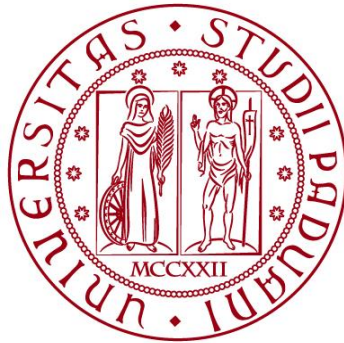


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea magistrale in Biologia Evoluzionistica**



**TESI DI LAUREA**

**Ottimizzazione ed utilizzo del DNA ambientale  
per il rilevamento della fauna ittica  
in laguna di Venezia**

**Relatore:** Prof. Lorenzo Zane  
Dipartimento di Biologia

**Correlatore:** Dott.ssa Elisa Boscari  
Dipartimento di Biologia

**Laureanda:** Irene Gregori

**ANNO ACCADEMICO 2021/2022**



# INDICE

1. <u>Introduzione</u>	1
1.1. La laguna di Venezia: un ambiente di transizione	1
1.2. Il DNA ambientale	4
1.3. Fasi fondamentali dell'analisi di DNA ambientale	8
1.3.1. Campionamento	9
1.3.2. Estrazione di DNA genomico da filtro	10
1.3.3. Amplificazione del DNA con primer gruppo specifici e DNA metabarcoding	11
1.3.4. Blocking primer	14
1.3.5. Analisi bioinformatiche	15
2. <u>Scopo della tesi</u>	16
3. <u>Materiali e metodi</u>	17
3.1. Campionamento	17
3.2. Estrazione di DNA genomico da filtro	20
3.3. Amplificazione di 12S rDNA (Tele02)	21
3.3.1. Blocking primer	24
3.4. Preparazione delle librerie e sequenziamento	26
3.5. Analisi informatiche	26
3.6. Analisi statistiche ed ecologiche	28
4. <u>Risultati</u>	30
4.1. Estrazione del DNA da filtro e prove di amplificazione	30
4.2. Sequenziamento	32
4.3. Identificazione di OTU	32
4.4. Analisi statistiche	40
4.4.1. Analisi delle componenti principali	41
4.5. Blocking primer	42

5. <u>Discussione</u>	46
6. <u>Conclusioni</u>	49
7. <u>Bibliografia</u>	50
8. <u>Sitografia</u>	58
9. <u>Appendici</u>	59

# 1. Introduzione

In questo lavoro di tesi è stato ottimizzato un approccio basato sull'analisi del DNA ambientale proveniente dalla laguna di Venezia, al fine di rilevare le specie appartenenti alla fauna ittica, analizzarne la distribuzione all'interno della laguna e la variazione nel tempo.

## 1.1. La laguna di Venezia: un ambiente di transizione

Con il termine acque di transizione si indica comunemente le acque salmastre originate dal mescolamento tra le acque costiere e le acque dolci dei fiumi, quali lagune, stagni costieri, laghi salmastri e zone di delta ed estuario. In accordo con questo, il D.Lgs. 152/06 definisce le acque di transizione come “corpi idrici superficiali in prossimità della foce di un fiume, che sono parzialmente di natura salina a causa della loro vicinanza alle acque costiere, ma sostanzialmente influenzati dai flussi di acqua dolce”. Tuttavia, tali definizioni non sono del tutto appropriate per alcuni ambienti, nei quali l'apporto di acque dolce è limitato o assente; una diversa definizione operativa di acque di transizione è stata proposta da McLusky e Elliot (2007), che parla di aree acquatiche che non sono né totalmente aperte alla costa, né aree d'acqua dolce chiuse, o correnti.

Indipendentemente dalla definizione, le acque di transizione rappresentano dei sistemi naturali d'interfaccia di elevato valore ecologico, strettamente connessi agli ambienti adiacenti (mare da una parte, fiumi dall'altra) tramite flussi di materia (Signa, 2011). Per le loro caratteristiche e localizzazione, sono importanti aree di riproduzione e *nursery* per numerose specie acquatiche, svolgendo così un ruolo decisivo per il mantenimento della biodiversità e della produttività degli ecosistemi marini e fluviali adiacenti. Grazie alla loro struttura fisica e biologica, gli ecosistemi di transizione fungono da filtri naturali tra terra e mare, attraverso cui sono metabolizzati materia organica e nutrienti di origine continentale (Tagliapietra *et al.* 2014).

La laguna di Venezia è uno dei più vasti sistemi di transizione del Mediterraneo; copre una superficie di 550km<sup>2</sup> (isole incluse) e un volume d'acqua di 600x10<sup>6</sup> m<sup>3</sup>. La laguna è in collegamento con il mare attraverso le bocche di porto di Lido (che misura 800m di larghezza) Malamocco (400m) e Chioggia (380m) ed è compresa tra le odierne foci del Piave a nord e quella del Brenta-Bacchiglione a sud. Gli ecosistemi acquatici di transizione e i tratti di mare ad essi contigui sono da sempre di grande interesse per lo sfruttamento delle risorse ittiche (Blaber *et al.* 2000; Perez-Ruzafa *et al.* 2012); questo vale anche la laguna di Venezia, che risulta molto influenzata dall'attività di pesca che si svolge al suo interno, oltre che da altri

impatti di origine antropica come scarico di acque reflue, trasporto acqueo, inquinamento da attività industriali.



**Fig. 1** - La laguna di Venezia vista dall'alto. Fonte: <https://www.italianostravenezia.org/2020/10/18/autorita-per-la-laguna-di-veneziale-nostre-considerazioni/>

La laguna di Venezia (Fig. 1) è un ambiente storicamente fortemente modificato dall'uomo, basti pensare alla diversione dei fiumi Piave, Sile e Brenta, iniziate con gli interventi trecenteschi della Repubblica Serenissima e concluse nel 1860: delle imponenti opere idrauliche che avevano come obiettivo quello di indirizzare la foce dei fiumi direttamente in mare, senza apportare acqua in laguna.

Queste opere sono state svolte con l'obiettivo di mantenere integro il bacino lagunare ed evitare che le foci dei fiumi minacciassero di interrare la laguna, rendendo impossibile la difesa contro la terraferma (Pavan, 1989) e, sempre al fine di renderla più difendibile, isolare Venezia consentendole maggiore protezione (Rinaldo, 1997). Queste costruzioni hanno portato, conseguentemente, al fatto che la laguna ha assunto caratteristiche più marine di quelle originarie. Nonostante ciò, la laguna non è tuttora da considerarsi un ambiente esclusivamente di acqua salata, in quanto gli apporti di acqua dolce non sono trascurabili (complessivamente, circa 40m<sup>3</sup>/s) (Quartarella, 2013).

In tempi più recenti, un ulteriore impatto a livello ambientale per la laguna di Venezia è legato alle opere volte a risolvere le problematiche legate alle "acque alte", alte maree particolarmente elevate caratterizzate da un'intensità tale da allagare larga parte dell'area urbana, che rappresentano

una grande minaccia per la città di Venezia, i suoi abitanti e l'inestimabile patrimonio storico, artistico e ambientale. Il progetto MOSE alle bocche della laguna di Venezia è senza dubbio la più grande opera di ingegneria marittima realizzata in tempi moderni in una laguna ([www.corila.it/](http://www.corila.it/)). Il MOSE consiste in 4 barriere costituite da 78 paratoie mobili tra loro indipendenti in grado di separare temporaneamente la laguna dal mare e di difendere Venezia sia dagli eventi di marea eccezionali e distruttivi, sia da quelli periodici più frequenti e comunque a lungo termine fortemente impattanti. Le barriere sono collocate in corrispondenza delle bocche di porto di Lido, Malamocco e Chioggia. Il MOSE può proteggere Venezia e la laguna da maree alte fino a 3 metri e da un innalzamento del livello del mare fino a 60 centimetri ([www.mosevenezia.eu/mose/](http://www.mosevenezia.eu/mose/)).

“Venezia2021, Programma di ricerca scientifica per una laguna “regolata”” è un programma di ricerca scientifica coordinato da CORILA, che ha come obiettivo quello di acquisire le conoscenze necessarie sull'ecosistema della laguna di Venezia, al fine di svolgere una gestione sostenibile della stessa. In particolare, si prefigge di regolare l'impatto del MOSE sulla comunità ecosistemica. Il CORILA, consorzio per il coordinamento delle ricerche inerenti al sistema lagunare di Venezia, è un'associazione tra Università Ca' Foscari di Venezia, Università IUAV di Venezia, Università di Padova, Consiglio Nazionale delle Ricerche e Istituto Nazionale di Oceanografia e Geofisica Sperimentale, vigilata dal Ministero dell'Istruzione Università e Ricerca.

Al fine di contribuire al mantenimento del corretto equilibrio dell'ecosistema lagunare, Venezia2021 integra al suo interno strumenti osservativi ed elaborazione dei dati raccolti per proteggere il naturale funzionamento ecologico della laguna, senza escludere dalle considerazioni l'importanza dei servizi ecosistemici forniti dalla stessa. Il programma di ricerca comprende un totale di 15 linee di ricerca, raggruppate in 5 tematiche omogenee ([www.corila.it/it/Venezia2021](http://www.corila.it/it/Venezia2021)).

