Yang X, Xiao L, Wang X, Li S, Chen M, Zhao H, Zhang Y. 2016. The complete mitochondrial genome of the *Rhabdosargus sarba* (Perciformes: Sparidae), Mitochondrial DNA Part A. 27:1606–1607.

Shi X, Su Y, Wang J, Ding S, Mao Y. 2012. Complete mitochondrial genome of black porgy *Acanthopagrus schlegelii* (Perciformes, Sparidae). Mitochondrial DNA. 23:310–312.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. Gene. 424:96–101.

Chapter 6

The complete mitochondrial genome of the Axillary seabream, *Pagellus acarne* (Perciformes: *Sparidae*)

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Sordino, P., Pepe, T. (2018 a). The complete mitochondrial genome of the Axillary seabream, *Pagellus acarne* (Perciformes: *Sparidae*). *Mitochondrial DNA Part B*, 3, 434-435. Doi: 10.1080/23802359.2018.1450674

6. Abstract

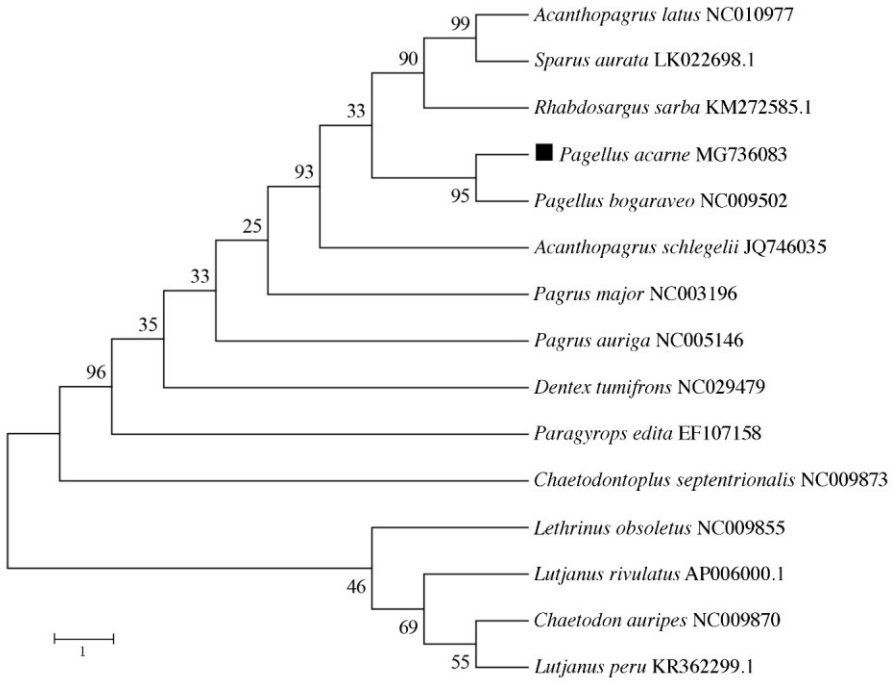
The Axillary seabream (*Pagellus acarne,* Risso 1827) belongs to the *Sparidae* family, order Perciformes. This highly-valued commercial fish species is distributed along the northern and eastern Atlantic coasts from Norway to Senegal, and throughout the Mediterranean Sea. Its complete mitochondrial genome is 16,486 bp in length, consisting of 13 protein-coding genes, 22 tRNA genes, two rRNA genes, and two non-coding regions region (D-loop, 808 bp and L-origin, 29 bp). Its overall base composition is A: 26,8%, C: 29%, G: 17.6%, and T: 26.6 %.

The axillary seabream (*Pagellus acarne,* Risso 1827) is a demersal fish belonging to the *Sparidae* family. Its geographical distribution extends from the Black Sea and the Mediterranean, along the west coast of Europe and Africa, to Norway and Angola (Santos et al. 1995). *P. acarne* is a very common species especially in the Mediterranean, where it is prevalent in commercial and artisanal catches. It is listed as Least Concern in the IUNC Red List of Threatened Species, with the recommendation of reducing current fishing efforts (Russell et al. 2014). In this study, the complete mitochondrial genome of *P. acarne* was sequenced with Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA) (GenBank MG736083). The specimen was collected in the Mediterranean Sea (FAO area 37; N 40°54'47.40", E 14°43'56.6") and identified based on morphological features. The complete sequence was 16,486 bp in length, including 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). This gene arrangement is similar to the typical vertebrate mitogenome (Wang et al. 2008). Most of the genes were encoded on the heavy strand, while the NAD dehydrogenase subunit 6 (*ND6*) and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (*UCN*), Glu, Pro] are encoded on the light strand. The nucleotide composition is A: 26,8%, C: 29%, G: 17.6%, and T: 26.6%, which is similar to other *Sparidae* mitogenomes (Xia et al. 2008; Shi et al. 2012; Dray et al. 2016). Protein-coding genes began with an ATG start codon, with the exception of *COI* and *ND4* that start with GTG. Four types of stop codons revealed are TAA (*ND1*, *ATP6*, *COIII*, *ND4L*, ND5), AGG (*COI*), T (*COII*, *ND4*, *Cytb*) and TAG (*ND2*, *ATP8, ND6*). The 12S and 16S rRNA genes were located between the $tRNA^{Phe}$ (GAA) and $tRNA^{Leu}$ (TAA) genes, and were separated by the

$tRNA^{Val}$ gene as in other vertebrates (Li et al. 2016). The 22 tRNA genes vary from 66 to 73 bp in length. The 808 bp-long control region is located between $tRNA^{Pro}$ (TGG) and $tRNA^{Phe}$ (GAA). The non-coding region (L-strand origin of replication) is 38 bp long and is located between $tRNA^{Asn}$ (GTT) and $tRNA^{Cys}$ (GCA).

The phylogenetic position of *P. acarne* was validated using MEGA6 software (Tamura et al. 2013) to construct a maximum-likelihood (ML) tree with 1000 bootstrap replicates, containing the complete mitogenomes of 10 *Sparidae* species (Figure 6.1). The resultant phylogeny shows that the *P. acarne* is closely related to *P. bogaraveo* with high bootstrap value supported. The study of the mitochondrial genome of *P. acarne* may reveal novel barcoding regions and deepen our knowledge of the evolution of sparid fishes

Figure 6.1. The phylogenetic position of Pagellus acarne was validated by ML method with the complete mitogenomes of 9 sparids and 5 arbitrary outgroup species (Lutjanus peru, Lutjanus rivulatus, Lethrinus obsoletus, Chaetodontoplus septentrionalis, Chaetodon auripes). Numbers above the nodes indicate 1000 bootstrap values. Mitogenome accession numbers are listed behind the species names.



References

Dray L, Neuhof M, Diamant A, Huchon D. 2016. The complete mitochondrial genome of the gilthead seabream Sparus aurata L. (Sparidae). Mitochondrial DNA A DNA Mapp Seq Anal.;27(1):781-2.

Li J, Yang H, Xie Z, Yang X, Xiao L, Wang X, Li S, Chen M, Zhao H, Zhang Y. 2016. The complete mitochondrial genome of the Rhabdosargus sarba (Perciformes: Sparidae), Mitochondrial DNA Part A, 27:3, 1606-1607.

Russell B, Carpenter K.E. Pollard D. 2014. Pagellus acarne. The IUCN Red List of Threatened Species 2014: e.T170229A1297432. http://dx.doi.org/10.2305/IUCN.UK.2014-3.RLTS.T170229A1297432.en

Santos M. N., C. Costa Monteiro, K. Erzini. 1995. Aspects of the biology and gillnet selectivity of the axillary seabream (*Pagellus acarne*, Risso) and common pandora (*Pagellus erythrinus*, Linnaeus) from the Algarve (south Portugal), Fisheries Research, 23, 223-236.

Shi X, Su Y, Wang J, Ding S, Mao Y. 2012. Complete mitochondrial genome of black porgy *Acanthopagrus schlegelii* (Perciformes, Sparidae). Mitochondrial DNA 23:310–2.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. Gene 424:96–101.

Xia J, Xia K, Jiang S. 2008. Complete mitochondrial DNA sequence of the yellowfin seabream *Acanthopagrus latus* and a genomic comparison among closely related sparid species. Mitochondrial DNA 19:385–93.

Chapter 7

The complete mitochondrial genome of the Angolan dentex *angolensis* (Perciformes: *Sparidae*)

The complete mitochondrial genome of the Angolan dentex *angolensis* (Perciformes: Sparidae). Mascolo C., Ceruso M., Chirollo C., Palma G., Anastasio A., Sordino P. and Pepe T. (2019). *Mitochondrial DNA Part B*. Doi: 10.1080/23802359.2019.1591248

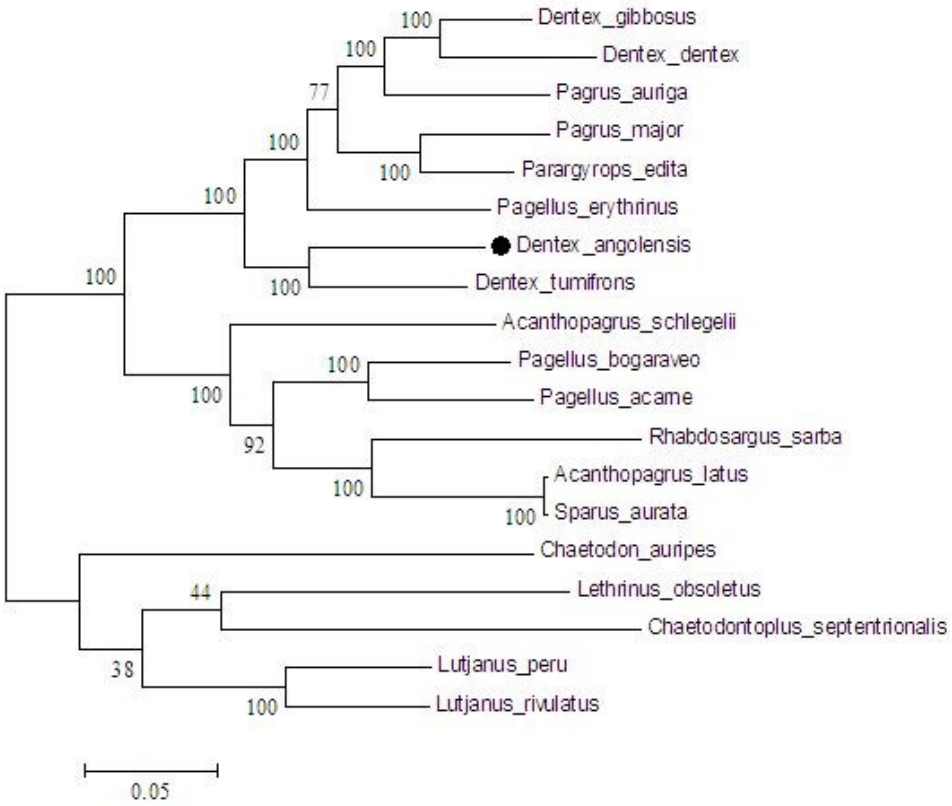
7. Abstract

The complete mitochondrial genome of the Angolan dentex (*Dentex angolensis,* Poll and Maul, 1953) is 16.581 bp in length and contains 13 protein-coding genes, 22 transfer RNA genes and 2 ribosomal RNA genes. The overall base composition of *D. angolesis mt*DNA is: 27,4 % for A, 28 % for C, 16,7 % for G, 27,9 % for T. Phylogenetic analysis indicated that D. angolensis is most closely related to D. tumifrons. Whole genome sequencing of *D. angolensis* mitochondrial *DNA* will contribute to improve knowledge about *Sparidae* evolution.

The Angolan dentex (*Dentex angolensis*) has one of the highest economic value among marine fishes (DM n. 19105 del 22 Settembre 2017). For consumption purposes, it is marketed fresh, frozen, and sometimes dried salted. D. angolensis is widely distributed in the Eastern Central Atlantic along the West Coast of Africa from Morocco to Angola, including Canary Islands (Rico et al., 1995). The economic value of D. angolensis is driving overfishing in some regions, with substantial (10%/year) population decline (Aheto et al. 2011). Thus, the species is now classified as "Near Threatened" in the Red List of Threatened Species (Pollard et al., 2014).
To make a step forward our understanding of the evolution of sparid mitogenomes, we sequenced the complete mitochondrial genome of *D. angolensis* (GenBank accession no.MH593823). A whole specimen caught in the South Atlantic Ocean (FAO area 47; S 8°27'07.1", E 13°00'05.0") was identified as D. angolensis based on morphological features. *DNA* extracted was stored at the Department of Veterinary Medicine and Animal Production, University "Federico II", Naples, Italy. D. angolensis mitogenome has been sequenced with Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA). The complete sequence is 16.581 bp long, containing 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). Mitochondrial array and gene distribution are in agreement with other vertebrate mitogenomes (Wang et al., 2008). Base composition is similar to one measured in other mitochondrial *DNA* sequences of *Sparidae* species, with 27,4 % for A, 28 % for C, 16,7 % for G, 27,9 % for T (Ceruso et al., 2018). Mitochondrial genes were distributed

in two groups: most of the genes were encoded on the heavy strand, while the NAD dehydrogenase subunit 6 (ND6), and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu, Pro] are on the light strand. All protein-coding genes begin with an ATG start codon except for COI that started with GTG. Stop codons were of 4 types, i.e. TAA (ND1, ND2, ATP8, ATP6, COIII, ND4L, ND5, ND6,), AGG (COI), T (COII, ND4, CYTB) and, TAG (ND3). Ribosomal subunit 12S (954 bp) was located between the tRNAPhe (GAA) and tRNALeu (TAA) genes, while 16S (1674 bp) was found between tRNAVal (TAC) and tRNALeu (TAA) genes, as in other vertebrates (Mascolo et al. 2018). The 22 tRNA genes vary from 66 to 74 bp in length. The control region is 904 bp long and is located between tRNAPro (TGG) and tRNAPhe (GAA). The non-coding region (L-strand origin of replication) is 35 bp long and is located between tRNAAsn (GTT) and tRNACys (GCA). To validate the phylogenetic position of D. angolensis, we constructed a phylogenetic tree using MEGA6 software (Tamura et al., 2013) (Figure 7.1). The resultant phylogeny shows that D. angolensis is closely related to D. tumifrons, in agreement with Chiba et al. (2009). This study can be used to evaluate genetic diversity and population structure of sparids, and to provide key information for phylogenetic studies.

Figure 7.1. Phylogenetic analysis of *D. angolensis* based on the entire *mt*DNA genome sequences of 13 sparid fishes available in GenBank (Acanthopagrus latus NC_010977.1; Acanthopagrus schlegelii JQ_746035.1; Dentex MG_727892.1; Dentex gibbosus MG_653593.1; Dentex tumifrons NC_029479.1; Pagellus acarne MG_736083.1; Pagellus bogaraveo NC_009502.1; Pagellus erythrinus MG_653592.1; Pagrus auriga AB_124801.1; Pagrus major NC_003196.1; Parargyrops edita EF_107158.1; Rhabdosargus sarba KM_272585.1; Sparus aurata LK_022698.1). Five outgroup species (Lutjanus peru KR362299.1, Lutjanus rivulatus AP006000.1, Lethrinus obsoletus NC009855, Chaetodontoplus septentrionalis NC009873 and Chaetodon auripes NC009870) were selected and the maximum likelihood method was used. Numbers above the nodes indicate 1000 bootstrap values.



References

Aheto, D.W., Asare, N.K., Quaynor, B., Tenkrang, E.Y., Asare, C. and Okyere, I. 2012. Profitability of Small-scale Fisheries in Elmina, Ghana. Sustainability 4: 2785-2794.

Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. 2018. The complete mitochondrial genome of the common dentex, Dentex (Perciformes: *Sparidae*). Mitochondrial DNA Part B. 3: 391-392.

Chiba SN, Iwatsuki Y, Yoshino T, Hanzawa N. 2009. Comprehensive phylogeny of the family *Sparidae* (Perciformes: Teleostei) inferred from mitochondrial gene analyses. Genes Genet Syst. 84:153–170.

Mascolo C, Ceruso M, Palma G, Anastasio A, Sordino P, Pepe T. 2018. The complete mitochondrial genome of the axillary seabream, Pagellus acarne (Perciformes: *Sparidae*). Mitochondrial DNA Part B. 3: 434-435.

Pollard D., Buxton C.D., Carpenter K.E. & Russell B. 2014. Dentex angolensis. The IUCN Red List of Threatened Species 2014: e.T170158A1283985. http://dx.doi.org/10.2305/IUCN.UK.2014-3.RLTS.T170158A1283985.en.

Rico V., J.I. Santana & J.A. González. 1995. Ocurrence of Dentex (Polysteganus) angolensis Poll and Maul, 1953 (*Sparidae*) in the Canary Islands. Cybium. 19 (4): 323-432.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (Ctenopharyngodon idella, Teleostei): Insight into its phylogenic position within Cyprinidae. Gene. 424:96–101.

Decreto n. 19105, 22 settembre 2017. Attribuzione delle denominazioni in lingua italiana delle specie ittiche di interesse commerciale. GU Serie Generale n.266 del 14-11-2017.

The complete mitochondrial genome of the Red porgy *Pagrus pagrus* (Perciformes: *Sparidae*), preliminary results.

Chapter 8

The complete mitochondrial genome of the Red porgy *Pagrus* (Perciformes: *Sparidae*), preliminary results.

The complete mitochondrial genome of the Red porgy *Pagrus pagrus* (Perciformes: *Sparidae*), preliminary results.

8. Abstract

The *Red porgy* (*Pagrus*, Linnaeus 1758) is a significant species of commercial and recreational fisheries in the Southeast Atlantic coast of the United States, Argentina, and the Mediterranean coasts. The complete mitochondrial genome of the Red porgie is 16,167 bp in length and contains 13 protein-coding genes, 22 transfer RNA genes and 2 ribosomal RNA genes.

The Red porgy or common seabream (*Pagrus pagrus*) is a significant species of commercial and recreational fisheries of the Sparidae family (DM n. 19105 del 22 Settembre 2017). Because of over-fishing populations have declined and in some areas minimum size restrictions have been established to try to compensate this. The Red porgy is also a species of great commercial importance for the aquaculture industry, because of its high market request, good advance rates and adaptability in culture environments (Russell et al, 2014). *P. pagrus* is distributed along warm costs of the Atlantic Ocean, in the western coast of Europe and the Mediterranean Sea, along the eastern coasts of North and South America and the Caribbean Sea (FAO FishFinder, 2019).
To make a step forward our understanding of the evolution of sparid mitogenomes, we sequenced the complete mitochondrial genome of *P. pagrus*. A whole specimen caught in the middle of Mediterranean Sea (FAO area 37.1) was identified as *P. pagrus* based on morphological features. DNA extracted was stored at the Department of Veterinary Medicine and Animal Production, University "Federico II", Naples, Italy. *P. pagrus* mitogenome was been sequenced with Illumina MiSeq 2500 System (Illumina, San Diego, CA, USA). The complete sequence is 16.167 bp long, containing 13 protein-coding genes, 2 ribosomal RNA genes (12S rRNA and 16S rRNA), 22 transfer RNA genes (tRNA) and two non-coding regions (D-loop and L-origin). The gene arrangement is in agreement to the typical vertebrate mitogenome (Wang *et al*., 2008). Mitochondrial arrangement and gene delivery are in pact with other vertebrate complete mitogenomes (Wang et al., 2008). Mitochondrial genes were scattered in two sets: greatest of the genes were encoded on the heavy strand, while the NADH dehydrogenase subunit 6 (ND6), and eight tRNA genes [Gln, Ala, Asn, Cys, Tyr, Ser (UCN), Glu, Pro] are on the light strand. Altogether protein-coding genes begin with an ATG start codon except for COI that

started with GTG. Stop codons were of 4 types, i.e. TAA (ND1, ATP8, ND4L, ND5, ND6,), AGG (COI), T (COII, ND4, CYTB) and TA (ND3, ATP6, COIII, ND2). Ribosomal subunit 12S (797 bp) was positioned between the tRNAPhe (GAA) and tRNALeu (TAA) genes, while 16S (1676 bp) was start between tRNAVal (TAC) and tRNALeu (TAA) genes, as in other vertebrates and sparids (Ceruso et al., 2018; Mascolo et al. 2018). The 22 tRNA genes vary from 66 to 74 bp in length.

D-loop is 712 bp long and is set between tRNAPro (TGG) and tRNAPhe (GAA). The non-coding region (L-strand origin of replication) is 36 bp long and is sited between tRNA Asn (GTT) and tRNACys (GCA).

References

Ceruso M, Mascolo C, Palma G, Anastasio A, Pepe T, Sordino P. 2018. The complete mitochondrial genome of the common dentex, Dentex (Perciformes: Sparidae). Mitochondrial DNA Part B. 3:391–392.

Decreto n. 19105, 22 settembre 2017. Attribuzione delle denominazioni in lingua italiana delle specie ittiche di interesse commerciale. GU Serie Generale n.266 del 14-11-2017.

FAO FishFinder - Web Site. About FAO FishFinder. FI Institutional Websites. In: *FAO Fisheries and Aquaculture Department* [online]. Rome. Updated 12 January 2017. [Cited 16 February 2019]. http://www.fao.org/fishery/

Mascolo C, Ceruso M, Palma G, Anastasio A, Sordino P, Pepe T. 2018. The complete mitochondrial genome of the axillary seabream, *Pagellus acarne* (Perciformes: Sparidae). *Mitochondrial DNA Part B*. 3: 434-435.

Russell B.; Pollard D.; Carpenter K.E. & Vega-Cendejas M. (2014). "Pagrus". The IUCN Red List of Threatened Species. IUCN. 2014: e.T15873A788483. doi:10.2305/IUCN.UK.2014-3.RLTS.T15873A788483.en.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Wang C, Chen Q, Lu G, Xu J, Yang Q, Li S. 2008. Complete mitochondrial genome of the grass carp (*Ctenopharyngodon idella*, Teleostei): Insight into its phylogenic position within Cyprinidae. *Gene*. 424:96–101.

Chapter 9

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

Ceruso, M., Mascolo C., Anastasio, A., Pepe, T., Sordino, P. (2019). *Food Control.* In press. doi: https://doi.org/10.1016/j.foodcont.2019.03.028.

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

9.1. Abstract

Sparids have different organoleptic properties that correlate with a wide variety of retail prices in the market. Components of the observed morphology are rarely sufficient for full identification of these fish species, whose authentication requires specialist knowledge. Genetic diversity or variation and their measurements enable molecular methods as one of the most suggested remedies for *aliud pro alio* frauds. Genetic approaches have the potential for reducing costs and providing correct identification for a large number of market products. Mitochondrial (*mt*) *DNA* sequences (*16S* and *12S* ribosome subunits, cytochrome b-*Cytb*, and cytochrome c oxidase I-*COI*) have been widely used for fish species identification. Yet, these *mt*DNA regions perform well for certain species but are less discriminating for others. Here, we report the first study of the whole *mt*DNA of the perciform fishes of the family *Sparidae* with the aim to select more efficient barcoding markers for taxonomical discrimination against frauds. For species-level sequence information, we analyzed and compared the whole sequence of thirteen *Sparidae* mitogenomes, nine publicly available and four recently sequenced ones. In particular, we searched for effective *DNA* barcode markers for the correct identification of sparid species by looking for interspecific variable regions flanked by conserved sequences for PCR primer design. We found that only four *mt*DNA genes are devoid of insertions or deletions, which can complicate the process of sequence alignment. Among them, *NAD* genes show encouraging utility in discriminating closely related sparid species owing to nucleotide sequence variability compared with classical barcodes for species. Discrimination capacity of *NAD* genes suggests their application as alternative *mt*DNA tools for the identification of *Sparidae* fishes. In particular, *NAD5* fragments with high interspecific nucleotide sequence divergence were amplified and appear flawless for *Sparidae* species identification.

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

9.2 Introduction

The growing global diffusion of seafood for human consumption requires always more accurate sanitary and quality controls. sparids, comprises about 38 genera and 159 species (https://www.fishbase.de/Summary/FamilySummary.php?ID=330), some of which are greatly appreciated as seafood (Tinacci *et al*., 2019). The identification of *Sparidae* species is not feasible even when external characters are preserved due to the morphological resemblance of these fishes (Antonucci *et al.*, 2009). Species identification becomes even more difficult to achieve after industry processing, when distinctive external traits are removed. Despite their similarity, seabream species have different organoleptic quality that corresponds to variable prices in fish markets. Consequently, substitution of commercially important sparid species seafood is common (Armani *et al*., 2015; Cawthorn *et al*., 2015). For example, the more expensive sparid species *Dentex* and *Pagrus* are often replaced with *D. gibbosus* (Katavic *et al.*, 2000).

All this premised, *Sparidae* species identification requires immediate solutions. Further development of molecular methods is instrumental in overcoming economically motivated mislabelling or substitution of high-value porgy species with less expensive or lower quality alternatives. Research on fish mitochondrial *DNA* (*mt*DNA, mitogenome) has led to substantial advances in the fields of species authentication and population biology (Miya *et al*., 2001). The main species-specific *DNA* sequences used as markers for fish species identification belong to the mitochondrial genes encoding ribosomal ***16S*** and ***12S*** subunits, cytochrome b (***Cytb***), and cytochrome c oxidase I (***COI***) (Pepe *et al*., 2007; Chiba *et al*., 2009; Cawthorn *et al.*, 2012; Giusti *et al*., 2019). *COI* is a moderately variable nucleotide sequence known as '*DNA* barcode' for taxa identification in many animal and vegetal groups (Hebert *et al*., 2003). However, current research shows that mitochondrial *DNA* markers perform well for certain species but may be less discriminating for others (Trotta *et al.*, 2005; Li *et al.*, 2018). Widely used *COI* markers may evolve too slowly and have low nucleotide sequence divergence in certain taxonomic groups, thus hampering discrimination of closely related fish species (*e.g.*, *Tuna*) (Rubinoff *et al*., 2006; Ward *et al*. 2009; Terio *et al*., 2010; Cawthorn *et*

al., 2015) when the gene does not contain effective regions for barcoding applications (Deagle *et al*., 2014).

Therefore, the ***COI*** barcode sequence cannot be used indiscriminately for the identification of all fish species. This situation has led to the formulation of the proposal to study and analyze the complete *mt*DNA sequence with the aim to identify mitochondrial markers or multiple marker approaches (Miya *et al*., 2015) with higher and more specific discrimination capacity.

For correct, quick and reliable species identification, it is essential to query the entire *mt*DNA genome for >100 bp mtDNA sequences with high interspecific divergence (Deagle *et al*., 2014; Shokralla *et al*., 2015).

At present, ***mt***DNA genomes of *Sparidae* fishes are not conspicuously represented in the scientific literature, making difficult to understand if currently used genetic markers are most effective for sparid species identification.

We have recently developed a fast and reliable method for sequencing the entire mitochondrial genome of fishes (Mascolo *et al*., 2019). To increase knowledge of the *Sparidae* family mitogenomes, we obtained the complete ***mt***DNA sequence of four commercially important seabream species (Ceruso *et al*., 2018a, 2018b; Mascolo *et al*., 2018a, 2018b). In this report, we analysed thirteen complete mitogenomes of **s**parid, nine available in GenBank, and four recently sequenced. The aim was to find the most effective gene fragments for sparid species discrimination.