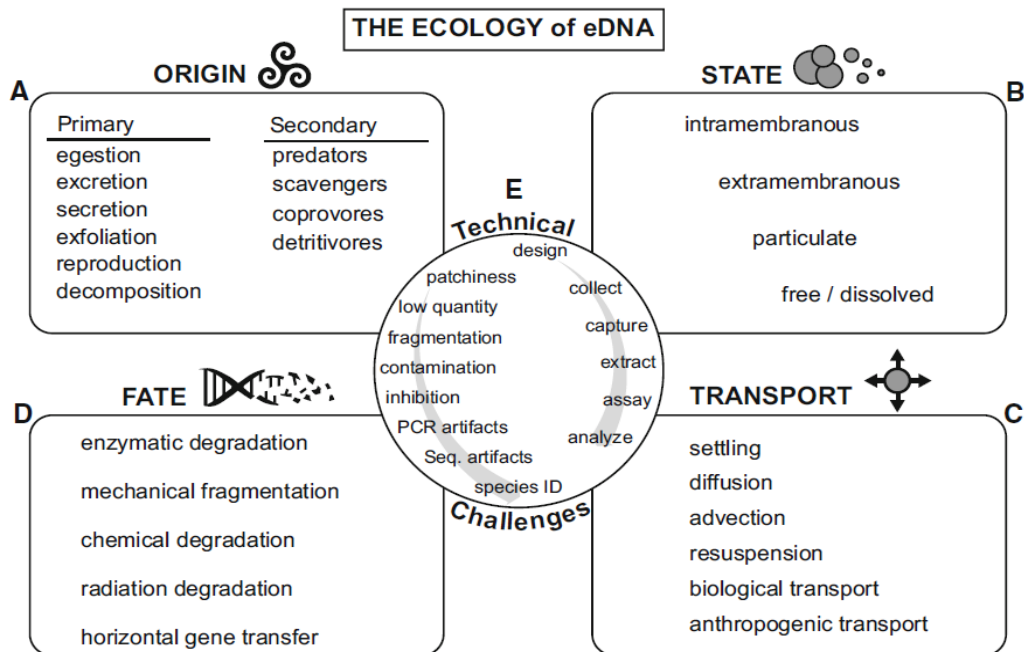
Uno degli *workpackage* del programma di ricerca Venezia2021, all'interno del quale si colloca anche il presente progetto di tesi, si occupa della messa a punto di approcci sperimentali volti alla valutazione dell'impatto del MOSE sulle comunità biologiche, mediante monitoraggio sistematico e a lungo termine della composizione in specie e della loro abbondanza reciproca nella laguna di Venezia. Tra gli obiettivi vi è la messa a punto di una metodica di analisi del DNA ambientale da campioni di acqua superficiale.

## 1.2. Il DNA ambientale

Negli ultimi anni i metodi di studio della biodiversità in diversi ambienti sono stati rivoluzionati dall'analisi dei frammenti di DNA purificati da campioni ambientali. L'analisi e il sequenziamento del DNA ambientale, internazionalmente noto come *environmental DNA* (abbreviato in "eDNA"), mediante metodi di sequenziamento massivo di seconda generazione, ha dato prova di essere estremamente valido e veloce nell'individuazione multi-specie in ecosistemi naturali e nella stima della loro abbondanza (Creer *et al.* 2016).

Le analisi di DNA ambientale prevedono una fase di campionamento e la successiva fase di estrazione di DNA dal campione ambientale; il DNA estratto viene poi amplificato mediante PCR, purificato e sequenziato tramite metodiche di *Next Generation Sequencing* (NGS).

Con DNA ambientale si intende un complesso insieme di frammenti di molecole di DNA presenti in diversi tipi di campioni ambientali (acqua, aria, terreno, ecc.) generalmente provenienti da diversi organismi e/o da degradazioni degli stessi (Taberlet *et al.* 2012); più precisamente, il DNA ambientale può contenere DNA cellulare proveniente da cellule o organismi viventi, e DNA extracellulare proveniente dalla naturale morte cellulare e la conseguente distruzione della struttura cellulare (Levy-Booth *et al.* 2007; Pietramellara *et al.* 2009). Uno schema generale dell'ecologia del DNA ambientale viene presentato in Fig. 2.



**Fig. 2** - Schema generale dell'ecologia del DNA ambientale, tratta da Barnes *et al.* 2015.



Normalmente, il campione ambientale viene processato al fine di catturare il DNA presente (ad esempio, i campioni d'acqua vengono filtrati attraverso appositi filtri) che viene successivamente purificato.

Se si considera l'analisi del DNA ambientale da campioni d'acqua, oggetto del presente lavoro di tesi, questa metodica può aumentare l'accuratezza e diminuire sensibilmente il costo delle normali indagini ecologiche, e allo stesso tempo permettere l'individuazione di specie molto rare o invasive (Valentini *et al.* 2009). L'enorme vantaggio della metodica infatti risiede nella possibilità di analizzare contemporaneamente il DNA proveniente da diverse specie e diversi individui senza la necessità di campionare direttamente ogni singolo individuo, riducendo così lo sforzo di campionamento. Inoltre, la possibilità di individuare più facilmente anche specie rare o elusive fornisce un ulteriore valore aggiunto, in quanto la loro individuazione e campionamento in acqua sono di solito molto difficoltosi (Mathon *et al.* 2021).

Al contempo, al di là del notevole interesse suscitato dalla metodologia, l'utilizzo del DNA ambientale è caratterizzato da un serie di limitazioni (Barnes & Turner, 2016). Uno degli aspetti primari da considerare nel momento in cui si svolge un'analisi di DNA ambientale, soprattutto in ambiente acquatico, è che individuare il DNA di una specie non corrisponde necessariamente al ritrovamento degli individui di quella specie. Il DNA infatti può essere trasportato lungo i corsi d'acqua e spostato dalle correnti (Barnes & Turner, 2016); si rileva dunque il DNA di una specie, ma il risultato va integrato con altri dati (es. rilevamenti precedenti della specie in modo inequivocabile) al fine di trarre delle considerazioni ecologiche coerenti, come svolto in Karahan e collaboratori (2017) e in Aglieri e collaboratori (2021).

Uno dei principali svantaggi della metodica è che il DNA ambientale si presenta degradato e poco concentrato; l'analisi richiede perciò messe a punto di protocolli ambiente specifici e la selezione di loci target di dimensioni ridotte, possibilmente in regioni conservate, in modo da favorire l'amplificazione simultanea da più specie grazie all'utilizzo di primer universali (Taberlet *et al.* 2018).

Il DNA ambientale inizia a decadere immediatamente dopo il suo distacco dall'organismo (Thomsen *et al.* 2012a, b) e questo processo può essere alimentato da forze meccaniche, attività microbiche e reazioni chimiche spontanee (Lindahl 1993; Nielsen *et al.* 2007). L'eDNA in acqua si presenta generalmente in basse concentrazioni (<200 pg/L) e può essere distribuito in maniera eterogenea all'interno dello specchio d'acqua (Takahara *et al.* 2012; Pilliod *et al.* 2013). Perciò, le probabilità di isolare il DNA ambientale a partire da campioni acquatici dipendono dall'efficacia dei metodi di

campionamento, di estrazione del DNA e dalla precisione dei saggi effettuati a valle (Goldberg *et al.* 2015). Oltre a tutti questi fattori, l'eDNA necessita di un database di riferimento adeguatamente costruito in funzione delle necessità dello studio (Taberlet *et al.* 2018).

In ogni caso, il DNA ambientale si è dimostrato, a parità di repliche, più efficiente rispetto ad altre metodiche di monitoraggio per l'attribuzione di specie ittiche (Aglieri *et al.* 2021).

Anche se l'utilità del DNA ambientale è stato ampiamente dimostrato, non si è ancora arrivati ad una standardizzazione dell'analisi ecologica dello stesso, soprattutto in ambiente marino (Aglieri *et al.* 2021). Le applicazioni più interessanti per il DNA ambientale riguardano non solo l'identificazione di specie presenti in un certo ambiente (Doi *et al.* 2017), ma anche la stima della taglia di una popolazione (Takahara *et al.* 2012) e le analisi genetiche delle popolazioni stesse (Barnes & Turner, 2016). Inoltre, l'individuazione di specie invasive dannose per l'ecosistema, che permette una risposta tempestiva alla gestione efficace dell'invasione (Lodge *et al.* 2006), rappresenta ormai un'applicazione fondamentale dell'analisi di DNA ambientale. Molti studi sono stati effettuati con successo per rilevare le specie invasive mediante eDNA in ecosistemi d'acqua dolce: ad esempio, sono state identificate le Lumache della Nuova Zelanda (*Potamopyrgus antipodarum*) in flussi d'acqua in Idaho, USA (Goldberg *et al.* 2013), cozze zebra del Caspio (*Dreissena polymorpha*) negli Stati Uniti centro-occidentali (Egan *et al.* 2013), i gamberi della Louisiana (*Procambarus clarkii*) in Francia (Tréguier *et al.* 2014), e il "Pesce gioiello" africano (*Hemichromis letourneuxi*) e pitone birmano (*Python bivittatus*) in Florida, negli stagni degli Stati Uniti (Díaz-Ferguson *et al.* 2014) e zone umide vicine a corsi d'acqua a Miami (Piaggio *et al.* 2014), rispettivamente.

Spear e collaboratori (2015) hanno utilizzato la quantificazione del DNA ambientale per investigare lo stato riproduttivo di rare specie anfibie; i risultati dimostrano che la quantità di eDNA è maggiore durante la stagione riproduttiva, suggerendo quindi la possibilità di utilizzare il DNA ambientale identificare gli aggregati di individui riproduttivi in natura (Goldberg *et al.* 2015).

In lavori precedenti è stata osservata una correlazione positiva statisticamente significativa tra la concentrazione di eDNA di una specie e il numero di individui di quella specie e biomassa. Infatti, assumendo che i vertebrati acquatici rilascino eDNA nell'acqua (da feci, secrezioni o tessuti) in proporzione alla loro biomassa, la misurazione del numero di copie di eDNA può essere usato per stimare la biomassa delle specie (Takahara *et al.* 2012).

Nonostante ciò, l'eDNA risulta utile soprattutto per dati di presenza/assenza, che possono essere usati per stimare la distribuzione delle specie. La stima della biomassa risulta infatti spesso difficile da effettuare in modo accurato, in particolare per organismi acquatici come i pesci (Begon *et al.* 2020). Nei pesci, la totalità della superficie del corpo è in diretto contatto con l'ambiente esterno e rilascia DNA nell'ambiente (Hansen *et al.* 2018). Per questo, a parità di biomassa ci aspettiamo che un grande numero di pesci piccoli propaghi più particelle di eDNA rispetto ad un unico pesce più grande. Inoltre, un principio generale da considerare è che il tasso metabolico è proporzionale alla massa; piccoli organismi avranno un tasso metabolico più grande rispetto agli organismi più grandi, considerando solo la massa (Maruyama *et al.* 2014), e rilasceranno quindi più DNA in acqua. Tuttavia, in alcune specie di pesci, negli stadi giovanili e larvarli la correlazione tra taglia e tasso metabolico è indipendente dalla massa: questo dato complica ulteriormente il tentativo di quantificare la correlazione tra scarto metabolico (eDNA incluso) e il numero degli individui, o la taglia del pesce (Post & Lee, 1996). Altri studi, come quello di Merkes e collaboratori (2014), hanno dimostrato che carcasse di pesci e feci dei predatori possono essere fonte di eDNA. Per tutti questi motivi, anche senza considerare gli effetti ambientali nella produzione, persistenza e trasporto del DNA, specialmente in ambienti marini risulta difficile stabilire una precisa relazione, spaziale e temporale, tra il DNA individuato e dati qualitativi/quantitativi di monitoraggio (Sassoubre *et al.* 2016).

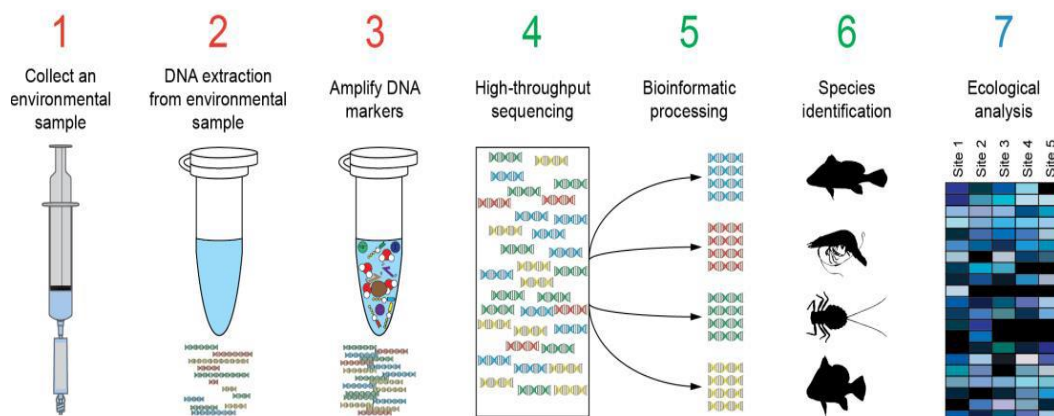
Inoltre, la maggior parte del DNA contenuto in un campione ambientale ha origini locali (del posto in cui è stato raccolto il campione); può accadere però che il DNA venga trasportato da siti vicini mediante dispersione attiva o passiva di organismi e propaguli, o mediante trasporto passivo di DNA extracellulare nell'acqua dovuto ai numerosi movimenti dell'acqua stessa. Per questo motivo, tenere in considerazione che il DNA ambientale si sposta è molto importante, soprattutto in ambienti acquatici. È stato visto infatti che il DNA ambientale di pesci e invertebrati può venire rilevato in acqua o sedimento raccolto anche a decine di chilometri a valle della fonte di eDNA, a causa della forza del flusso e della turbolenza dell'acqua (Deiner & Altermatt 2014; Jane *et al.* 2015). Così negli ambienti di transizione è possibile campionare eDNA che viene poi ricondotto sia a specie marine che a specie d'acqua dolce, trasportato dagli affluenti. Essendo gli ambienti acquatici influenzati dalle correnti, quando si svolge una analisi di DNA ambientale in acqua bisogna anche chiedersi se, una volta individuato il DNA di una specie, questo sia effettivamente localizzato nell'area di campionamento, o sia DNA trasportato dalle correnti (Hansen *et al.* 2018). Specialmente in ambienti acquatici bisogna quindi considerare gli effetti dell'ambiente sulla produzione, persistenza e trasporto del DNA ambientale; dovendo considerare tutti questi fattori, può essere difficoltoso

stabilire delle relazioni temporali e spaziali affidabili tra l'individuazione del DNA e precisi dati qualitativi e quantitativi (Sassoubre *et al.* 2016). Ad esempio, una variazione nel contenuto di ossigeno in acqua (Herbert & Steffensen, 2005) o di temperatura (Schurmann & Steffensen, 1997) sono fattori che causano cambiamenti fisiologici e che possono portare ad un'alterazione nel rilascio di DNA a livello ambientale (Hansen *et al.* 2018).

### 1.3. Fasi fondamentali dell'analisi di DNA ambientale

Il DNA ambientale può essere usato per diversi scopi: identificazione di una specie in particolare, identificazione di tutti i taxa appartenenti ad uno specifico gruppo, o per studiare le caratteristiche funzionali dei genomi mediante sequenziamento *shotgun*, senza quindi un target specifico, approccio utilizzato principalmente per i microrganismi (Simon & Daniel 2011).

In questo lavoro di tesi verrà approfondito il secondo protocollo, ovvero quello gruppo-specifico, che verrà utilizzato per gli esperimenti. In Fig. 3 si riportano le fasi dell'analisi di DNA ambientale: la raccolta dei campioni, l'estrazione del DNA e l'amplificazione del marcatore. Si procede poi con un sequenziamento massivo (HTS, *High-throughput sequencing*) e il risultato viene analizzato con strumenti bioinformatici, fino ad arrivare ad una identificazione delle specie. A questo punto è possibile procedere ad una analisi ecologica.



**Fig. 3** - Fasi fondamentali dell'analisi del DNA ambientale. Fonte: Taberlet *et al.* 2018

Le piattaforme di HTS risultano particolarmente adatte per l'analisi del DNA ambientale, che risulta spesso degradato; permettono infatti di sequenziare rapidamente il DNA da diversi tipi di campioni di DNA ambientale (Deiner *et al.* 2017).

Il sequenziamento di ampliconi tramite HTS, fornendo come risultato milioni di sequenze, presenta la capacità di rivelare tipicamente molte varianti di ogni sequenza presente nel mix di amplificati; oltre alla naturale variabilità delle sequenze, le varianti comprendono anche errori di PCR

(Schloss *et al.* 2011; Coissac *et al.* 2012), quali errori puntuali (sostituzioni nucleotidiche o piccole *indels*) o sequenze chimeriche (Boyer *et al.* 2016). È necessario quindi, per tutte le fasi dell'analisi di DNA ambientale, prevedere anche una serie di controlli di campionamento e sperimentali (Ficetola *et al.* 2016).

### 1.3.1. Campionamento

Un piano di campionamento appropriato è cruciale per il successo di qualsiasi studio. Per il DNA ambientale non c'è ancora un approccio universale, e la strategia di campionamento deve essere modificata rispetto al tipo di ambiente (per esempio, sarà diverso tra acquatico e terrestre), il quesito scientifico, considerazioni statistiche, logistiche e finanziarie, ed eventualmente i risultati di uno studio pilota (Taberlet *et al.* 2018). Inoltre, i campioni devono essere raccolti con l'obiettivo di poterli confrontare tra di loro per rispondere alla domanda biologica di interesse: se il quesito comprende la variabile temporale, sarà importante campionare in momenti diversi dello stesso anno, o nello stesso momento, ma in anni diversi. Se il quesito è maggiormente legato alla variazione del DNA ambientale in spazi diversi, il campionamento si dovrà concentrare maggiormente sulla variazione dei siti di campionamento. In qualsiasi caso, in studi precedenti (McDonald *et al.* 2004) si è notata una correlazione tra il numero di repliche di campionamento e le specie individuate: prevedere un adeguato numero di repliche di campionamento, anche di uno stesso luogo e momento, risulta quindi fondamentale per una stima maggiormente accurata della biodiversità presente; un numero insufficiente di campioni raccolti può infatti portare ad una sottostima delle specie presenti in una comunità (Evans *et al.* 2017).

Nell'ottica di fornire delle linee guida per gli studi di DNA ambientale in ambienti acquatici, studi precedenti (Golberg *et al.* 2015, Taberlet *et al.* 2018) sottolineano l'importanza di evitare il più possibile le fonti di contaminazione. È consigliabile quindi utilizzare equipaggiamenti monouso che non siano stati a contatto con fonti di DNA ad alta concentrazione, come gli ambienti e gli strumenti di laboratorio post-PCR. In alternativa, viene consigliato l'utilizzo ripetuto degli strumenti necessari al campionamento (per esempio, le taniche per la raccolta dell'acqua), previo risciacquo con candeggina diluita 1:10 (Taberlet *et al.* 2018). Le contaminazioni, seppur in minima parte, risultano comunque inevitabili e al fine di quantificarne l'entità e adattare le analisi successive di conseguenza, è necessario includere in fase di campionamento anche dei bianchi di filtrazione, che corrispondono a filtri nei quali viene fatta passare acqua deionizzata; il DNA amplificato che deriva dall'estrazione del DNA da tali filtri rappresenta

un'indicazione del livello di contaminazione presente in sede di campionamento. (Ficetola *et al.* 2016)

A differenza di un campione di eDNA proveniente da feci o suolo, quello proveniente dall'acqua è molto diluito e lavorare con piccoli volumi d'acqua non garantirebbe la cattura di un numero amplificabile di molecole. È quindi molto comune filtrare da 10 a 100 litri di acqua utilizzando filtri o capsule filtranti (Valentini *et al.* 2016; Zinger *et al.* 2012).

Una volta ultimata la filtrazione, i filtri possono essere conservati in vari modi: congelamento (Jerde *et al.* 2011), immersione in etanolo (Goldberg *et al.* 2011), disidratazione, o immersione in un buffer di lisi (Renshaw *et al.* 2015). La quantità finale di dati ottenuti dall'analisi può dipendere dai metodi di preservazione utilizzati, dal materiale e dalla porosità dei filtri (Deiner *et al.* 2015; Majaneva *et al.* 2018).

### 1.3.2. Estrazione di DNA genomico da filtro

Nella maggior parte degli studi di DNA ambientale proveniente da campioni acquatici viene utilizzato il kit di estrazione *DNeasy Blood & Tissue Kit* (Qiagen), che si è dimostrato essere il più efficiente per l'estrazione di DNA ambientale da filtro (Jeunen *et al.* 2019; Goldberg *et al.* 2016). I protocolli di estrazione da DNA ambientale sono stati indagati a fondo, e secondo gli esperimenti documentati in letteratura, sembra che le estrazioni di eDNA si possano ottenere con successo senza particolare attenzione riguardo il pH dell'acqua e che l'enzima più adatto all'estrazione di DNA da filtro sia la proteinasi K (Tsuji *et al.* 2017).