9.3. Materials and Methods

9.3.1. Sparidae mtDNA genome data and sample processing

A total of thirteen complete *Sparidae* mitogenomes were analyzed and compared (**Tab. 1**). Nine mitochondrial genome sequences considered in this study were available in GenBank and belong to: *Acanthopagrus latus*, *Acanthopagrus schlegelii*, *Dentex tumifrons*, *Pagellus bogaraveo*, *Pagrus major*, *Pagrus auriga*, *Parargyrops edita*, *Rhabdosargus sarba*, and *Sparus aurata*. In addition, this research group has recently generated the complete mtDNA sequence of other four sparid species, *Dentex, Dentex gibbosus*, *Pagellus acarne* and *Pagellus erythrinus* (Ceruso *et al*., 2018a, 2018b; Mascolo *et al*., 2018a, 2018b). *D. dentex* specimen was captured in

middle Adriatic Sea, *D. gibbosus* and *P. auriga* fishes were caught in Thyrrhenian Sea, and *P. major* in the Northwest Pacific, northward to Japan. All these species were supplied by municipal fish market in Pozzuoli (Naples, Italy). *P. acarne, P. erythrinus*, *S. aurata*, and *P. bogaraveo* specimens were caught in middle Thyrrhenian sea and supplied at fish market in Salerno (Italy). Fish were frozen on board at -20°C and shipped in insulated boxes to the laboratory. Whole specimens were classified at the species level according to their anatomical and morphological features at the Department of Veterinary Medicine and Animal Production, University of Naples Federico II, Naples (Italy).

9.3.2 Total genomic DNA extraction.

Total genomic *DNA* was extracted from the skeletal muscle using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The extracted *DNA* was quantified using Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA). *DNA* concentration was 40 ng/µl and purity was in the range of 1.8-2.0 ratio at A260/A280. Electrophoretic analysis in 1% agarose gel was performed to examine the extracted *DNA*.

9.3.3. Mitochondrial DNA extraction, sequencing and assembly

A tissue sample was collected from the dorsal portion of the skeletal muscle of each specimen. Then, 25 mg of muscle were weighted and collected into 2 ml vials with 80 µL ice-cold 1% PBS and 5-mm Stainless Steel Beads (Qiagen, Valencia, CA, USA). Samples were grinded at -18 °C (TissueLyser Adapter Sets were cooled at -18 °C for 12 h) with oscillation frequencies of 30 Hz for 5 minutes in Qiagen Retsch Tissuelyser Mixer Mill Grinder DNA-RNA Homogenization 85210 MM301 (Qiagen, Cridersville, OH, USA). Next, lysates were pelleted by centrifugation at 8000×*g* for 8 min at room temperature. *Mt*DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions and eluted into 50 µL of elution buffer (Mascolo *et al.*, 2019). *DNA* quality and quantity was determined using Qubit™ 4 Fluorometer (Thermo Scientific Inc., Waltham, MA, USA) and Agilent 4200 TapeStation system (Agilent Technologies, Santa Clara, CA, USA) respectively (Genomix4life, Baronissi, Italy). Genome sequencing was performed at Bio-Fab Research (Rome, Italy), where libraries were prepared from 1 ug of each purified

*mt*DNA sample, using XT kit (Illumina, San Diego, CA, USA) according to manufacturer's instructions. Libraries were quantified using an Agilent Bioanalyzer 2100 chip HS *DNA* (Agilent Technologies, Santa Clara, CA, USA) and Qubit™ 4 Fluorometer (Thermo Scientific Inc., Waltham, MA, USA), and pooled in equimolar amounts, 4 nM being the final concentration. Pooled samples were then subjected to cluster generation and sequenced using an Illumina MiSeq 2500 System with v3 chemistry (Illumina, San Diego, CA, USA) in a 2x300 paired-end format at a final concentration of 16pM pmol. The quality of raw sequencing reads was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), followed by removal of Illumina adapters and trimming using Trimmomatic (v0.33) (Bolger *et al*., 2014). Trimming was performed at both 5' and 3' ends, eliminating bases with a Phred score of +33. Paired reads were processed simultaneously, and orphan reads were removed. High quality reads were used as input to perform mitogenome *de novo* assembly with a hybrid assembler method using Velvet (Zerbino *et al.*, 2008) and MIRA (Chevreux *et al.*, 1999). Scaffolds were obtained with Scaffolding Pre-Assemblies After Contig Extension (SSPACE) (Boetzer *et al.*, 2010), and annotation was made with Infernal: INFERence of RNA ALignments, and BLAST/BLASTX for coding sequences using RefSeq and MitoFish (Mitochondrial Genome Database of Fish; Iwasaki *et al.*, 2013).

9.3.4 Comparative analysis of mtDNA data

The complete mitogenome of the thirteen *Sparidae* fishes was subjected to comparative analysis using several bioinformatics tools in order to find the most effective genes for species identification.

The alignment image of the thirteen *Sparidae* complete mitogenomes considered in this study was obtained using Circos 0.69 software (Alikhan *et al.*, 2011). Hamming Distance algorithm was used to evaluate in percent the genetic dissimilarity among species and genes (Mitchell, 1998). Overall mean *p*-genetic distance analyses were conducted using the Maximum Composite Likelihood model (Kumar *et al*., 2018; Tamura *et al.*, 2004). The nucleotide sequence variability was determined by aligning gene-by-gene sequences of *Sparidae* species using MEGA 6.0 (Saitou *et al.*, 1987). Pairwise and multiple alignments were performed with

ClustalW. The parameters of gap open penalty and gap extension penalty were 15 and 6.66, respectively. The *DNA* weight matrix was the IUB matrix, and the transition weight was 0.5. The negative matrix was off and the delay divergent cutoff was 30%. Gene divergence was determined gene-by-gene in terms of variable sites using analysis modality of MEGA 6.0. The number of variable sites is displayed with the total number of sites (Variable sites/Total # of sites). Variable sites were calculated after removal of missing/gap sites from all the nucleotide and amino acid sequences. Protein coding gene (PCG) sequences were translated into amino acid sequences, whose variability was calculated with MEGA 6.0 with the same modality used for nucleotide sequences. Each coding sequence of the thirteen *Sparidae* mitogenomes was subjected to comparative analysis in order to find conserved regions (CR), using BioEdit software, version 7.2.5. The following parameters have been set: minimum segment length (actual for each sequence): 19, maximum average entropy: 0.2, gaps limited to 2 per segment, contiguous gaps limited to 1 in any segment.

9.3.5. NAD5 sequence analysis and comparison with databases

NAD5 primers were designed by eye after multiple alignment of the *Sparidae* complete mitochondrial genome sequences (Table 9.1) using BioEdit Sequence Alignment Editor (Hall, 1999). To verify melting temperature (Tm), secondary structure, self-annealing and inter-primer binding, Multiple Primer Analyzer was used (Thermo Fisher Scientific, Waltham, MA, USA). Then, the efficiency of *NAD5* primers for sparid species identification was tested *in silico*, with Unipro UGENE software (Kumar and Chordia, 2015). Amplicon sizes of 505 bp and 265 bp were amplified from fresh specimens of the following *Sparidae* species: *D. dentex, D. gibbosus, P. acarne, P. bogaraveo, P. erythrinus, P. auriga, P. major,* and *S. aurata*. PCR reactions were carried out in a 25 µl reaction volume containing 16.75 µl of sterile distilled water, 2.5 µl of dNTP mix (2 mM), 2.5 µl of $MgCl_2$ (10x), 0.5 µl of each primer (50 pMol/µl), 0.25 µl of 5 U/µl Taq (Roche, Basel, Switzerland), and 2 µl of *DNA* template containing approximately 80 ng of *DNA*. The thermal cycling profile was: an initial denaturation step of 95 °C for 5 min followed by 45 cycles of denaturation at 95 °C for 1 min, annealing at 55 °C (505 bp fragment) and 51 °C (265 bp fragment) for 1 min, and extension at 72 °C for 1 min. *COI*

primers (Armani *et al*., 2015) were used as positive control. PCR products were electrophoresed on a 1% agarose gel, visualized via ultraviolet transillumination and purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany).

Amplicons were sequenced by Sanger method with the Automated Capillary Electrophoresis Sequencer 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA) at the Molecular Biology Service at the Stazione Zoologica Anton Dohrn. *NAD5* sequences were analyzed using the BioEdit Sequence Alignment Editor (Hall, 1999). All the obtained sequences were used to run a BLAST analysis on GenBank for species identification and to assess the concordance between morphological and molecular analyses (Ratnasingham & Hebert, 2013).

Table 9.1. List of the Sparidae species with complete mtDNA sequence considered in this study.

N°	SPECIES	AC. NUMBER	FAO FISHING AREAS	REFERENCES
1	*Acanthopagrus latus*	NC_010977.1	FAO 71	Xia *et al*., 2008
2	*Acanthopagrus schlegelii*	JQ_746035.1	FAO 71	Shi *et al*., 2012
3	*Dentex*	MG_727892.1	FAO 37	Ceruso *et al*., 2018a
4	*Dentex gibbosus*	MG_653593.1	FAO 37	Mascolo *et al*., 2018b
5	*Dentex tumifrons*	NC_029479.1	FAO 71	Zeng *et al*., unpublished
6	*Pagellus acarne*	MG_736083.1	FAO 37	Mascolo *et al*., 2018a
7	*Pagellus bogaraveo*	NC_009502.1	FAO 27	Ponce *et al*., 2008
8	*Pagellus erythrinus*	MG_653592.1	FAO 37	Ceruso *et al*., 2018b
9	*Pagrus auriga*	AB_124801.1	FAO 37	Ponce *et al*., unpublished
10	*Pagrus major*	NC_003196.1	Farmed and supplied from Andalusia (Spain) fish market	Miya *et al*., 2001
11	*Parargyrops edita*	EF_107158.1	FAO 71	Xia *et al*., 2007
12	*Rhabdosargus sarba*	KM_272585.1	Farmed Daya Bay Aquaculture Center, Guangdong (China)	Li *et al*., 2016
13	*Sparus aurata*	LK_022698.1	Farmed and supplied from Jaffa (Israel) fish market	Dray *et al*., 2014

9.4 Results and Discussion

9.4.1 DNA extraction and NGS sequencing

For Next Generation Sequencing (NGS), *mt*DNA was separated from the nuclear *DNA* by using a Miniprep Kit (see M&M). This approach yielded good quantity and quality *mt*DNA samples from muscle tissue (Table 9.2). Quality control of NGS raw sequencing data eliminated low-quality portions while preserving high quality portion of reads. Then, high quality reads were normalized and assembled for each sample separately by calculating coverage using *de novo* scaffolds. NGS data and read numbers are reported in Table 9.3.

Table 9.2. MtDNA sample quantity and quality.

	D. dentex	*D. gibbosus*	*P. acarne*	*P. erythrinus*
DIN	4.7	5.5	4.2	6.2
DNA concentration(ng/µl)	3.92	3.59	6.36	6.24

DIN: DNA Integrity Number.

Table 9.3. NGS analysis

	D. dentex	***D. gibbosus***	***P. acarne***	***P. erythrinus***
NGS reads number	2567324	2300388	3487934	3532942
STATISTICS FOR DE NOVO ASSEMBLY				
Longest scaffold length	16907	16771	17127	12091
Raw reads	2626640	2300388	3487934	3659022
EC reads	2567324	2247342	3377514	2414609
% EC passing reads	97.74	97.69	96.83	65.99
Raw nt	529475558	547583072	798005139	918414522
STATISTICS FOR COVERAGE				
Raw Cov	227.32	242.26	170.84	771.69
EC cov	175.95	169.14	121.52	266.04
Median cov	48	19	27	17
10th percentile cov	6	3	4	3
Bases>= Q40 % GC	2114109	1851830	3935921	951847
% GC content	41.7	43.2	43.2	42.2

. Summary statistics for *de novo* assembly includes: raw reads - number of raw reads used as input; EC reads - number of reads used for assembly, after quality control and Error Correction (EC); % EC passing reads - percentage of the raw reads that the EC reads represent; raw nt - number of nucleotides used as input. Summary statistics for coverage includes: raw cov - average depth of sequence coverage provided by the raw data; EC cov - coverage after quality control and error correction; median cov - median actual depth of coverage in the assembly; 10th percentile cov - 10th percentile depth of coverage -- 90% of sites have greater coverage; Bases >= Q40 % GC - number of bases that have a PHRED-scale quality greater or equal to Q40.

9.4.2. Sparidae mtDNA analysis

9.4.2.1. Sparidae mitogenome structure

All thirteen complete sparid *mt*DNA genomes analyzed in this study contained 37 genes (13 protein coding, 22 *tRNA*, and 2 *rRNA* genes) and 2 non-coding regions (*L*-origin of replication and *D-loop*), as typically found in fish species (Satoh *et al.*, 2016). Mitochondrial genes of the studied sparid species were arranged in the same order as in the typical vertebrate mitogenome (Wang et al. 2008). Most mitochondrial genes were encoded on the heavy strand, with only *NAD dehydrogenase subunit 6* (*NAD6*) and eight *tRNA* genes [$tRNA^{Gln}$, $tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Tyr}$, $tRNA^{Ser}$ *(UCN)*, $tRNA^{Glu}$, $tRNA^{Pro}$] encoded on the light strand. The *12S* and *16S rRNA* genes were located between the $tRNA^{Phe}$ (GAA) and $tRNA^{Leu}$ (TAA) genes and were separated by the $tRNA^{Val}$ gene as in other vertebrates (Li *et al.*, 2016). The control region is located between $tRNA^{Pro}$ (TGG) and $tRNA^{Phe}$ (GAA). The non-coding region (*L*-strand origin of replication) is located between $tRNA^{Asn}$ (GTT) and $tRNA^{Cys}$ (GCA). Mitogenomes of the newly sequenced *Sparidae* species (*D. dentex*, *D. gibbosus*, *P. acarne*, and *P. erythrinus*) possess the same organization described above (Tables 9.4-9.7). Likewise, their size and base composition are similar to those determined for fish within the *Sparidae* family and other teleost families (Broughton *et al.*, 2001; Inoue *et al.*, 2001b; Miya *et al.*, 2001) (Table 9.8). Genome length of sparid mitochondria varied from 16486 (*P. acarne*) to 17031 bp (*P. major*). Protein coding and non-coding gene lengths are shown in Tables 9.9 and 9.10.