Al fine di evitare i contaminanti che possono portare a falsi positivi, che rappresentano uno dei limiti del DNA ambientale, si rendono necessari specifici accorgimenti (Ficetola *et al.* 2016).

È necessario infatti che ci sia completa separazione (stanze diverse) tra l'area di pre-PCR e post-PCR e in queste due bisogna usare anche diversi camici e guanti; come regola generale, nulla deve passare dalla post-PCR alla pre-PCR. È necessario tenere sempre a mente il rischio di contaminazione, e per questo cambiare guanti frequentemente, pulire in modo sistematico tutti gli strumenti, in particolare le pipette, che sono la principale fonte di contaminazione; per questo vanno usati sempre puntali con filtro, sia per le estrazioni che per la preparazione dei campioni per la PCR. In aggiunta a queste misure preventive, poiché un certo livello di contaminazione è inevitabile durante la procedura di estrazione, è consigliato effettuare dei controlli negativi di estrazione; questi corrispondono ad una “falsa” estrazione, svolta contemporaneamente a quelle “vere”, usando gli stessi reagenti delle estrazioni normali, ma senza inserire alcun campione (Aglieri *et al.* 2021; Taberlet *et al.* 2018; De Barba

*et al.* 2014): in questo modo, eventuali specie identificate nei controlli negativi di estrazione corrisponderanno a contaminazioni di fondo presenti nei reagenti utilizzati e a contaminazioni legate all'operatore e/o al laboratorio nel quale sono stati svolti gli esperimenti.

### 1.3.3. Amplificazione del DNA con primer gruppo specifici e DNA *metabarcoding*

Mentre con il termine *barcoding* si intende l'identificazione tassonomica di una specie basata su un marker diagnostico (Hebert *et al.* 2003), per le analisi di DNA ambientale risulta più informativa la metodica di DNA *metabarcoding*, con la quale si intende l'identificazione tassonomica di più specie estratte da un campione di DNA ambientale, che è stato amplificato tramite PCR e sequenziato mediante una piattaforma *high-throughput* (Deiner *et al.* 2017). Il *metabarcoding* si caratterizza per l'identificazione di numerosi taxa in un campione ambientale mediante il sequenziamento di ampliconi generati da primer universali (Deagle *et al.* 2009). Associando un HTS con la possibilità di associare le sequenze di eDNA ad un'unità tassonomica, il DNA *metabarcoding* offre uno strumento molecolare potente in grado di indagare in modo non invasivo l'abbondanza di più specie da molti ecosistemi diversi (Deiner *et al.* 2017). L'eDNA *metabarcoding* è considerata una tecnica di valutazione biologica sensibile, perché riesce a individuare l'abbondanza dei taxa tramite l'identificazione di frammenti taxa specifici raccolti in campioni ambientali (Ficetola *et al.* 2008; Bohmann *et al.* 2014; Rees *et al.* 2014) ed è risultata in grado di individuare numerose specie (Thomsen & Willerslev, 2015).

Risulta cruciale in questa fase la scelta del metabarcode più appropriato. Il DNA metabarcode perfetto è un frammento di DNA più corto possibile che mostri una sequenza altamente variabile, affiancata da due regioni conservate. La regione centrale più variabile risulta utile per discriminare tra tutte le specie appartenenti al gruppo target, perché quella regione è specie-specifica e non viene condivisa con altre specie. Le due regioni laterali sono invece conservate all'interno del gruppo target, ma differiscono tra gruppi. Le regioni maggiormente conservate corrispondono a quelle con le quali il primer per il DNA *metabarcoding* si deve appaiare perfettamente. Quando si sceglie il metabarcode bisogna anche considerare che la tecnica di *metabarcoding*, per essere efficace, necessita di un database di riferimento sul quale confrontare i dati. In assenza, all'interno del database, di dati di sequenze relativi al marcatore selezionato, non è possibile utilizzare questo metodo con la massima efficacia. (Taberlet *et al.* 2018)

Per quanto riguarda l'aspetto dell'efficacia della reazione di PCR, AmpliTaq Gold® 360 Master Mix (Applied Biosystems, Foster City, CA, USA), testata precedentemente per studi di *metabarcoding*, è una *hot start, non-proofreading polymerase* che permette di ottenere risultati ottimali per quanto riguarda l'aspetto quantitativo. Le *Taq* polimerasi rappresentano un buon compromesso tra il livello di artefatti che possono derivare dall'amplificazione, l'aspetto quantitativo e il costo per la procedura. In una *hot start Taq* polimerasi viene utilizzato uno specifico anticorpo per bloccare l'azione dell'enzima a basse temperature. Per questo, è necessario uno step iniziale a 95-96°C per rimuovere l'anticorpo e attivare la polimerasi. Questo processo previene l'allungamento di frammenti aspecifici di DNA durante l'iniziale aumento della temperatura, prima dello step di denaturazione, e in questo modo riduce significativamente la formazione di dimeri di primer (Taberlet *et al.* 2018).

Anche durante la reazione di PCR è molto importante evitare l'incorrere di falsi positivi o falsi negativi. Una delle strategie per limitare il numero di entrambi è quella di aumentare il numero di replicati tecnici, ovvero prodotti di diverse reazioni di PCR che presentano lo stesso estratto di DNA. L'inclusione di replicati tecnici nel protocollo porta ad una valutazione nettamente migliore delle differenze tra le diverse metodologie e riduce le incertezze durante l'analisi dei dati (Prosser 2010); inoltre, offre l'aumento della probabilità di individuazione di specie in situazioni in cui il DNA è estremamente diluito (Giguet-Covex *et al.* 2014; Pansu *et al.* 2015b; Valentini *et al.* 2016).

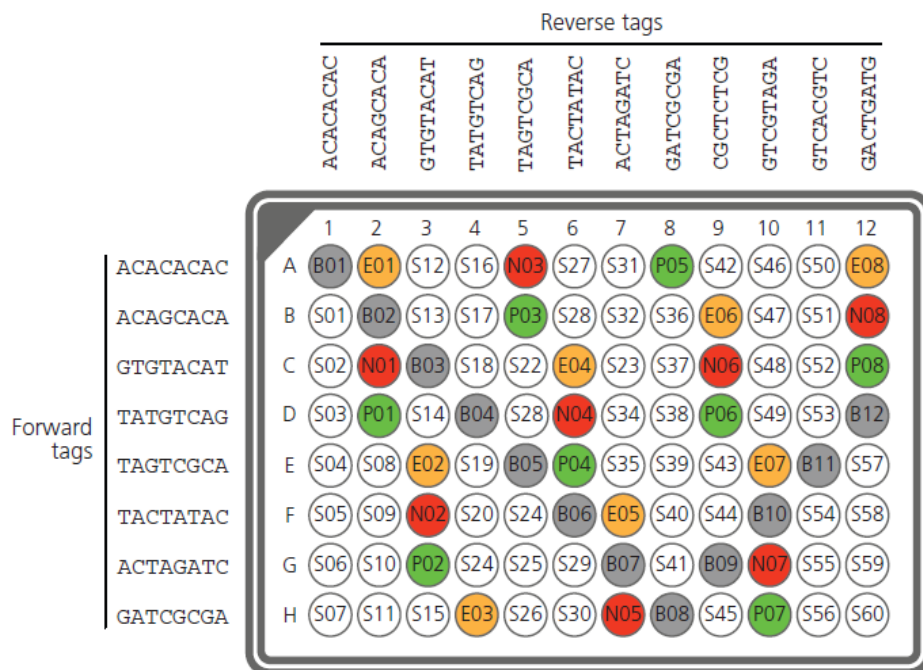
L'utilizzo della più adatta *Taq* polimerasi e l'inclusione di replicati tecnici però non previene del tutto la formazione di aspecifici o chimere. Il rischio di contaminazione infatti va considerato per ogni reazione di PCR e può portare alla formazione di falsi positivi (Kwok & Higuchi 1989). Per questo motivo è cruciale, in ogni esperimento di DNA *metabarcoding*, l'inclusione di appropriati controlli al fine di interpretare correttamente i risultati. In un esperimento si dovrebbero inserire dal 10% al 20% di controlli, di 5 diversi tipi (Taberlet *et al.* 2018). I controlli negativi di estrazione, di cui si è parlato nel paragrafo precedente; i controlli negativi di PCR, nei quali l'aggiunta di DNA template, come il DNA estratto, viene sostituito da una addizione di acqua DNA-free. Questi permettono di individuare potenziali contaminanti che possono insorgere durante lo step di amplificazione. I controlli positivi di PCR, che devono essere campioni con un DNA il più simile possibile a quello amplificato nei campioni e servono per verificare che sia avvenuta correttamente l'amplificazione del DNA estratto.



Per sfruttare le capacità del sequenziamento di seconda generazione di sequenziare contemporaneamente ampliconi da numerosi campioni, gli estratti vengono amplificati tramite PCR usando primer generici, ai quali vengono aggiunti dei tag (Binladen *et al.* 2007) costituiti da corte sequenze di circa 4-8 basi. In questo modo, i tag lasciano “un’impronta” specifica negli ampliconi derivati dalla PCR, i quali possono quindi essere uniti per il sequenziamento e successivamente ricondotti, tramite il tag specifico, al campione dal quale sono stati originati (Schnell *et al.* 2015).

Numerosi studi che utilizzano il *metabarcoding* hanno notato il manifestarsi di false combinazioni di tag usati nel sequenziamento Illumina (Esling *et al.* 2015), fenomeno riconducibile a sporadici eventi di contaminazione incrociata di primer portatori di indici diversi e chimere, che si sono verificate esclusivamente negli esperimenti dove si trovavano le librerie di sequenziamento da più campioni amplificati contemporaneamente (Kircher *et al.* 2011).