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

*Table 9.4. Mitochondrial genome organization of Dentex. * Numbers correspond to nucleotides contributing to gene length variation. Negative numbers indicate overlapping*

GENE	ANTICODON	POSITION	SIZE	STRAND	START	END	3'*
tRNA-Phe	GAA	11559-11626	68	H			0
12S rRNA		**11627-12581**	**955**	**H**			**-1**
tRNA-Val	TAC	12581-12652	72	H			0
16s rRNA		**12653-14347**	**1695**	**H**			**0**
tRNA-Leu (UUR)	TAA	14348-14420	73	H			0
Nad1		**14421-15395**	**975**	**H**	**ATG**	**TAA**	**+5**
tRNA-Ile	GAT	15401-15470	70	H			-1
tRNA-Gln	**TTG**	**15470-15540**	**71**	**L**			**-1**
tRNA-Met	CAT	15540-15609	70	H			0
Nad2		**15610-16656**	**1047**	**H**	**ATG**	**TAA**	**+2**
tRNA-Trp	TCA	3-72	70	H			0
tRNA-Ala	**TGC**	**73-141**	**69**	**L**			**+1**
tRNA-Asn	GTT	143-215	73	L			+5
Rep-origin L strand		**221-255**	**35**	**-**			**0**
tRNA-Cys	GCA	256-321	66	L			0
tRNA-Tyr	**GTA**	**322-391**	**70**	**L**			**+1**
COI		393-1958	1566	H	GTG	AGG	-13
tRNA-Ser(UCN)	**TGA**	**1946 -2016**	**71**	**L**			**+2**
tRNA-Asp	GTC	2019-2091	73	H			+8
COII		**2100-2790**	**691**	**H**	**ATG**	**T**	**0**
tRNA-Lys	TTT	2791-2864	74	H			+1
Atp8		**2866-3033**	**168**	**H**	**ATG**	**TAA**	**-10**
Atp6		3024-3707	684	H	ATG	TAA	-1
COIII		**3707-4492**	**786**	**H**	**ATG**	**TAA**	**-1**
tRNA-Gly	TCC	4492-4563	72	H			0
Nad3		**4564-4914**	**351**	**H**	**ATG**	**TAG**	**-2**
tRNA-Arg	TCG	4913-4984	72	H			0
Nad4l		**4985-5281**	**297**	**H**	**ATG**	**TAA**	**-7**
Nad4		5275-6655	1381	H	ATG	T	0
tRNA-His	**GTG**	**6656-6724**	**69**	**H**			**0**
tRNA-Ser (AGY)	GCT	6725-6792	68	H			+6
tRNA-Leu (CUN)	**TAG**	**6799-6871**	**73**	**H**			**0**
Nad5		6872-8710	1839	H	ATG	TAA	-4
Nad6		**8707-9228**	**522**	**L**	**ATG**	**TAA**	**0**
tRNA-Glu	TTC	9229-9297	69	L			+4
Cytb		**9302-10442**	**1141**	**H**	**ATG**	**T**	**0**
tRNA-Thr	TGT	10443-10516	74	H			-1
tRNA-Pro	**TGG**	**10516-10586**	**71**	**L**			**0**
Control region (D-loop)		15587-11558	973	-			0

nucleotides between adjacent genes; H and L indicate genes transcribed on the heavy and light strands, respectively

*Table 9.5. Mitochondrial genome organization of Pagellus acarne. * Numbers correspond to nucleotides contributing to gene length variation. Negative numbers indicate*

GENE	ANTICODON	POSITION	SIZE	STRAND	START	END	3'*
tRNA-Phe	GAA	27-94	68	H			0
12S rRNA		**95-1046**	**952**	**H**			**0**
tRNA-Val	TAC	1046-1117	72	H			-1
16s rRNA		**1118-2814**	**1697**	**H**			**0**
tRNA-Leu (UUR)	TAA	2815-2887	73	H			0
Nad1		**2888-3862**	**975**	**H**	**ATG**	**TAA**	**0**
tRNA-Ile	GAT	3867-3936	70	H			+6
tRNA-Gln	**TTG**	**3936-4006**	**71**	**L**			**-1**
tRNA-Met	CAT	4006-4075	70	H			-1
Nad2		**4076-5122**	**1047**	**H**	**ATG**	**TAG**	**0**
tRNA-Trp	TCA	5121-5190	70	H			0
tRNA-Ala	**TGC**	**5191-5259**	**69**	**L**			**0**
tRNA-Asn	GTT	5261-5333	73	L			+2
Rep-origin L strand		**5334-5370**	**38**	**-**			**0**
tRNA-Cys	GCA	5371-5437	67	L			0
tRNA-Tyr	**GTA**	**5438-5507**	**70**	**L**			**0**
COI		5509-7074	1566	H	GTG	AGG	+2
tRNA-Ser(UCN)	**TGA**	**7062-7132**	**71**	**L**			**-12**
tRNA-Asp	GTC	7136-7208	73	H			+4
COII		**7217-7907**	**691**	**H**	**ATG**	**T**	**+9**
tRNA-Lys	TTT	7908-7981	74	H			0
Atp8		**7983-8147**	**165**	**H**	**ATG**	**TAG**	**+2**
Atp6		8141-8824	684	H	ATG	TAA	-7
COIII		**8824-9609**	**786**	**H**	**ATG**	**TAA**	**-1**
tRNA-Gly	TCC	9609-9680	71	H			-1
Nad3		**9681-10029**	**349**	**H**	**ATG**	**T**	**0**
tRNA-Arg	TCG	10030-10101	72	H			0
Nad4l		**10102-10398**	**297**	**H**	**ATG**	**TAA**	**0**
Nad4		10392-11772	1381	H	ATG	T	-7
tRNA-His	**GTG**	**11773-11841**	**69**	**H**			**0**
tRNA-Ser (AGY)	GCT	11842-11910	69	H			0
tRNA-Leu (CUN)	**TAG**	**11918-11990**	**73**	**H**			**+9**
Nad5		11991-13829	1839	H	ATG	TAA	0
Nad6		**13826-14347**	**522**	**L**	**ATG**	**TAG**	**-4**
tRNA-Glu	TTC	14348-14416	69	L			0
Cytb		**14422-15562**	**1141**	**H**	**ATG**	**T**	**+7**
tRNA-Thr	TGT	15563-15634	72	H			-2
tRNA-Pro	**TGG**	**15634-15704**	**71**	**L**			**-1**
Control region (D-loop)		15705-26	808	-			0

overlapping nucleotides between adjacent genes; H and L indicate genes transcribed on the heavy and light strands, respectively.

*Table 9.6. Organization of the Dentex gibbosus mitochondrial genome. * Numbers correspond to nucleotides contributing to gene length variation. Negative numbers indicate overlapping nucleotides between adjacent genes; H and L indicate genes transcribed on the heavy and light strands, respectively.*

GENE	ANTICODON	POSITION	SIZE	STRAND	START	END	3'*
tRNA-Phe	GAA	1-68	68	H			0
12S rRNA		**69-1022**	**955**	**H**			**0**
tRNA-Val	TAC	1023-1094	72	H			0
16s rRNA		**1095-2790**	**1696**	**H**			**0**
tRNA-Leu (UUR)	TAA	2791-2863	73	H			0
Nad1		**2864-3838**	**975**	**H**	**ATG**	**TAA**	**+7**
tRNA-Ile	GAT	3844-3913	70	H			-1
tRNA-Gln	**TTG**	**3913-3983**	**71**	**L**			**-1**
tRNA-Met	CAT	3983-4052	70	H			0
Nad2		**4053-5098**	**1047**	**H**	**ATG**	**TA**	**0**
tRNA-Trp	TCA	5099-5168	70	H			0
tRNA-Ala	**TGC**	**5169-5237**	**69**	**L**			**+3**
tRNA-Asn	GTT	5239-5311	73	L			0
Rep-origin L strand		**5312-5350**	**40**	**-**			**0**
tRNA-Cys	GCA	5351-5416	66	L			0
tRNA-Tyr	**GTA**	**5417-5486**	**70**	**L**			**+3**
COI		5488-7053	1566	H	GTG	AGG	-13
tRNA-Ser (UCN)	**TGA**	**7041-7111**	**71**	**L**			**+4**
tRNA-Asp	GTC	7114-7186	73	H			+10
COII		**7195-7885**	**691**	**H**	**ATG**	**T**	**0**
tRNA-Lys	TTT	7886-7959	74	H			+2
Atp8		**7961-8128**	**168**	**H**	**ATG**	**TAA**	**-9**
Atp6		8119-8801	684	H	ATG	TAA	0
COIII		**8802-9586**	**786**	**H**	**ATG**	**TA**	**0**
tRNA-Gly	TCC	9587-9658	72	H			0
Nad3		**9659-10007**	**351**	**H**	**ATG**	**T**	**0**
tRNA-Arg	TCG	10008-10078	71	H			0
Nad4l		**10079-10375**	**297**	**H**	**ATG**	**TAA**	**-7**
Nad4		10369-11749	1380	H	GTG	T	0
tRNA-His	**GTG**	**11750-11818**	**69**	**H**			**0**
tRNA-Ser (AGY)	GCT	11819-11886	68	H			+8
tRNA-Leu (CUN)	**TAG**	**11893-11965**	**73**	**H**			**0**
Nad5		11966-13804	1839	H	ATG	TAA	-4
Nad6		**13801-14322**	**522**	**L**	**ATG**	**TAA**	**0**
tRNA-Glu	TTC	14323-14391	69	L			+6
Cytb		**14396-15536**	**1141**	**H**	**ATG**	**T**	**0**
tRNA-Thr	TGT	15537-15610	74	H			-1
tRNA-Pro	**TGG**	**15610-15679**	**70**	**L**			**0**
Control region (D-loop)		15680-16771	1091	-			0

*Table 9.7. Organization of the Pagellus erythrinus mitochondrial genome. * Numbers correspond to nucleotides contributing to gene length variation. Negative numbers*

GENE	ANTICODON	POSITION	SIZE	STRAND	START	END	3'*
tRNA-Phe	GAA	161-228	68	H			0
12S rRNA		**229-1183**	**955**	**H**			**0**
tRNA-Val	TAC	1183-1254	72	H			-1
16s rRNA		**1294-2950**	**1696**	**H**			**0**
tRNA-Leu (UUR)	TAA	2951-3023	73	H			0
Nad1		**3024-3998**	**975**	**H**	**ATG**	**TAA**	**0**
tRNA-Ile	GAT	4004-4073	70	H			+6
tRNA-Gln	**TTG**	**4073-4143**	**71**	**L**			**-1**
tRNA-Met	CAT	4143-4212	70	H			-1
Nad2		**4213-5258**	**1046**	**H**	**ATG**	**TA**	**0**
tRNA-Trp	TCA	5259-5328	70	H			0
tRNA-Ala	**TGC**	**5329-5397**	**69**	**L**			**0**
tRNA-Asn	GTT	5399-5471	73	L			+1
Rep-origin L strand		**5472-5506**	**31**	**-**			**0**
tRNA-Cys	GCA	5507-5572	66	L			0
tRNA-Tyr	**GTA**	**5573-5642**	**69**	**L**			**0**
COI		5644-7209	1566	H	GTG	AGG	+2
tRNA-Ser(UCN)	**TGA**	**7197-7267**	**71**	**L**			**-12**
tRNA-Asp	GTC	7270-7342	73	H			+2
COII		**7351-8041**	**691**	**H**	**ATG**	**T**	**+8**
tRNA-Lys	TTT	8042-8115	74	H			0
Atp8		**8117-8284**	**168**	**H**	**ATG**	**TAA**	**+1**
Atp6		8275-8958	684	H	ATG	TAA	-9
COIII		**8958-9743**	**786**	**H**	**ATG**	**TAA**	**-1**
tRNA-Gly	TCC	9743-9813	71	H			-1
Nad3		**9814-10164**	**351**	**H**	**ATG**	**TAG**	**0**
tRNA-Arg	TCG	10163-10233	71	H			-2
Nad4l		**10234-10530**	**297**	**H**	**ATG**	**TAA**	**0**
Nad4		10524- 11904	1380	H	GTG	T	-7
tRNA-His	**GTG**	**11905-11973**	**69**	**H**			**0**
tRNA-Ser (AGY)	GCT	11974-12041	68	H			0
tRNA-Leu (CUN)	**TAG**	**12048-12120**	**73**	**H**			**+6**
Nad5		12121-13959	1839	H	ATG	TAG	0
Nad6		**13956-14477**	**522**	**L**	**ATG**	**TAA**	**-4**
tRNA-Glu	TTC	14478-14546	69	L			0
Cytb		**14551-15691**	**1141**	**H**	**ATG**	**T**	**+4**
tRNA-Thr	TGT	15692- 15765	74	H			0
tRNA-Pro	**TGG**	**15765-15834**	**70**	**L**			**-1**
Control region (D-loop)		15835-160	914	-			0

Species	Base composition %					bp total
	T	**C**	**A**	**G**	**A/T**	
Acanthopagrus latus	27,3	28,8	26,8	17,1	54,1	16609
Acanthopagrus schlegelii	27,9	27,9	28,0	16,2	55,9	16649
Dentex	26,9	28,7	27,5	16,9	53,59	16656
Dentex gibbosus	26,05	28,6	27,8	16,5	54,8	16771
Dentex tumifrons	26,4	29,5	27,1	17,1	53,4	16616
Pagellus acarne	26,6	29,0	26,8	17,6	53,4	16486
Pagellus bogaraveo	26,7	29,1	26,9	17,2	53,7	16941
Pagellus erythrinus	27,5	28,2	27,5	16,8	54,9	16828
Pagrus auriga	26,9	28,7	27,7	16,7	54,6	16628
Pagrus major	27,0	28,8	27,5	16,7	54,5	17031
Parargyrops edita	27,0	29,0	26,0	18,0	54,3	16640
Rhabdosargus sarba	27,3	28,8	26,8	17,1	54,1	16644
Sparus aurata	26,9	28,8	27,6	16,7	53,3	16652

indicate overlapping nucleotides between adjacent genes; H and L indicate genes transcribed on the heavy and light strands, respectively.