Quando diversi campioni condividono gli stessi tag, forward o reverse, è molto importante quindi includere nell’esperimento alcune combinazioni di tag non utilizzate. Guardando poi, durante l’analisi informatica, alle sequenze attribuite a quella combinazione di tag che non è stata utilizzata, è possibile stimare l’occorrenza di chimere lungo i campioni, chiamati *tag jumps* (Schnell *et al.* 2015). In Fig.4, un esempio schematizzato di una piastra per l’amplificazione.



**Fig. 4** - Esempio esplicativo di una piastra per un esperimento di PCR. In grigio, i bianchi di sequenziamento contenenti i tag non utilizzati. In arancione, i bianchi di estrazione; in rosso i controlli negativi di PCR e in verde i controlli positivi. In bianco i campioni, numerati da 01 a 60. Sopra e a lato vengono raffigurati i tag campione-specifici (Taberlet *et al.* 2018).

#### 1.3.4. *Blocking primer*

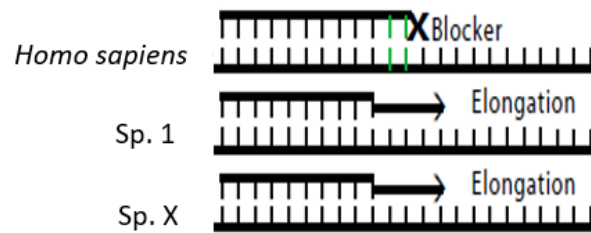
Nonostante l'eDNA *metabarcoding* generalmente sia più performante rispetto ai metodi convenzionali di monitoraggio per i pesci (Shaw *et al.* 2016; Cilleros *et al.* 2019), la diversità di sequenze in un campione che può essere individuato da primer universali è spesso compromesso da concentrazioni molto alte di qualche template di DNA; le sequenze più rare non vengono quindi amplificate, perché la reazione di PCR favorisce le dominanti (Harper *et al.* 2019; Bylemans *et al.* 2019). Se il DNA presente a concentrazioni più alte risulta essere quello di contaminanti, può competere in PCR con il DNA di interesse; questo si riflette in una eccessiva individuazione delle specie più abbondanti, a discapito di quelle rare (Shehzad *et al.* 2012). Un modo per ridurre l'amplificazione di alcune specie molto abbondanti, al fine di rivelare quelle più rare, è l'aggiunta di un *PCR blocking oligonucleotide* (Vestheim & Jarman 2008), o *blocking primer*. Questo approccio viene utilizzato spesso per bloccare l'amplificazione di DNA contaminante (Boessenkool *et al.* 2012) o, in studi sulla dieta degli organismi, per bloccare il DNA del predatore (Vestheim & Jarman, 2008; Shehzad *et al.* 2012).

Studi svolti da Rojahn e collaboratori (2021) dimostrano che l'utilizzo del *blocking primer* rappresenta un modo per diminuire la dimensione campionaria necessaria all'individuazione di tutte le specie anche del 25%, mentre Boessenkool e collaboratori (2012) hanno dimostrato chiaramente come l'utilizzo di un *blocking primer* per l'uomo (*Homo sapiens*) abbia permesso l'amplificazione di tre specie in più, che non erano state identificate senza l'utilizzo di questo primer.

Il *blocking primer* deve sovrapporsi ad uno (F o R) dei primer universali di amplificazione, e in modo ottimale dovrebbe trovarsi all'estremità del frammento di PCR, in cui c'è maggiore variazione tra la sequenza della specie che deve essere bloccata e la sequenza della specie che invece si vuole amplificare. Particolare attenzione va riservata alla concentrazione del *blocking primer* di utilizzo, per permettere il blocco dell'amplificazione del contaminante, ma allo stesso tempo permettere al primer universale di amplificare le specie di interesse; quella ottimale sembra essere 10 volte superiore al primer universale utilizzato per l'amplificazione (Taberlet *et al.* 2018; Vestheim & Jarman 2008; Wilcox *et al.* 2014; Su *et al.* 2018).

Ci sono modi diversi per costruire un *blocking primer*, ma il metodo più efficace risulta essere la modifica del primer che si lega alla sequenza contaminante mediante uno *spacer 3C* al 3' dell'oligonucleotide; la modifica previene l'*elongation* durante la PCR, senza però influenzare in modo

significativo le proprietà di *annealing* (Vestheim & Jarman 2008; Huggins *et al.* 2020) (Fig. 5).



**Fig. 5** - Immagine esplicativa del funzionamento di un blocking primer. Mentre le sequenze delle altre specie vengono amplificate mediante l'utilizzo del primer universale, il blocking primer blocca l'amplificazione del DNA di *Homo sapiens*. Tratta da Vestheim & Jarman, 2008.

Il *blocking primer* si allinea così alla sequenza della specie di interesse, competendo con il primer universale e bloccando l'amplificazione della stessa grazie alla modifica, che permette di interrompere la fase di allungamento nella reazione di PCR (Huggins *et al.* 2020).

### 1.3.5. Analisi bioinformatiche

Qualsiasi sia il quesito ecologico alla base dell'analisi, le metodiche di DNA *metabarcoding* fanno affidamento a sequenziamenti NGS e generano un considerevole numero di sequenze di DNA (tipicamente milioni di *reads*). La manipolazione di un così grande numero di dati richiede l'utilizzo di pacchetti dedicati (Boyer *et al.* 2016). Mathon e collaboratori (2021) hanno identificato in Obitools un pacchetto che restituisce dei dati affidabili dall'analisi di DNA *metabarcoding*; l'affidabilità del pacchetto si basa soprattutto sulla possibilità di utilizzare numerosi algoritmi per filtrare e ordinare le *reads* derivanti dal sequenziamento, possibilità che consente di adattare i comandi alle numerose applicazioni del *metabarcoding* (Boyer *et al.* 2016). Obitools è stato usato in numerosi studi (es. Ficetola *et al.* 2010; Kowalczyk *et al.* 2011; Yoccoz *et al.* 2012; Bellemain *et al.* 2013), dimostrando di essere consono alle numerose applicazioni del DNA *metabarcoding*.

Obitools permette di svolgere le analisi in diverse fasi; la prima prevede un allineamento delle sequenze derivate dal sequenziamento, in base a punteggi di similarità definiti dall'operatore. Successivamente, le sequenze allineate vengono incrociate con un file di input che presenta ciascun campione e il relativo tag utilizzato per ricondurre le sequenze ad uno specifico campione. Da questa prima analisi derivano quindi le sequenze identificate, ovvero quelle che presentano il metabarcode utilizzato. Successivamente, queste sequenze vengono incrociate con un database

specifico, al fine di ottenere l'assegnazione tassonomica più precisa possibile (Boyer *et al.* 2016). Particolarmente interessante risulta il programma *ecoPCR* al suo interno, che permette di costruire un database specifico per il *metabarcoding* utilizzato durante le analisi estraendo le regioni di DNA di interesse da un database scaricato precedentemente (EMBL) (De Barba *et al.* 2014).

Una volta avvenuta l'assegnazione tassonomica, il pacchetto fornisce in output un file in cui ad ogni sequenza presenta esplicitate le sue caratteristiche (count della sequenza, riconduzione al campione, assegnazione tassonomica con percentuale di identità).

A partire da questo, le analisi statistiche ed ecologiche possono essere effettuate con R (RStudio Team 2021. RStudio: Integrated Development for R. RStudio, PBC, Boston), mediante utilizzo di pacchetti specifici, come *vegan*, *phyloseq*, *cluster*.

## 2. Scopo della tesi

Il presente lavoro di tesi ha come scopo l'analisi del DNA ambientale a partire da campioni d'acqua prelevati da diversi siti della Laguna di Venezia, un ambiente di transizione che riveste una notevole rilevanza dal punto di vista ecologico e commerciale, al fine di censire la fauna ittica presente, fornendo informazioni sulla biodiversità presente e, potenzialmente, sulla "salute" di questo habitat. In particolare, attraverso l'analisi di campioni prelevati dal 2018 al 2021, in una fase precedente alla piena operatività del sistema di paratie mobili del MOSE, si intende standardizzare la metodica di analisi del DNA ambientale e fornire una prima indicazione delle specie presenti e della loro variazione spazio-temporale. I risultati ottenuti potrebbero rappresentare la base per il monitoraggio a lungo termine della composizione della fauna ittica attraverso uno strumento innovativo.

Nella fase di ottimizzazione del protocollo ci si è concentrati anche su un aspetto tecnico importante per limitare l'amplificazione di contaminanti, problema frequentemente riscontrato con l'analisi del DNA ambientale con l'approccio passivo, come nel presente caso, mediante l'utilizzo di primer conservati e analisi *metabarcoding*. Una parte del lavoro si è quindi concentrata sulla messa a punto e la valutazione dell'efficienza di *blocking primer* contro frammenti mitocondriali di origine umana.

### 3. Materiali e metodi

#### 3.1. Campionamento

Nel presente lavoro di tesi sono stati analizzati campioni d'acqua superficiale raccolti periodicamente in diversi siti della Laguna di Venezia. I campioni analizzati sono relativi a due differenti cicli di campionamento (1 e 2). Il primo ciclo di campionamento è stato effettuato dal 13/11/2018 al 18/02/2020, ha avuto una cadenza mensile, ed è stato purtroppo interrotto in seguito alle restrizioni introdotte dall'emergenza COVID-19. Il secondo è stato svolto dal 15/11/2018 al 19/05/2021, ed ha avuto una cadenza quadrimestrale. Come risultato si hanno due serie di campioni, parzialmente sovrapposte a livello temporale e spaziale.

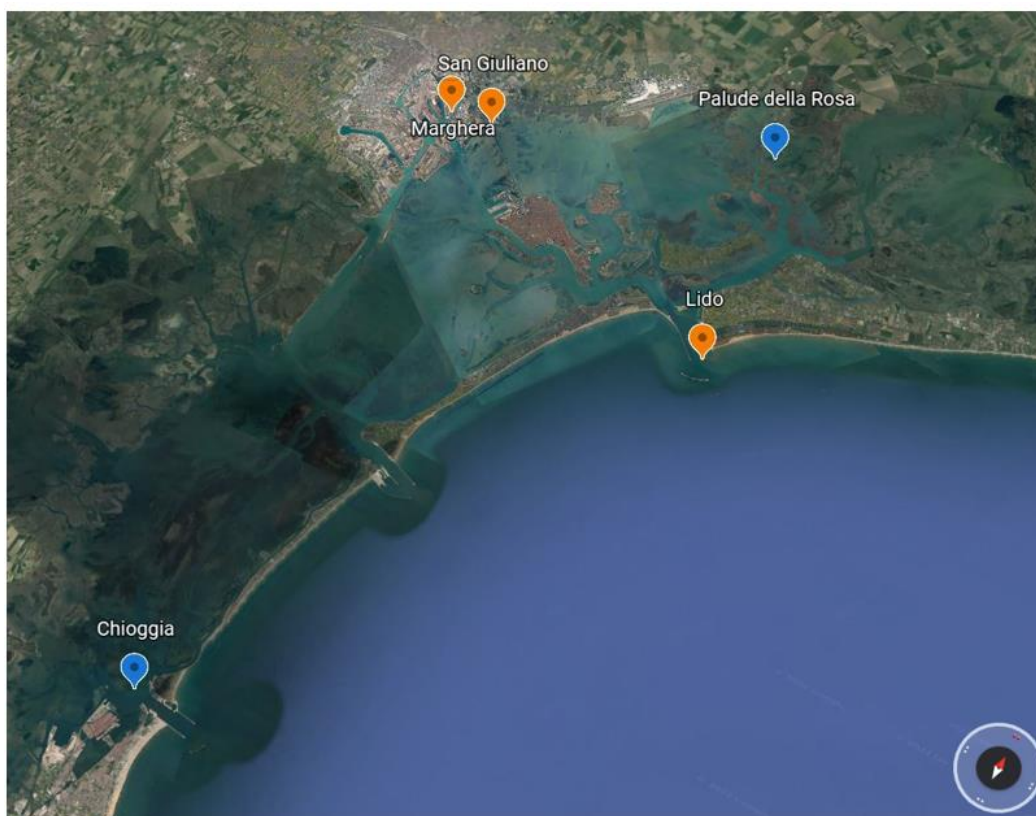
Il primo ciclo, campionamento 1, ha interessato due siti della Laguna: Palude della Rosa (Stazione 5) e Chioggia (Stazione 15) (Fig. 6).

In totale, 124 campioni (filtri) sono stati raccolti con cadenza mensile in collaborazione con il gruppo di ricerca della Dottoressa Camatti dell'ISMAR-CNR di Venezia; 69 campioni sono stati processati in un precedente lavoro di tesi (Cananzi 2019) e inclusi nel presente lavoro per le analisi dei dati di sequenziamento, i restanti campioni, non processati in precedenza, sono stati processati e analizzati in questa tesi.

Per il campionamento 1, i campioni di acqua sono stati raccolti in taniche precedentemente lavate con candeggina diluita 1:10 e risciacquate con acqua deionizzata (milliRO), al fine di evitare la loro contaminazione con altro DNA eventualmente presente nelle taniche in precedenza. Le taniche sono state poi trasportate al dipartimento di Biologia dell'Università di Padova e l'acqua al loro interno è stata filtrata entro 24 ore dalla raccolta in laguna. Ogni campione è stato filtrato utilizzando due tipologie di filtro: uno in fibra di vetro (*Whatman<sup>TM</sup>*, GF/C) con maglia 1.2  $\mu\text{m}$ ; l'altra tipologia è l'acetato di cellulosa (*Sartorius<sup>TM</sup>*) con maglia 0.45  $\mu\text{m}$ ; entrambe le tipologie di filtro presentano un diametro di 47 mm. Per ogni data di campionamento e per ogni sito sono stati prodotti anche dei bianchi di filtrazione, ovvero filtri nei quali è stata filtrata solo acqua deionizzata.

I filtri sono stati successivamente conservati, avvolti in carta stagnola e chiusi in una scatola, in un freezer a -80°C.

Il secondo ciclo, campionamento 2, ha interessato tre siti aggiuntivi, corrispondenti a Lido, Marghera e San Giuliano (Fig. 6). I 157 campioni totali sono stati raccolti esclusivamente nei mesi di novembre, febbraio, maggio e agosto in collaborazione con il gruppo di ricerca del Dottor Vezzi (che si ringrazia) del Dipartimento di Biologia dell'Università di Padova, da novembre 2018 a maggio 2021.



**Fig. 6** - Siti di campionamento in laguna di Venezia. In blu, i siti di campionamento in comune tra campionamento 1 e 2. In arancione, i siti di campionamento aggiuntivi del campionamento 2. Modificata da: <https://www.google.it/intl/it/earth/>

Per ogni giornata di campionamento, i campioni d'acqua sono stati filtrati in replicato utilizzando tre tipologie di filtri: filtri a membrana in polietersulfone *Millipore Express PLUS*, con maglia 0.22  $\mu\text{m}$  e diametro 142 mm, filtri a membrana *Isopore* in policarbonato, con maglia 3.0  $\mu\text{m}$  e diametro 142 mm, filtri a rete in Nylon, con maglia 20  $\mu\text{m}$  e diametro 90 mm. A differenza dei precedenti, questi campioni sono stati conservati in freezer a  $-80^{\circ}\text{C}$  in provette falcon con capacità 50 ml.

Nel presente lavoro di tesi sono stati processati a partire dall'estrazione di DNA da filtro la quasi totalità dei campioni del campionamento 2 e i rimanenti non ancora processati del campionamento 1. In totale sono stati analizzati 166 filtri, dei quali 36 sono stati sequenziati in duplicato al fine di verificare l'efficienza del blocking primer disegnato.

Inoltre, un totale di 20 campioni di controllo sono stati inseriti sia in fase di preparazione della libreria, sia in fase di sequenziamento. I campioni di controllo includono bianchi di filtrazione e bianchi di estrazione.

In Tabella 1 viene indicata la numerosità campionaria per periodo e sito dei due campionamenti.

**Tabella 1** - Tabella riassuntiva dei filtri derivati dai campionamenti 1 e 2.

<b>Periodo</b>	<b>Sito</b>	<b>Filtri</b>
	Chioggia	38
	Palude della Rosa	10
Inverno 2018 - 2019	Lido	6
	Marghera	6
	San Giuliano	6
	Chioggia	10
	Palude della Rosa	24
Primavera 2019	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	7
	Palude della Rosa	15
Estate 2019	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	5
	Palude della Rosa	15
Autunno 2019	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	5
	Palude della Rosa	23
Inverno 2019 – 2020	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	3
	Palude della Rosa	3
Primavera 2020	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	3
	Palude della Rosa	3
Estate 2020	Lido	3
	Marghera	3
-	San Giuliano	3

	Chioggia	3
	Palude della Rosa	3
Autunno 2020	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	3
	Palude della Rosa	3
Inverno 2020 – 2021	Lido	3
	Marghera	3
	San Giuliano	3
	Chioggia	3
	Palude della Rosa	3
Primavera 2021	Lido	3
	Marghera	3
	San Giuliano	3
<b>10 periodi</b>	<b>5 siti</b>	<b>281 filtri</b>

Per i dettagli sui campioni analizzati, fare riferimento all'appendice I.

### 3.2. Estrazione di DNA genomico da filtro

Per 36 campioni del campionamento 1 e 104 campioni del campionamento 2 si è proceduto con l'estrazione di DNA da mezzo filtro utilizzando il kit di estrazione *DNeasy Blood & Tissue Kit* (Qiagen) il cui protocollo è stato adattato, al fine di renderlo più compatibile con l'estrazione da filtro. Ogni metà filtro è stata tagliata e privata del bordo bianco (che non ha filtrato l'acqua), tagliata in piccoli pezzi e immersa nel buffer di lisi. In particolare, il volume del buffer di lisi è stato aumentato di un fattore 3 in modo che tutto il filtro, precedentemente tagliato in piccoli pezzetti, fosse completamente sommerso. Di conseguenza, è stato adattato il volume della proteinasi K. Per il resto del protocollo sono state seguite le indicazioni standard riportate nel manuale, tranne l'eluizione che è stata effettuata in 100 µl, al fine di concentrare il campione. Per la fase di lisi è stato utilizzato un Thermomixer (Thermomixer 2.0, Eppendorf) e i passaggi che richiedono la centrifugazione stati eseguiti, nei tempi indicati dal protocollo Qiagen, utilizzando una centrifuga 5415C (Eppendorf).

Durante ogni estrazione sono stati sempre inclusi due bianchi di estrazione, corrispondenti a campioni processati nello stesso modo senza il filtro di partenza. L'esito dell'estrazione è stato valutato mediante quantificazione al Nanodrop 2000 (ThermoFisher). Gli estratti sono stati conservati a -20°C sino al momento dell'analisi.



Il DNA di 53 campioni appartenenti al campionamento 2 era stato estratto precedentemente presso il laboratorio del Dottor Vezzi mediante il kit di estrazione *AllPrep DNA/RNA/miRNA Universal Kit* (Qiagen), e di questi sono state prelevate delle aliquote per il successivo processamento.

Nei campioni di DNA estratti da filtro la concentrazione di DNA attesa è relativamente bassa, e la possibilità di incorrere in contaminazioni aumenta (Taberlet *et al.* 2018). Al fine di limitare la contaminazione dei campioni con eventuale DNA di altre specie analizzate in laboratorio, tutte le estrazioni sono state condotte in una stanza completamente separata dal laboratorio, dedicata esclusivamente all'estrazione di campioni per analisi di DNA ambientale, munita di ricambio d'aria a pressione positiva. Sono stati utilizzati camice, guanti e mascherina chirurgica per tutta la durata degli esperimenti. Prima dell'inizio e dopo la fine di ogni giornata di esperimenti la stanza è stata pulita da eventuali tracce di DNA presente con candeggina diluita 1:10 e carta pulita. La pulizia ha compreso anche i mobili (vetrine e sedie) sempre presenti nella stanza. Nel caso in cui, nella stessa giornata, siano stati estratti campioni provenienti da diversi siti o date, la stanza è stata pulita completamente anche tra un'estrazione e l'altra. Allo stesso modo sono stati puliti con candeggina tutti gli strumenti utilizzati per l'estrazione (Termomixer, centrifuga, pipette) e gli strumenti spostabili (forbici, pinze, rack ecc.) sono stati, prima di ogni estrazione, lasciati almeno 15 minuti sotto una cappa a raggi UV prima della pulizia. Durante tutti gli esperimenti effettuati sono stati utilizzati esclusivamente puntali sterili con filtro.