Table 9.8. Base composition of Sparidae mtDNA.

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

Table 9.9. Length (bp) of protein-coding genes.

Species	NAD1	NAD2	COI	COII	ATP8	ATP6	COIII	NAD3	NAD4L	NAD4	NAD5	NAD6	Cyt-b
Acanthopagrus latus	975	1046	1566	691	165	684	786	349	297	1381	1839	522	1120
Acanthopagrus schlegelii	975	1047	1566	691	165	684	785	349	297	1381	1839	522	1141
Dentex dentex	975	1047	1566	691	168	684	786	351	297	1381	1839	522	1141
Dentex gibbosus	975	1047	1566	691	168	684	786	351	297	1381	1839	522	1141
Dentex tumifrons	975	1047	1566	691	168	684	785	349	297	1381	1839	522	1141
Pagellus acarne	975	1047	1566	691	165	684	786	349	297	1381	1839	522	1141
Pagellus bogaraveo	975	1047	1566	691	165	684	786	349	297	1380	1839	522	1141
Pagellus erythrinus	975	1047	1566	691	168	684	786	351	297	1381	1839	522	1141
Pagrus auriga	975	1047	1566	691	168	684	786	351	297	1381	1839	522	1141
Pagrus major	975	1047	1563	691	168	683	785	349	297	1381	1839	522	1141
Pargyrops edita	975	1047	1563	691	168	684	785	349	297	1381	1839	522	1141
Rhabdosargus sarba	975	1050	1566	691	165	684	785	349	297	1381	1842	522	1141
Sparus aurata	975	1046	1566	691	165	683	785	349	297	1381	1842	522	1141

Table 9. 10. Length (bp) of non-coding genes.

Species	*D-loop*	*L-origin*
Acanthopagrus latus	943	30
Acanthopagrus schlegelii	945	33
Dentex	973	35
Dentex gibbosus	1091	40
Dentex tumifrons	936	35
Pagellus acarne	808	38
Pagellus bogaraveo	1195	124
Pagellus erythrinus	914	31
Pagrus auriga	948	53
Pagrus major	1354	39
Parargyrops edita	964	36
Rhabdosargus sarba	978	36
Sparus aurata	982	36

9.4.2.2. MtDNA comparative data

The first alignment of the thirteen *Sparidae* complete mitogenomes currently present in Genbank showed that the similarity of the translated sequences is higher than 70% compared to the reference mitogenome (*Sparus aurata*). This result suggests, as expected, that the thirteen species are very similar each other.

In particular, regarding the structure of the thirteen mitogenomes, among 37 mitochondrial protein-coding genes (PCG), four genes (*NAD1*, *NAD4L*, *NAD6*, *COII*) showed no changes in nucleotide sequence length. The *Cytb* gene sequence was shorter only in *A. latus* (1120 instead of 1141 bp) due to a 5' 12 bp deletion and a 3' 9 bp deletion (Figure 9.1). The *NAD2* gene is 1046/1047 bp long in all sparid mitogenomes except for *R. sarba* (1050 bp) (Table 9.9). *COI*, *ATP8* and *NAD5* genes showed a 3 bp insertion in some sparids, with sequence length ranging from 1563 to 1566 bp (*COI*), from 165 to 168 bp (*ATP8*), and from 1839 to 1842 bp (*NAD5*). Some species showed a 1-2 bp deletion in *ATP6*, *COIII*, *NAD4* and *NAD3* genes due to the presence of an incomplete stop codon (Satoh *et al.*, 2016). As expected, sequence length of non-coding regions showed a variation more consistent than that of coding genes. Remarkably, the origin of light-strand replication (OL) of *P. bogaraveo* is 2.5/3.0 times longer than in other sparids (124 bp *vs.* 30-53 bp). The displacement loop (*D-loop*), or non-coding control region, contains the origin of heavy-strand replication. The length of the *D-loop* has changed repeatedly and independently across sparid orders and genera, with the smallest one in *P. acarne* (808 bp) and the longest ones in *P. bogaraveo* (1195 bp) and *P. major* (1354 bp) (Table 9.10).

All protein-coding genes begin with an ATG start codon, except for *COI* that starts with GTG in all the studied species, and *NAD4* that starts with GTG in *D. gibbosus*, *D. tumifrons*, *P. erythrinus*, *P. major*, and *P. edita* (Table 9.11). Stop codons were of 6 types, *i.e.* TAA (*NAD1*, *NAD2*, *ATP8*, *ATP6*, *COIII*, *NAD4L*, *NAD5*, *NAD6*), TAG (*NAD1*, *NAD2*, *ATP8*, *NAD3*, *NAD5*, *NAD6*), TA (*NAD2*, *ATP6*, *COIII*), AGA (*NAD2*, *NAD4*), AGG (*COI*), and T (*COII*, *NAD3*, *NAD4*, *Cytb)* (Table 9.11).

The mitochondrial phylogenetic tree of *Sparidae* species (Figure 9.7, supplementary material) shows a close relationship between *D. dentex* and *D. gibbosus*, in agreement with previous work (Day, 2002; Orrell *et* al., 2002; Chiba *et al*., 2009). Also considering the genetic dissimilarity among species, the Hamming Distance comparison showed that *D. dentex* and *D.*

gibbosus are the closest species, with 6% of genetic dissimilarity (Figure 9.2). Therefore, species discrimination in these two co-generic species may be more difficult than in other *Sparidae* species. Hamming Distance analysis was also carried out gene by gene, with the aim to find genes with higher divergence among species. Results showed that *D. dentex* and *D. gibbosus* have more genetic dissimilarity in *NAD4* (9%), *NAD5* (10%), and *NAD2* (12%) than in *COI* and *Cytb* (8%). Accordingly, analysis of the *p*-genetic distance among all mitogenomes showed that the *NAD* group genes display more sequence distance than *COI* and *Cytb* genes (Figure 9.3).

In addition, higher values of nucleotide sequence variability were observed in the *NAD* gene group (*NAD1* 39%, *NAD2* 50%, *NAD3* 39%, *NAD4L* 39%, *NAD4* 43%, *NAD5* 41%, *NAD6* 44%) (Table 9.12**)**. Of note, *NAD* gene sequence variability is higher compared with classical markers currently used for species identification, *i.e. Cytb* (36%), *COI* (32%), *16S* (24%) and *12S* (21%). About amino acidic variability, *NAD2* (35%) and *ATP8* (32%) have the higher values (Table 9.12).

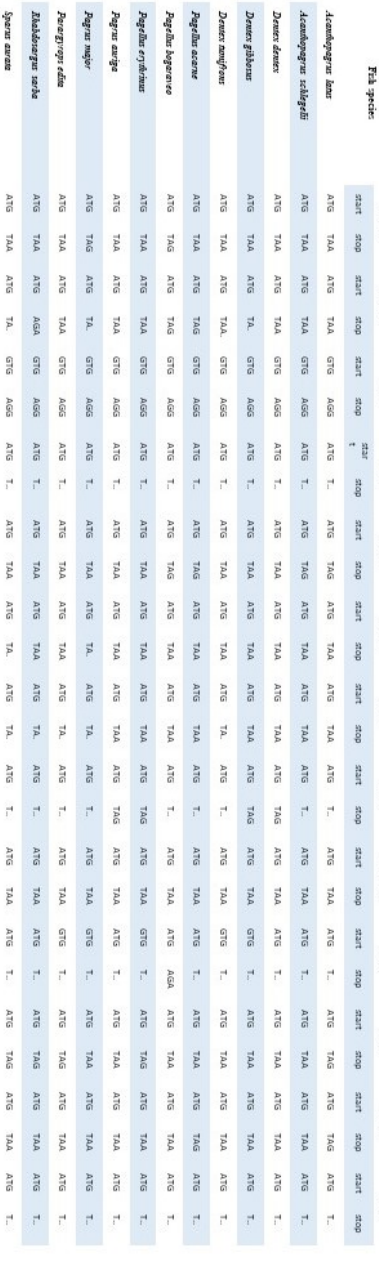
A comparative analysis of *Sparidae mt*DNA sequences was carried out in order to find conserved regions (CR) that may be used to design primers. This analysis showed that the higher number of CRs occurred within four genes, *i.e. Cytb* (11 CR), *NAD5* (10), *16S* (10) and *COI* (8) (Table 9.13). Among them, the *COI* coding sequence has 1398 bp conserved on a total of 1566 (89%), featuring the longest CR (1061 bp) found in this analysis. Analogous situation occurs in the *16S* gene, with 10 CRs containing 1621 bp distributed over a total of 1697 bp (95%). *NAD5* and *Cytb* show a slightly different situation. *NAD5* presents 10 conserved areas that contain only 378 conserved nucleotides on a total of 1842 (20%). However, all *NAD5* CRs vary in length between 23 and 74 bp, dimensions compatible with typical primer length (18-30 bp). On the other hand, *Cytb* possesses 11 CRs that range between 23 and 106 bp and compose 49% of the gene nucleotide coding sequence length.

Although *NAD* genes show higher levels of sequence variability among sparids, it is only *NAD5* that is associated with a congruous number of CRs to be targeted for primer design evaluation.

Therefore, the *NAD5* gene appears more appropriate for barcoding as it provides species-level information on *Sparidae* family and is associated with conserved areas that could be functional for primer design (Figure 9.4). Among the most used markers for fish species identification (*Cytb*,

COI, *16S* and *12S*), the *Cytb* gene also shows some conserved zones followed by areas with species-specific sequence variability, therefore useful for the purposes of species characterization. Regarding classical barcoding markers, it is important to note that *COI* and *16S* are particularly conserved among the studied *Sparidae*, which could represent a limit for species discrimination. Indeed, the *12S* gene does not seem to be very effective for *Sparidae* molecular barcoding, due to its low coding sequence variability (21%) and low number of CRs.

Table 9.11. Start/stop codons of protein-coding genes



Fish species	NAD1		NAD2		COI		COII		ATP8		ATP6		COIII		NAD3		NAD4L		NAD4		NAD5		NAD6		Cyb	
	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop	start	stop
Acanthopagrus latus	ATG	TAA	ATG	TAA	GTG	AGG	ATG	T..	ATG	TAG	ATG	TAA	ATG	TAA	ATG	T..	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAG	ATG	T..
Acanthopagrus schlegelii	ATG	TAA	ATG	TAA	GTG	AGG	ATG	T..	ATG	TAG	ATG	TAA	ATG	TAA	ATG	T..	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAA	ATG	T..
Dentex dentex	ATG	TAA	ATG	TA.	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAA	ATG	T..
Dentex gibbosus	ATG	TAA	ATG	TA.	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAA	ATG	T..
Dentex maufeni	ATG	TAA	ATG	TAA.	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TA.	ATG	T..	ATG	TAA	GTG	T..	ATG	TAA	ATG	TAA	ATG	T..
Pagellus acarne	ATG	TAA	ATG	TAG	GTG	AGG	ATG	T..	ATG	TAG	ATG	TAA	ATG	TAA	ATG	T..	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAG	ATG	T..
Pagellus bogaraveo	ATG	TAG	ATG	TAG	GTG	AGG	ATG	T..	ATG	TAG	ATG	TAA	ATG	TAA	ATG	T..	ATG	TAA	ATG	AGA	ATG	TAA	ATG	TAA	ATG	T..
Pagellus erythrinus	ATG	TAA	ATG	TAA	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	GTG	T..	ATG	TAG	ATG	TAA	ATG	T..
Pagrus auriga	ATG	TAA	ATG	TAA	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TAA	ATG	TAG	ATG	TAA	ATG	T..	ATG	TAA	ATG	TAA	ATG	T..
Pagrus major	ATG	TAG	ATG	TA.	GTG	AGG	ATG	T..	ATG	TAA	ATG	TA.	ATG	TA.	ATG	T..	ATG	TAA	GTG	T..	ATG	TAA	ATG	TAA	ATG	T..
Pargyrops edita	ATG	TAA	ATG	TAA	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TA.	ATG	T..	ATG	TAA	GTG	T..	ATG	TAG	ATG	TAA	ATG	T..
Rhabdosargus sarba	ATG	TAA	ATG	AGA	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TA.	ATG	T..	ATG	TAA	ATG	T..	ATG	TAG	ATG	TAA	ATG	T..
Sparus aurata	ATG	TAA	ATG	TA.	GTG	AGG	ATG	T..	ATG	TAA	ATG	TAA	ATG	TA.	ATG	T..	ATG	TAA	ATG	T..	ATG	TAG	ATG	TAA	ATG	T..

Table 9.12. Nucleotide and aminoacid sequence variability of Sparidae mtDNA of Acanthopagrus latus, Acanthopagrus schlegelii, Dentex, Dentex gibbosus, Dentex tumifrons, Pagellus acarne, Pagellus bogaraveo, Pagellus erythrinus Pagrus auriga, Pagrus major, Parargyrops edita, Rhabdosargus sarba, Sparus aurata. Number of variable nucleotide/amino acid sites of single genes/proteins on total sequence alignment. The total nucleotide and amino acid numbers of single genes and proteins mean the numbers obtained after alignment.