### 3.3. Amplificazione di 12S rDNA (Tele02)

Per l'analisi del DNA ambientale è stato scelto di amplificare una piccola porzione di un gene mitocondriale. Nel dettaglio, per l'amplificazione degli estratti è stata utilizzata una coppia di primer disegnata specificatamente in modo tale da amplificare preferenzialmente l'Infraclasse dei Teleostei (Tele02) presentata da Taberlet e collaboratori (2018) nel libro "*Environmental DNA for biodiversity research and monitoring*", nell'appendice relativa ai primer. La coppia di primer Tele02 (Tabella 2) amplifica la regione mitocondriale del 12S rDNA, producendo frammenti che vanno da 180bp a 300bp circa. Il profilo termico è stato ottimizzato mediante simulazioni di PCR *in silico*, ma non è ancora stato sufficientemente testato empiricamente.

**Tabella 2** - In tabella è riportato il primer utilizzato con il nome, la sequenza, la regione genomica che amplifica e le dimensioni attese del frammento indicate in paia di basi (minime/medie/massime).

Nome	Sequenza	Regione	Dim. Frammenti
Tele02	F: 5'-AAACTCGTGCCAGCCACC-3' R: 5'-GGGTATCTAATCCCAGTTTG-3'	12S <u>rDNA</u>	182/247/298

Il protocollo di preparazione della libreria è stato selezionato tra quelli proposti da Taberlet *et al.* (2018) e consiste in un'unica reazione di amplificazione del frammento target da ogni filtro. Gli ampliconi filtro-specifici sono stati poi uniti in un unico *pool* nel loro volume totale, purificati e sequenziati.

Al fine di permettere, in fase di post-processamento dei dati, di differenziare gli ampliconi provenienti da ogni singolo campione, per ogni filtro sono stati utilizzati in questa tesi dei primer Tele02 modificati con l'aggiunta di indici specifici (*barcode* di 8 basi) in posizione 5' di ogni primer. Gli ampliconi di ogni campione includono quindi alle estremità una coppia di barcode univoca e specifica per ogni filtro, denominata in seguito "tag".

In totale sono stati ordinati 32 Primer Tele02 Forward e 12 Primer Tele02 Reverse, per un totale di 348 possibili combinazioni di tag.

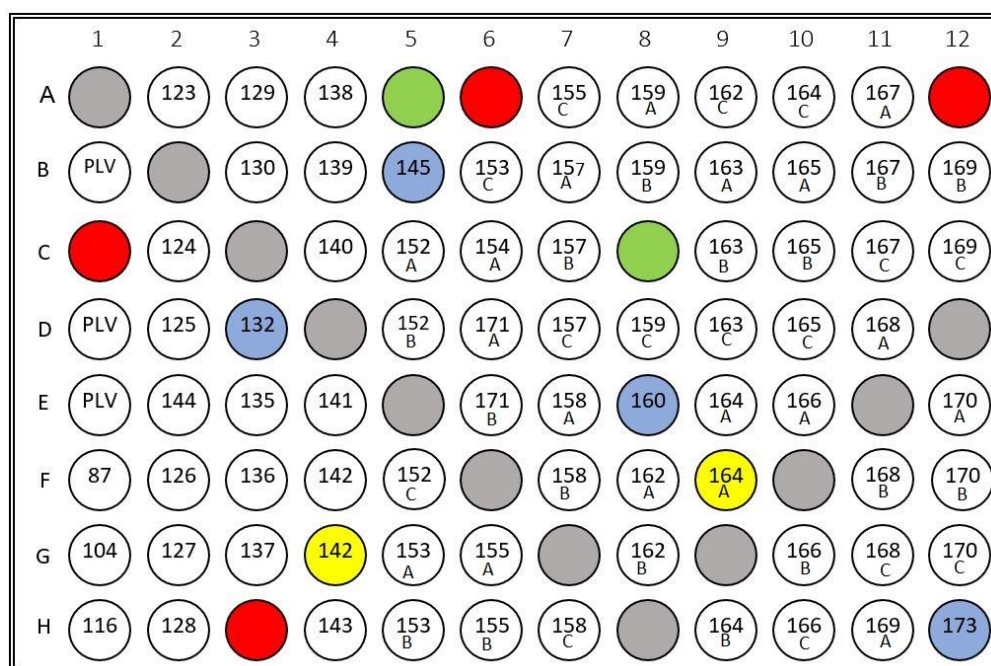
In questo lavoro di tesi sono state sequenziate due piastre da 96 pozzetti ciascuna; ogni campione presente nelle piastre è stato amplificato in triplicato e unito in un solo pozzetto dopo l'amplificazione. Al fine di evitare l'incorrere di fenomeni di amplificazione preferenziale casuale di alcuni frammenti durante il processo di PCR (Prosser 2010; Searle *et al.* 2016) o di mancata amplificazione di altri, sono state effettuate tre repliche tecniche per ogni campione delle due piastre, che presentano quindi anche la stessa combinazione di tag; una volta avvenuta l'amplificazione, le repliche tecniche state unite, nel loro volume totale, nella stessa colonna della piastra.

Per il primo sequenziamento, l'amplificazione di 69 campioni è avvenuta in un volume finale di 10 µl contenente AmpliTaq Gold™ 360 Master Mix 1X (Applied Biosystems), ciascun primer Tele02, forward e reverse, a concentrazione finale 0.5 µM. Ai reagenti sono stati aggiunti 2µl di DNA template estratto dal campione. Per i restanti 36 campioni, processati nel secondo sequenziamento, l'amplificazione è avvenuta agli stessi volumi e con le stesse concentrazioni dei precedenti, ma è stata dimezzata la quantità di DNA template presente (1µl).

Nel secondo sequenziamento sono stati inoltre inseriti gli stessi 36 campioni amplificati attraverso un protocollo ottimizzato, che includeva l'utilizzo di Blocking primer all'interno della reazione di PCR. L'amplificazione è avvenuta in un volume finale di 10 µl contenente AmpliTaq Gold™ 360

Master Mix 1X (Applied Biosystems), ciascun primer Tele02, forward e reverse, a concentrazione finale 0.5  $\mu$ M. Il blocking primer è stato aggiunto a concentrazione finale 5  $\mu$ M, 10X rispetto ai primer universali. Ai reagenti è stato aggiunto 1  $\mu$ l di DNA template estratto dal campione.

Nella scelta dei campioni sono stati inseriti anche 10 bianchi di estrazione e di filtrazione, 6 e 4 rispettivamente nel primo e nel secondo sequenziamento. Seguendo le indicazioni riportate in Taberlet e colleghi (2018), sono stati inoltre inseriti nella piastra 12 bianchi di sequenziamento, nei quali sono presenti solo AmpliTaq Gold™ 360 Master Mix 1X (Applied Biosystems) e acqua. In altri pozzetti sono stati invece inseriti soltanto i reagenti, senza il template: questi rappresentano i controlli negativi di PCR. In Fig. 7 si riporta lo schema di una delle piastre di amplificazione effettuate.



**Fig. 7** - Schema della piastra del primo sequenziamento. In grigio, i bianchi di sequenziamento; in blu i bianchi di estrazione e filtrazione; in rosso i controlli negativi di PCR; in verde i controlli positivi. In bianco, i campioni di estratto da filtro il relativo codice identificativo.

Il profilo termico utilizzato per l'amplificazione è il seguente:

10' 95°C	} 35 cicli
30" 95°C	
1' 50°C	
30" 72°C	
7' 72°C	

L'amplificazione è stata svolta nello stesso ambiente pre-PCR delle estrazioni, utilizzando gli stessi accorgimenti di pulizia prima e dopo l'esperimento.

Il termociclatore utilizzato per tutte le prove, comprese quelle relative al blocking primer, e per la preparazione della libreria di sequenziamento, è il SimpliAmp™ (Applied Biosystems).

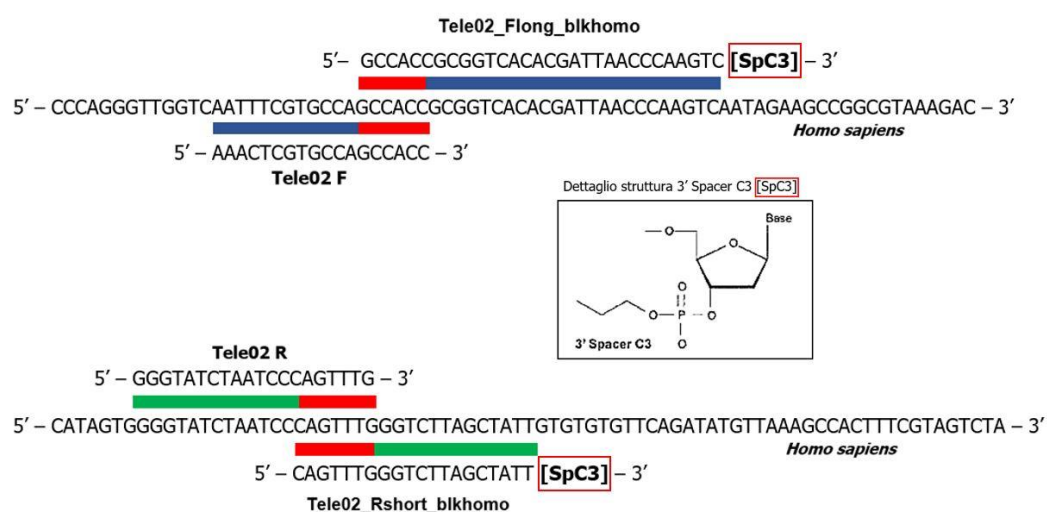
Al fine di valutare l'avvenuta amplificazione del DNA degli estratti, è stata svolta una analisi mediante elettroforesi su gel d'agarosio all'1,8% in TBE 1X. Al gel è stato aggiunto l'intercalante GelRed (Biotium). Sono stati utilizzati per l'analisi 3 µl di amplificato e 2 µl di loading Dye.