Gene	Variable sites/Total sites			
	Nucleotide		**Amino acid**	
NAD1	391/975	39%	41/325	12%
NAD2	528/1050	50%	125/350	35%
COI	510/1567	32%	32/522	6%
COII	235/691	34%	27/230	11%
COIII	267/786	33%	35/262	13%
ATP8	72/168	42%	18/56	32%
ATP6	313/684	45%	59/228	25%
NAD3	137/351	39%	16/117	13%
NAD4L	116/297	39%	14/99	14%
NAD4	607/1386	43%	94/462	20%
NAD5	769/1842	41%	130/614	21%
NAD6	234/522	44%	48/174	27%
Cytb	413/1141	36%	35/380	10%
12s	209/967	21%	/	/
16s	419/1714	24%	/	/

Table 9.13. Distribution of conserved regions in Sparidae mtDNA genes.

Gene	N° of conserved regions	Conserved sequence length (bp)	Total bp conserved/Total gene length
NAD1	1	29	29/975
NAD2	1	29	29/1047
COII	8	1061, 59, 26, 25, 119, 38, 70	1398/1566
COIII	3	437, 20, 187	644/691
ATP8	3	30, 42, 20	92/168
ATP6	3	46, 29, 28	103/684
COIIII	4	356, 38, 74, 125	593/786
NAD3	1	29	29/351
NAD4L	2	29-23	52/297
NAD4	1	23	23/1381
NAD5	10	25, 24, 26, 47, 50, 74, 53, 32, 24, 23	378/1842
NAD6	0	/	/
Cytb	11	106, 51, 47, 44, 83, 29, 23, 69, 29, 47, 35	563/1141
12s	6	129, 64, 126, 308, 110, 180	917/955
16s	10	128, 107, 23, 242, 145, 141, 107, 344, 330, 54	1621/1697

Figure 9.1. Sequence identity on alignments of two extracts of the Cytb gene in Sparidae mtDNA (from 1 to 68 bp and from 209 to 283 bp). In Acanthopagrus latus, this gene is 1120 bp long due to sequence gaps at 5' (12 to 23 bp) and 3' (232 to 240 bp) ends. This image was obtained with Unipro UGENE 1.29.

Frauds and fish species authentication: study of the complete mitochondrial genome of some *Sparidae* to provide specific barcode markers

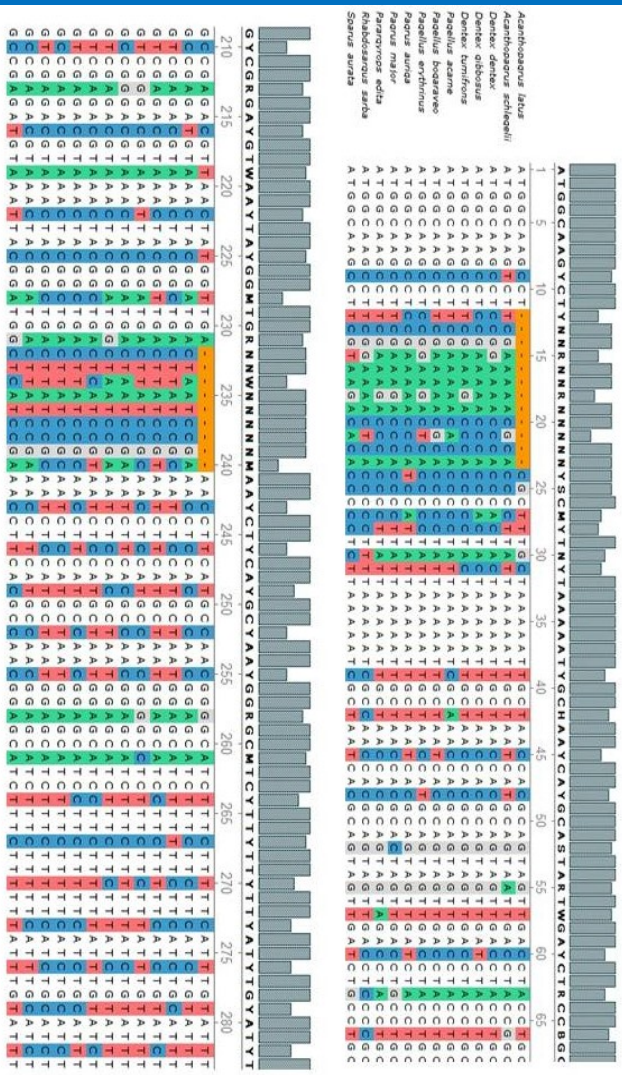


Figure 9.2. Hamming genetic dissimilarity in Sparidae full mtDNA genome comparison.

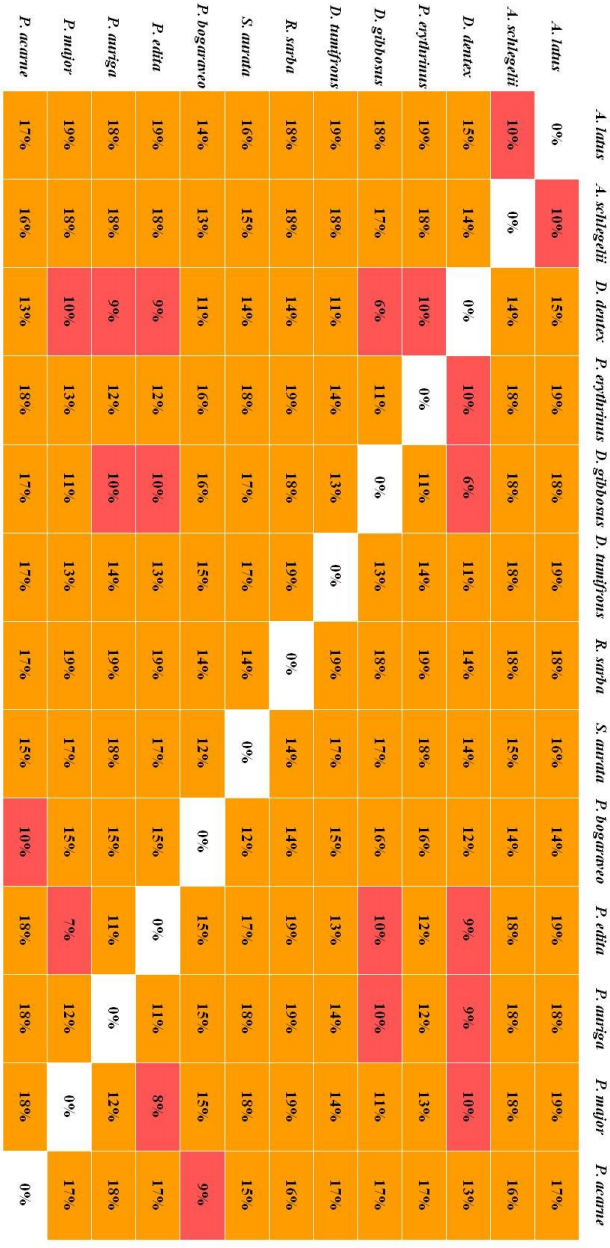


Figure 9.3. Overall mean p-genetic distance in Sparidae gene-by-gene comparison. Number of base substitutions per site from averaging over all sequence pairs is shown.

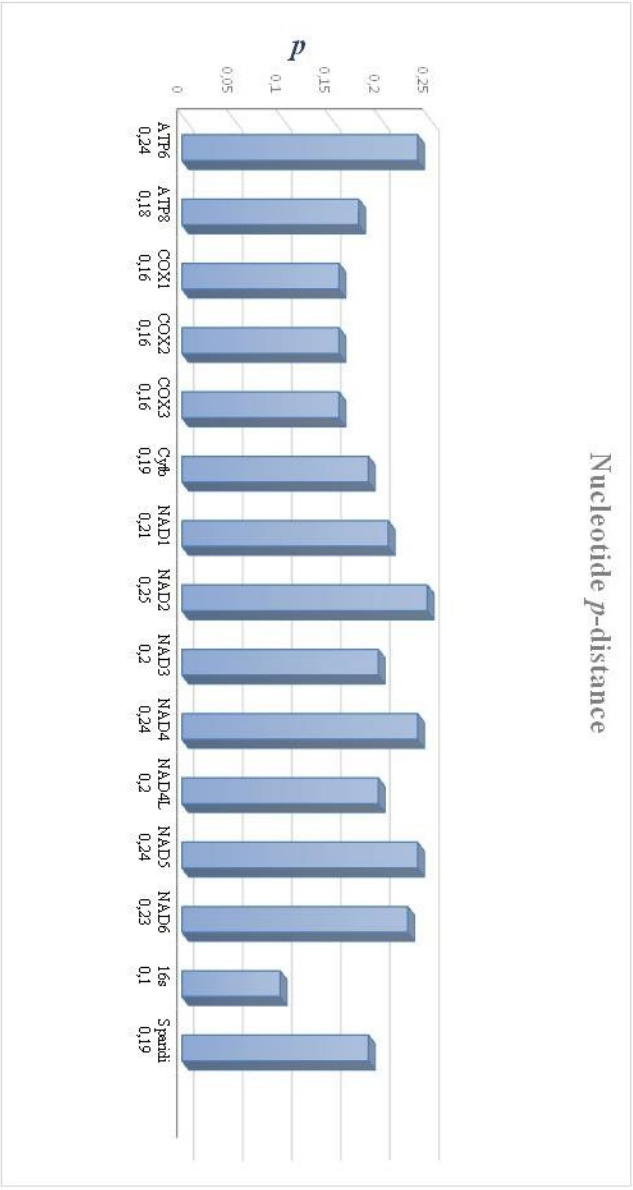
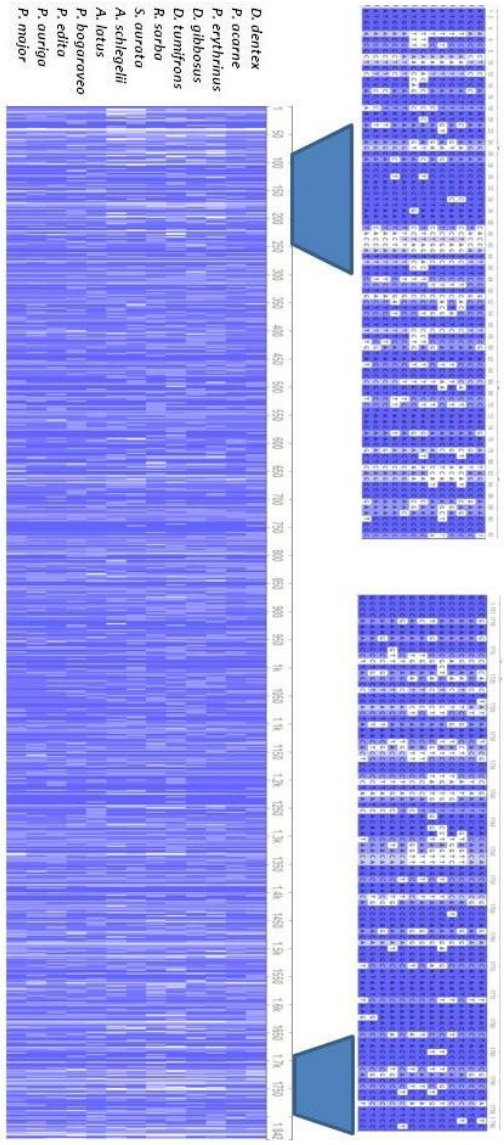


Figure 9.4. Typical composition of the NAD5 gene, with alternation of conserved (blue) and variable (white) regions. In the "Percentage Identity" coloring scheme, residues are colored according to the percentage of each residue in the alignment column: white characters 40% or less, light blue 41-60%, blue 61-80%, dark blue 81% or more. Variable and conserved regions are distributed in a uniform way throughout the gene.



9.4.2.3. NAD5 amplification and analysis

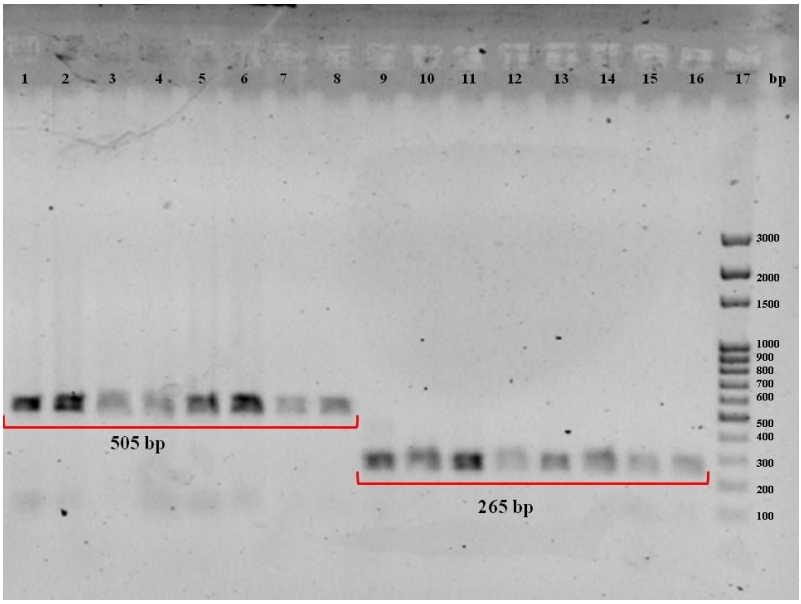
Following multiple alignment of the *Sparidae* complete *mt*DNA, primers for amplifying species-specific nucleotide sequences were designed on two fragments of the *NAD5* gene. The primer set is reported in Table 9.14. Two reverse primers correspond to fragments of different sizes: revnad5.1, with an expected PCR product of 505 bp, and revnad5.2, with an expected PCR product of 265 bp.

After testing *in silico* specificity of the designed primers for all *Sparidae* species (Table 9.1), PCR results showed that both primer pairs allow empirical amplification in all species, as confirmed by BLAST analysis (Figure 9.5). Sequence comparison with databases allowed correct species identification for all tested species, with sequence similarity scores between 98% and 100%. Species identification was made possible by using both 505 bp and 265 bp fragments.

Table 9.14. NAD5 PCR primers.

Primer name	Sequence	Tm °C	CG%	nt	A	T	C	G
Fwnad5	ACACCGGTCTCTGCCCTACT	62.02	54.02	20	3.08	5.03	8.03	2.05
Revnad5.1	AGGGCTCAGGCGTTTAGGT	69.05	64.00	19	3.00	3.08	3.08	8.03
Revnad5.2	GAAGGCAAGTTGTGGTTG	63.05	53.07	18	4.03	4.01	2.01	7.06

Figure 9.5. PCR amplification of NAD5 fragments. Both primer pairs allowed correct amplification of the eight Sparidae species currently tested. Lane 1-8 and 9-16: D. dentex, D. gibbosus, P. acarne, P. bogaraveo, P. erythrinus, P. auriga, P. major, and S. aurata. Lane 17: control ladder 1 Kb.



9.5. Conclusions

Our results demonstrated the importance of analyzing the whole *mt*DNA genome of fish species for unequivocal and effective identification. This study focused on *Sparidae* family that comprises highly commercially valued species, in order to guarantee their correct species authentication. In our work, the alignment of thirteen complete *Sparidae* mitogenomes allowed to identify two genes, *NAD5* and *NAD2*, whose nucleotide sequence showed more interspecific phylogenetic divergence than standard markers.

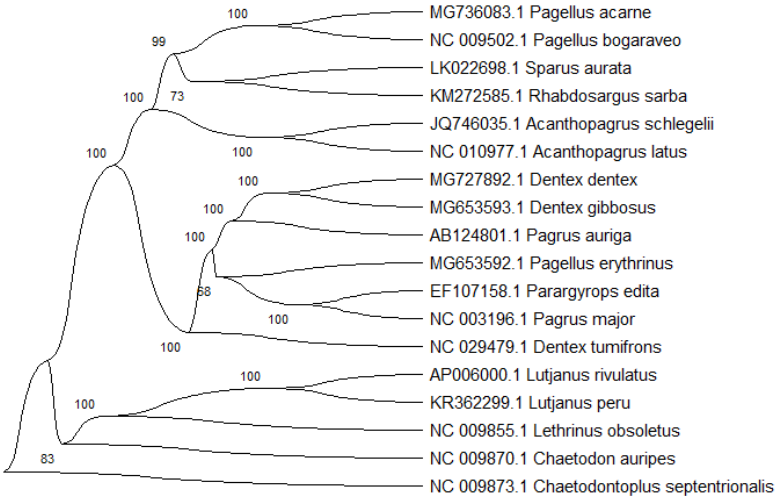
The selected regions of the *NAD5* gene have adequately high interspecific nucleotide dissimilarity (Table 9.12**,** Figures 9.3 and 9.4) to provide unambiguous results for *Sparidae* species authentication, also among particularly similar species from a genetic point of view, such as *D. dentex* and *D. gibbosus*.

Another gene that, in our opinion, deserves attention in sparid species identification field is *NAD2*. This idea is supported by results of Hamming Distance, nucleotide sequence variability and *p*-genetic distance. However, *NAD2* gene has only one CR (29 bp long) for primers design. Further studies are underway to verify the efficacy of the *NAD2* gene for this purpose.

This study contributes to the molecular traceability of fishery products, in agreement with Regulation (EU) 1379/2013 (European Commission, 2013), on the common organization of the markets in fishery and aquaculture products. Mislabelling of fish such as less valuable species that are sold under the names of more expensive ones is a growing problem in the production and distribution chain. For consumer protection and safety, competent national authorities responsible for monitoring and enforcing could improve and make full use of available technology, including *DNA*-testing, in order to deter operators from false labelling of seafood.

Supplementary material

Figure 9.6. Evolutionary relationships of Sparidae species



The evolutionary history was inferred using the UPGMA method. The optimal tree with the sum of branch length = 1,36665277 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved 18 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair. There was a total of 23471 positions in the final dataset. Evolutionary analyses were conducted in MEGA X.

References

Alikhan, N.F., Petty, N.K., Ben Zakour, N.L. & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12: 402. https://doi.org/10.1186/1471-2164-12-402.

Antonucci, F., Costa, C., Aguzzi, J., & Cataudella, S. (2009). Ecomorphology of Morpho Functional relationships in the family of Sparidae: a quantitative statistic approach. *Journal of Morphology*, 270, 843e855. https://doi.org/10.1002/jmor.10725

Armani, A., Guardone L., Castigliego, L., D'Amico, P., Messina, A., Malandra, R., Gianfaldoni D., & Guidi A. (2015). DNA and Mini-DNA barcoding for the identification of Porgies species (family *Sparidae*) of commercial interest on the international market. *Food Control*, 50, 589-596. https://doi.org/10.1016/j.foodcont.2014.09.025

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, 27, 578-9. https://doi.org/10.1093/bioinformatics/btq683

Bolger, A. M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30, 2114-20.

Cawthorn, D. M., Steinman, H. A., Witthuhn, R. C. (2012). Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene*, 491, 40-8. https://doi.org/10.1016/j.gene.2011.09.009

Cawthorn, D. M., Duncan, J., Kastern, C., Francis, J. & Hoffman, L. C. (2015). Fish species substitution and misnaming in South Africa: An economic, safety and sustainability conundrum revisited. *Food Chem*, 185, 165-81. https://doi.org/10.1016/j.foodchem.2015.03.113

Ceruso, M., Mascolo, C., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2018a). The complete mitochondrial genome of the common dentex, *Dentex* (Perciformes: Sparidae). *Mitochondrial DNA Part B*, 3, 391-392. https://doi.org/10.1080/23802359.2018.1450675

Ceruso, M., Mascolo, C., Lowe, E. K., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2018b). The complete mitochondrial genome of the

common Pandora *Pagellus erythrinus* (Perciformes: Sparidae), *Mitochondrial DNA Part B*, 3:2, 624-625. https://doi.org/10.1080/23802359.2018.1467235

Chevreux, B., Wetter, T. & Suhai, S. (1999). Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, 99, 45-56.

Chiba, S. N., Iwatsuki, Y., Yoshino, T., Hanzawa, N. (2009). Comprehensive phylogeny of the family Sparidae (Perciformes: Teleostei) inferred from mitochondrial gene analyses. *Genes & Genetic Systems*, 84, 153-70. https://doi.org/10.1266/ggs.84.153

Day, J. (2002). Phylogenetic relationships of the Sparidae (Teleostei: Percoidei) and implications for convergent trophic evolution. *Biological Journal of the Linnean Society*, 76, 269-301. https://doi.org/10.1111/j.1095-8312.2002.tb02088.x

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., Taberlet, P. (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* 10(9). pii: 20140562. https://doi.org/10.1098/rsbl.2014.0562

Dray, L. M., Neuhof, M., Diamant, A., & Huchon, D. (2016). The complete mitochondrial genome of the gilthead seabream *Sparus aurata* L. (Sparidae). *Mitochondrial DNA*, 27, 781-2. https://doi.org/10.3109/19401736.2014.928861

Felsenstein J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39, 783-791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Giusti, A., Tinacci, L., Sotelo, C. G., Acutis, P. L., Ielasi, N., Armani, A. (2019). Authentication of ready-to-eat anchovy products sold on the Italian market by BLAST analysis of a highly informative cytochrome b gene fragment. *Food control*, 97, 50-57. https://doi.org/10.1016/j.foodcont.2018.10.018

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT *Nucleic Acids Symposium Series*, Vol. 41, pp. 95-98 Key: citeulike:691774

Hebert, P. D., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003) Biological identifications through DNA barcodes. *Proceedings.*

Biological sciences Royal Society, 270, 313–321. https://doi.org/10.1098/rspb.2002.2218

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., & Nishida, M. (2013). MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Molecular Biology and Evolution*, 30, 2531-2540. https://doi.org/10.1093/molbev/mst141

Katavic, I., Grubisic L. & Skakelja N. (2000). Growth performance of pink dentex as compared to four other sparids reared in marine cages in Croatia. Short Communication. *Aquaculture International*, 8, 455–461.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35:1547-1549.

Kumar, A., Chordia, N. (2015). In Silico PCR Primer Designing and Validation. In: Basu C. (eds) PCR Primer Design. *Methods in Molecular Biology*, vol 1275. Humana Press, New York, NY, 143-151. https://doi.org/10.1007/978-1-4939-2365-6_10

Larkin, M.A., Blackshields, G., Brown ,N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948.

Li, J., Yang, H., Xie, Z., Yang, X., Xiao, L., Wang, X., Li, S., Chen, M., Zhao, H., & Zhang, Y. (2016). The complete mitochondrial genome of the *Rhabdosargus sarba* (Perciformes: Sparidae). *Mitochondrial DNA A DNA MappSeq Anal,* 27, 1606-7. https://doi.org/10.3109/19401736.2014.958690

Li, X., Shen, X., Chen, X., Xiang, D., Murphy, R. W., Shen, Y. (2018). Detection of Potential Problematic Cytb Gene Sequences of Fishes in GenBank. *Frontiers in Genetics*, 9, 30. https://doi.org/10.3389/fgene.2018.00030

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Sordino, P., Pepe, T. (2018 a). The complete mitochondrial genome of the axillary seabream, *Pagellus acarne* (Perciformes: Sparidae). *Mitochondrial*

DNA Part B, 3, 434-435. https://doi.org/10.1080/23802359.2018.1450674

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Sordino, P., Pepe, T. (2018 b). The complete mitochondrial genome of the Pink dentex *gibbosus* (Perciformes: Sparidae), *Mitochondrial DNA Part B*, 3:2,525-526. https://doi.org/10.1080/23802359.2018.1467230

Mascolo, C., Ceruso, M., Palma, G., Anastasio, A., Pepe, T., Sordino, P. (2019). Comparison of mitochondrial DNA enrichment and sequencing methods from fish tissue. *Food Chemistry*, accepted for publication.

Mitchell, M. An introduction to genetic algorithms. *MIT press*, 1998.

Miya, M., Kawaguchi, A., Nishida, M. (2001). Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, 18, 1993-2009. https://doi.org/10.1093/oxfordjournals.molbev.a003741

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M. & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2:150088. https://doi.org/10.1098/rsos.150088

Okonechnikov K, Golosova O, Fursov M, (2012). The UGENE team. 28: 1166-1167. doi:10.1093/bioinformatics/bts091

Orrell, T.M., Carpenter, K.E., Musick, J.A., & Graves, J.E. (2002). Phylogenetic and biogeographic analysis of the Sparidae (Perciformes: Percoidei) from Cytochrome *b* sequences. *Copeia*, 3, 618-631. https://doi.org/10.1643/0045-8511(2002)002[0618:PABAOT]2.0.CO;2

Pepe, T., Trotta, M., Di Marco, I., Anastasio, A., Bautista, J. M., & Cortesi, M. L. (2007). Fish species identification in surimi-based products. *Journal of Agricultural and Food Chemistry,* 55, 3681-5. https://pubs.acs.org/doi/abs/10.1021/jf063321o

Ponce, M., Infante, C., Jimenez-Cantizano, R.M., Perez, L., & Manchado, M. (2008). Complete mitochondrial genome of the blackspot seabream, *Pagellus bogaraveo* (Perciformes: Sparidae),

with high levels of length heteroplasmy in the WANCY region *Gene*, 409, 44-52. https://doi.org/10.1016/j.gene.2007.11.004

Ponce, M., Infante, C., Catanese, G., Cardenas, S., Manchado, M. Complete mitochondrial DNA of redbanded seabream (*Pagrus auriga*). Unpublished.

Ratnasingham, S., & Hebert, P. D. (2013). A DNA-based registry for all animal species:

the Barcode Index Number (BIN) System. *PLoSOne*, 8(7), e66213. https://doi.org/10.1371/journal.pone.0066213

Rubinoff, D., Cameron, S. & Will, K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. *Journal of Heredity*, 97:581-94. https://doi.org/10.1093/jhered/esl036

Saitou, N. & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4, 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Satoh, T. P., Miya, M., Mabuchi, K., & Nishida, M. (2016). Structure and variation of the mitochondrial genome of fishes. *BMC Genomics*, 17, 719. https://doi.org/10.1186/s12864-016-3054-y

Shi, X., Su, Y., Wang, J., Ding, S., & Mao, Y. (2012). Complete mitochondrial genome of black porgy *Acanthopagrus schlegelii* (Perciformes, Sparidae). *Mitochondrial DNA*. 23, 310-2. https://doi.org/10.3109/19401736.2012.683182

Shokralla, S., Hellberg, R. S., Handy, S. M., King, I., Hajibabaei, M. (2015). A DNA Mini-Barcoding System for Authentication of Processed Fish Products. *Sci Rep*, 5:15894.

Tamura K., Nei M., and Kumar S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* (USA) 101:11030-11035. https://doi.org/10.1073/pnas.0404206101

Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. (2013). Mega 6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution,* 30, 2725-2729.

Terio, V., Di Pinto, P., Decaro, N., Parisi, A., Desario, C., Martella, V., Buonavoglia, C., Tantillo, M.G. (2010). Identification of tuna species in commercial cans by minor groove binder probe real-time

polymerase chain reaction analysis of mitochondrial DNA sequences. *Molecular and Cellular Probes*, 24, 352-6. https://doi.org/10.1016/j.mcp.2010.07.006

Thompson J. D., Higgins D. G., Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22, 4673-4680.

Tinacci, L., Giusti, A., Guardone, L., Luisi, E., Armani, A. (2019). The new Italian official list of seafood trade names (annex I of ministerial decree n. 19105 of September the 22nd, 2017): Strengths and weaknesses in the framework of the current complex seafood scenario. *Food Control*, 96, 68-75. https://doi.org/10.1016/j.foodcont.2018.09.002

Trotta, M., Schönhuth, S., Pepe, T., Cortesi, M. L., Puyet, A., Bautista, J. M. (2005). Multiplex PCR method for use in real-time PCR for identification of fish fillets from grouper (Epinephelus and Mycteroperca species) and common substitute species. *Journal of Agricultural and Food Chemistry*, 53, 2039-45. https://pubs.acs.org/doi/10.1021/jf048542d

Wang, C., Chen, Q., Lu, G., Xu, J., Yang, Q., Li, S. (2008). Complete mitochondrial genome of the grass carp (Ctenopharyngodon idella, Teleostei): Insight into its phylogenic position within Cyprinidae. *Gene*. 424:96–101. https://doi.org/10.1016/j.gene.2008.07.011

Ward, R. D., Hanner, R., Hebert, P. D. (2009). The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, 74:329-56. https://doi.org/10.1111/j.1095-8649.2008.02080.x

Xia, J., Xia, K., Gong, J., & Jiang, S. (2007). Complete mitochondrial DNA sequence, gene organization and genetic variation of control regions in *Parargyrops edita*. *Fisheries Science*, 73, 1042-1049. https://doi.org/10.1111/j.1444-2906.2007.01434.x

Xia, J., Xia, K., Jiang, S. (2008). Complete mitochondrial DNA sequence of the yellowfin seabream *Acanthopagrus latus* and a genomic comparison among closely related sparid species. *DNA Sequence*, 19, 385-393. https://doi.org/10.1080/19401730802350998

Zeng, L., Jiang, L., Wu, C. Complete mitochondrial genome of yellowback seabream (*Evynnis tumifrons*). Unpublished.

Zerbino, D. R., & Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs, *Genome Research*, 18, 821 – 829. http://www.genome.org/cgi/doi/10.1101/gr.074492.107

https://www.fishbase.de/Summary/FamilySummary.php?ID=330, Accessed 14.12.18.

Conclusions

Currently, the supply chain of seafood is one of the most involved sectors of the food industry. Therefore, fish products are subjected to several production phases. This is the main cause that fuels the existence of fraud in this sector.

Despite their similarity, sparid species have different organoleptic quality that corresponds to different commercial value. In fact, they are marketed all over the world and they are among the most expensive and appreciated seafood. Consequently, substitution of commercially important Sparidae species is common. Falsification, also acknowledged as aliud pro alio, is one of the most common types of fraud and it generally comprises the substitution of expensive species with less valuable ones. The more expensive sparid species, like *Dentex*, are often replaced with less expensive species or lower quality alternatives (e.g. *Dentex gibbosus, Pagrus auratus*). Frouds are favoured by the extremely wide number of new sparid species coming from the Atlantic and Indo-Pacific oceans. Moreover, species identification becomes even harder to achieve after industry processing, when distinctive external traits are removed. Based on bibliographic fonts, frauds by species substitution involve each year an alarming part of the traded products, implying an actual damage for the sector. The European Union made more stringent Regulations aimed at preventing fraud occurrences and protecting consumers, particularly concerning the mandatory information on products' label. An official and valid tool to support European legislation is required. *DNA* has been shown to be the most diagnostic and helpful molecular tool for this purpose. Some sequence regions clearly characterize a unique species and allow to discriminate it from others. However, mitochondrial markers cannot be randomly used for species identification in different fishes species *DNA* markers may perform well for certain species but may be less discriminating for others, in particular when fish species are closely related. Sparids are very similar each other from a genetic and morphological point of view. This led to the proposal to study and analyze the complete *mt*DNA sequence with the aim to identify mitochondrial markers or multiple marker approaches with the high and more specific discrimination capacity for each fish species. Species of the genus *Dentex*,

*Pagellu*s and *Pagrus* were studied to address sparid phylogenetic interrelationships.

In accordance with the stated aims (as reported in the respective section), the results of my Ph.D. project during the 3 years of study, were:

In the **Chapter 1**, the aim was to isolate the fish whole mitochondrial genome. The most common approach, which is based on a combination of long and short PCR reactions followed by Sanger sequencing method, was applied. The *mt*DNA was correctly amplified in in five species (*Dentex, Dentex gibbosus, Dentex nufar, Pagellus acarne* and *Pagellus erythrinus*) of commercial interest belonging to the *Sparidae* family. The best outcomes were obtained for two species analized (*D. gibbosus* and *P. erythrinus*). The amplification was also possible for the other three species (*D. dentex, D. nufar* and *P. acarne*) but with a low degree of PCR and sequencing reproducibility.

The **Chapter 2** focused on the finding of a most effective technique in terms of time, costs and efficiency to obtain the complete mitogenome sequence from fish tissues. With this aim, we compared three protocols in two species *(D. gibbosus* and *P. erythrinus*): 1) Illumina MiSeq sequencing of long amplicon (7.3-9.3 Kb) (Chapter 1); 2) 'traditional' Sanger sequencing of short amplicons (1.4-3.2 Kb), and 3) shotgun Illumina MiSeq approach to miniprep kit-enriched *mt*DNA.

The results showed that the best method to extract and sequence the complete *mt*DNA from fish tissue is the third method (shotgun sequencing, based on *mt*DNA enrichment with miniprep columns followed by MiSeq NGS). This procedure maximizes efficiency in separation and sequencing of the *mt*DNA also allowing to balance time, costs and sequence quality. The selected method allowed obtaining the complete mitogenome sequences from *Dentex gibbosus* (MG653593) and *Pagellus erythrinus* (MG653592).

The protocol chosen in the Chapter 2, was subsequently used to obtain the other sparids mitogenomes as reported in the following chapters:

In the **Chapter 3**, we report the complete mitochondrial genome of *Dentex gibbosus*. The mitogenome is 16.771 bp in length and contained 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and 2 non-coding regions. The overall base composition of *D. gibbosus mt*DNA is: 27,8 % for A, 28,60 % for C, 16,5 % for G, 27,05 % for T.

In the **Chapter 4,** we describe *Pagellus erythrinus* complete mitochondrial genome. The sequence is comprised of 16,828 bp and consists of 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and two non-coding regions (D-loop and L-origin). The overall nucleotide composition is: 27.5% of A, 28.2% of C, 27.5% of T and 16.8% of G.

In the **Chapter 5**, we sequenced *Dentex* complete mitochondrial genome. The sequence is comprised of 16,907 bp and consists of 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes and two non-coding regions (D-loop and L-origin). The overall nucleotide composition is: 27.5% of A, 28.7% of C, 26.9% of T and 16.9% of G.

In the **Chapter 6**, we obtained *Pagellus acarne* complete mitochondrial genome is 16,486 bp in length, consisting of 13 protein-coding genes, 22 tRNA genes, two rRNA genes, and two non-coding regions region (D-loop, 808 bp and L-origin, 29 bp). Its overall base composition is A: 26,8%, C: 29%, G: 17.6%, and T: 26.6 %.

In the **Chapter 7**, the complete mitochondrial genome of the *Dentex angolensis* was sequenced. It is 16.581 bp in length and contains 13 protein-coding genes, 22 tRNA genes and 2 rRNA genes. The overall base composition of *D. angolesis mt*DNA is: 27,4 % for A, 28 % for C, 16,7 % for G, 27,9 % for T. Phylogenetic analysis indicated that *D. angolensis* is most closely related to *D. tumifrons.*

In the **Chapter 8,** preliminary results showed that the complete mitochondrial genome of the *Pagrus* was obtained. It is 16.167 bp in length and contains 13 protein-coding genes, 22 tRNA genes and 2 rRNA genes.

In the **Chapter 9**, we report the first study of the whole *mt*DNA of some species of the *Sparidae* family. The aim was to select more efficient barcoding markers for taxonomical discrimination against frauds. We

analyzed and compared thirteen *Sparidae* mitogenomes (*Acanthopagrus latus, Acanthopagrus schlegelii, Dentex, Dentex gibbosus, Dentex tumifrons, Pagellus acarne, Pagellus bogaraveo, Pagellus erythrinus, Pagrus auriga, Pagrus major, Parargyrops edita, Rhabdosargus sarba, Sparus aurata*) nine publicly available and four recently sequenced ones by our research group. In particular, we studied for effective *DNA* barcode markers regions for the correct identification of sparid species by observing for interspecific variable regions flanked by conserved sequences useful for PCR primer design. The alignment of the thirteen Sparidae complete mitogenomes demonstrated a similarity of the translated sequences higher than 70% compared to the reference genome (*Sparus aurata*, LK022698.1, mitogenome). This suggests, that the thirteen species are very similar each other.

The gene-by-gene sequence analysis showed that the *NAD* genes group have encouraging utility in discriminating closely related sparid species. In fact, their nucleotide sequence variability, compared with classical barcodes markers, was higher. In particular, *NAD5* gene appeared with the high interspecific nucleotide sequence variability. The study of this gene allowed to detect conservative regions useful to design of specie-specific primers. The amplification and sequencing of the selected *NAD5* fragments allowed species identification. Therefore, when blasted with databases, the obtained sequences allowed correct species identification for all tested species, with sequence similarity scores between 98 % and 100%. In conclusion, the selected regions of the *NAD5* gene have high interspecific nucleotide divergence to give unambiguous results for *Sparidae* species authentication. The species identification was possible also in case of particularly similar species from a genetic point of view, such as *D. dentex* and *D. gibbosus* (6% of genetic divergence). Another gene that deserves attention in sparid species identification field is *NAD2*. This idea is supported by results of Hamming Distance, nucleotide sequence variability and p-genetic distance. Further studies are underway to verify the efficacy of the *NAD2* gene for this purpose.

This thesis demonstrated that **1)** the study of whole *mt*DNA of the fish species of the family *Sparidae* is a fruitful method for identifying more effective barcoding markers for taxonomical and phylogenetic

discrimination against frauds; **2)** the use of *mt*DNA enrichment with miniprep columns followed by Next Generation Sequencing (NGS) overcomes problems deriving from PCR and shotgun sequencing; **3)** the compared thirteen sparids species are very similar each other from a genetic point of view; **4)** *D. dentex* and *D. gibbosus* are particularly similar, with 6% of genetic dissimilarity.

In conclusion, for the first time, the analysis of the complete *mt*DNA sequences in fishes allowed to identify a new molecular barcode marker. The in-depth analysis of *Sparidae* mitogenomes allowed to resolve a taxonomic dispute persisting for decades.

The development of new methods described above appear in agreement with the provisions of Regulation (EC) No. 178/2002 (European Commission, 2002) and Regulation (EU) 1379/2013 (European Commission, 2013). The European law aim to ensure a high level of protection of human life and health, establishing a comprehensive system of traceability with a "farm to fork" approach. Therefore, traceability has to be assured at all production steps. Scientific and technical methods have always to support Community legislation. This research provides the basis for the molecular traceability of fish products.

Currently, there is a need for harmonization and standardization of analytical techniques and for universal access to a standard database. Nowdays, there is not an official and univocal method used internationally for the genetic identification of species. In Italy, official laboratories (Ministry of health, Zooprofilattici Institutes, eg.) do not use all the same protocol of analysis for species identification. A better cooperation between food control authorities and law enforcement agencies is required to prevent and struggle food fraud activities. The introduction of new and performing analytical method for fish species identification means that food inspectors and consumers can know that how declared on the label exactly corresponds to the real purchased food product.

Prospects for future research

The sequence's analysis of thirteen *Sparidae* complete mitogenomes (*Acanthopagrus latus, Acanthopagrus schlegelii, Dentex, Dentex gibbosus, Dentex tumifrons, Pagellus acarne, Pagellus bogaraveo, Pagellus erythrinus, Pagrus auriga, Pagrus major, Parargyrops edita, Rhabdosargus sarba, Sparus aurata*) displayed that the similarity of the sequences is higher than 70% compared to the reference genome (*Sparus aurata*, LK022698.1, mitogenome).

Moreover, the mitochondrial phylogenetic tree of Sparidae species shows a close relationship between *D. dentex* and *D. gibbosus*, in agreement with previous works present in bibliography (Orrel, 2002; Orrel & Carpenter, 2004; Chiba, 2009). The Hamming Distance comparison displayed that *D. dentex* and *D. gibbosus* are the closest species, with 6% of genetic dissimilarity in alignement of whole *mt*DNA and with 12% of genetic dissimilarity in *NAD2* gene. Molecular markers *COI* and *Cytb,* currently used in *Sparidae* for studies of fish species identification and philogenetic analysis, showed an even lower genetic dissimilarity. Further studies are proceeding to verify the usefulness of the *NAD2* as specie-specific gene, throught a simple and rapid amplification, without sequencing. In particular, food industries could use this method for the identification of the most important species, like *D. dentex, D. gibbosus, P. erythrinus* (the preparation of the manuscript related to this topic is in progress).

Considering the commercial importance and the entity of frauds of *Dentex* in agreement with Regulation (EU) 1379/2013 (European Commission, 2013), on the common organization of the markets in fishery and aquaculture products, this study will contribute to the molecular traceability of fishery products and to the molecular characterization of the catch based on FAO areas.

Additional studies could contribute to the development of a bio-geographical markers to better characterize the Mediterranean catch from that Atlantic using the FAO Major Fishing Areas.

Several studies of other authors, realized with the use of molecular markers, have demonstrated how the taxonomic relationships of the *Sparidae* family necessity a reconsideration. In fact, they have highlighted that the *Sparidae* family is not monophyletic, since it includes the genus Spicara, attributed to the family *Centracanthidae*. In this regard, in future researches it would be appropriate to study the mitogenome sequence of the species belonging both to the *Lutjanidae, Lethrinidae* that are morphologically very similar to the sparids (as already mentioned in Section 1.4.4.3), and *Centracanthidae* families. Supplementary studies are proceeding to sequence other specimens of different species belonging to the *Sparidae* family in order to get an overview of intra- and interspecific variability as complete as possible, to verify the correct taxonomic position of different species belonging to the *Sparidae* family.