Il gel, una volta pronto, è stato visualizzato attraverso un transilluminatore (Gel Doc™ XR+, Bio-Rad).

### 3.3.1. Blocking primer

Sulla base delle informazioni riportate nel paragrafo 1.3.4, per una parte dei campioni è stata testata l'aggiunta di *blocking primer* (Vestheim & Jarman 2008) alla miscela di reazione, al fine di ridurre l'amplificazione di contaminanti in fase di PCR. In particolare, ci si è concentrati sulla costruzione e la messa a punto di un *blocking primer* contro la specie umana che, dal sequenziamento precedente, è risultata come contaminante ad elevata frequenza.

Sono state effettuate numerose prove con due blocking primer diversi disegnati, mediante l'uso del software *oligo explorer*, sul DNA umano (*Homo sapiens*), come riportato in Fig. 8.



**Fig. 8** - Dettaglio dei due blocking primer disegnati. Un blocking primer lungo (30 basi) con una sovrapposizione di 6 basi sulla sequenza del primer forward Tele02. Il secondo più corto (20 basi) con una sovrapposizione di 7 basi con il primer revers Tele02. In dettaglio, la modifica apportata 3' Spacer C3.

I *blocking primer* si appaiano perfettamente alla sequenza umana presente a valle e nella porzione 3' del sito di *annealing* del primer Tele02, competendo con esso durante l'amplificazione, ma presentano in posizione 3' una base modificata con 3' *Spacer C3* (dettaglio della Fig. 8), modifica in grado di inibire l'allungamento senza influire sulla temperatura di *annealing*.

Le prove per l'ottimizzazione del *blocking primer* sono state effettuate nelle condizioni il più possibile simili a quelle previste per il protocollo di amplificazione dei campioni con il solo primer Tele02. Sono state quindi effettuate mediante tecniche di amplificazione (PCR) in un volume di 10 µl. Per l'amplificazione delle sequenze di teleosteo sono stati usati come primer principale, sia forward che reverse, due primer Tele02 muniti di tag. Nella miscela di PCR sono stati inoltre inseriti due tipi di controllo positivo: alcuni campioni di teleosteo e alcuni campioni di estratto da filtri. Per verificare l'effetto bloccante del primer, è stato inserito DNA umano partendo da una concentrazione di 218 ng/µl, che è stato diluito diverse volte, fino a 1:1000; si ringrazia il Professor Gianluca Occhi (Dipartimento di Biologia - Unipd) per aver fornito l'estratto di DNA.

Sono state svolte numerose prove mediante PCR per ottimizzare la concentrazione del *blocking primer* rispetto al primer Tele02 e il profilo di amplificazione che permettesse a entrambi la resa migliore; si riportano di seguito le condizioni testate di aliquote dei diversi primer, aliquote di PCR, profilo di amplificazione ottimizzati per l'utilizzo del *blocking primer*.

L'amplificazione mediante PCR per testare l'efficienza dei blocking primer è avvenuta in 10µl, concentrazione dei reagenti: AmpliTaq Gold™ 360 Master Mix 1X (Applied Biosystems), Tele02 F 0.5 µM, Tele02 R 0.5 µM. La concentrazione finale del *blocking primer* è stata testata a 2µM, 5µM e 8µM. La temperatura di annealing del profilo di amplificazione è stata testata a 50°C, 51°C, 52°C, 53°C.

I risultati della PCR sono stati successivamente confrontati tra di loro per verificare la mancata amplificazione del DNA umano di controllo mediante elettroforesi su gel d'agarosio all'1,8% in TBE1X. Al gel è stato aggiunto l'intercalante GelRed (Biotium). Sono stati utilizzati per l'analisi 2 µl di amplificato e 2.5 µl di loading Dye. Il gel, una volta pronto, è stato visualizzato attraverso un transilluminatore (Gel Doc™ XR+, Bio-Rad). Insieme ai campioni è stato utilizzato un marcatore di peso molecolare ExcelBand™ 100 bp+3K (Smobio®).

### 3.4. Preparazione delle librerie e sequenziamento

Gli ampliconi ottenuti dai diversi filtri di una piastra sono stati uniti in una sola provetta e purificati prima di essere inviati al servizio esterno di sequenziamento.

Lo step di purificazione è stato effettuato per eliminare dalla soluzione contenente il frammento amplificato altre molecole, che potrebbero interferire con il processo di sequenziamento, come residui di AmpliTaq Gold™ 360 Master Mix 1X (Applied Biosystems) e vari buffer di amplificazione. Per la purificazione è stato utilizzato il kit *MinElute PCR Purification Kit* (Qiagen) ed è stata svolta seguendo il protocollo del fornitore. Il prodotto di purificazione è stato visualizzato con un gel all'1,8% attraverso un transilluminatore (Gel Doc™ XR+, Bio-Rad), per verificare l'avvenuta purificazione.

I purificati sono stati poi quantificati, sia tramite Nanodrop 2000 (ThermoFisher), sia mediante Qubit™ 4 Fluorometer (ThermoFisher).

Gli ampliconi uniti e purificati sono stati inviati a *BMR Genomics S.R.L* (Padova) per la preparazione delle librerie Illumina per la cura delle estremità e la ligazione degli adattatori con il kit NEBNext® Ultra™ II (NEB), e il successivo sequenziamento.

Le librerie delle due piastre sono state sequenziate con 2 diverse corse Illumina Miseq 2x150 bp Paired-End. Il sequenziamento ha restituito come risultato una media di 12.5 milioni di sequenze in un file grezzo in formato FastQ, utilizzato per le successive analisi informatiche. Le analisi hanno anche incluso i risultati di un sequenziamento precedente 2x150 bp Paired-End, effettuato con piattaforma Illumina Hiseq.

### 3.5. Analisi bioinformatiche

Il pacchetto utilizzato per l'analisi informatica del prodotto del sequenziamento è ObiTools3 Version 3.0.1b141, specifico per il DNA *metabarcoding*, che permette, con diversi passaggi, di filtrare e ordinare le sequenze grezze prodotte dal sequenziamento e di fornire una assegnazione tassonomica alle stesse sulla base di un database di riferimento.

Il primo passaggio prevede di effettuare l'import (- *obi import*) dei dati grezzi forniti dal sequenziamento in formato FastQ. Obitools3 acquisisce le sequenze creando contestualmente un DMS (Sistema di Gestione Documentale). Una volta importate le sequenze grezze si effettua l'allineamento delle *reads paired ended* ottenute dal sequenziamento dello stesso filamento (Forward e Reverse) con il comando - *obi align*. Questo

comando permette anche la filtrazione qualitativa delle sequenze sulla base dell'allineamento.

Successivamente, il software richiede l'importazione di un file nel quale ci sia una corrispondenza, per ogni campione, dei primer utilizzati (Tele02 F e R) e dei tag associati ad essi, che forniscono una combinazione specifica per ogni campione (filtro). In questo modo viene fornito un file di output nel quale ogni sequenza, pulita ed eventualmente assegnata tassonomicamente, può essere ricondotta al filtro nel quale è stata trovata.

Con il comando - *obi ngsfilter* viene quindi associata ogni sequenza prodotta dall'allineamento alla sua specifica combinazione di tag.

Successivamente, mediante il comando -*obi uniq -m sample*, le sequenze vengono filtrate tenendo una sola copia di ogni sequenza; se questa è presente più volte nel dataset, il numero di volte viene evidenziato mediante il parametro *count*, misura dell'abbondanza di quella sequenza. Questo permette di ottenere un dataset semplificato, composto da sequenze uniche. Un altro parametro, *samplecount*, permette di identificare in quanti campioni è stata identificata quella specifica sequenza.

A questo punto, si effettua un ulteriore filtraggio qualitativo sui dati, nello specifico riguardo ai valori di *seq\_lenght* e *count*, rispettivamente impostati su minimo 80bp di lunghezza e minimo 10 sequenze in *count*, al di sotto dei quali le sequenze vengono scartate e non procedono nell'analisi.

L'ultimo passaggio di pulizia delle sequenze viene effettuato mediante il comando - *obiclean*, che permette di eliminare dall'analisi le varianti di una sequenza presenti al di sotto di una percentuale stabilita di campioni (tipicamente il 5%), in quanto considerate possibili errori di sequenziamento e fonte di rumore di fondo.

L'associazione tassonomica viene effettuata grazie all'utilizzo di un database appositamente scaricato da EMBL (*European Molecular Biology Laboratory*). Nello specifico, in data 20 maggio 2021 è stato scaricato un database che include le sequenze standard di tutti i gruppi tassonomici presenti in EMBL ([www.ebi.ac.uk/](http://www.ebi.ac.uk/)).

Una volta importato tale database nell'ambiente di OBI, viene svolta una PCR *in silico* mediante il comando -*obi ecopcr*, indicando le sequenze dei primer utilizzati nelle amplificazioni in vitro, al fine di estrarre esclusivamente le possibili sequenze omologhe al frammento amplificato con associazione tassonomica nota, riscontrabili nei campioni in analisi. Il comando taglia le sequenze ed elimina da queste il sito dei primer, conservando soltanto la regione amplificata compresa tra i due primer. Successivamente, si confrontano le sequenze ottenute sperimentalmente con quelle presenti nel database di riferimento; in presenza di una

percentuale d'identità del 97%, si ottiene l'attribuzione tassonomica delle sequenze in analisi.

L'assegnazione tassonomica viene effettuata tramite il comando - *ecotag*, a partire dal file delle sequenze presenti nel database, incrociato al file *my\_tax* (che comprende la nomenclatura tassonomica da associare alle sequenze) e alle sequenze “pulite”, ovvero quelle uniche risultanti dalla fase “di scrematura” iniziale delle sequenze allineate. La nomenclatura tassonomica è quella standard, basata su NCBI (*National Center for Biotechnology Information*), che in alcuni casi differisce dalla nomenclatura corrente.

Quello che si ottiene alla fine come output dal pacchetto di *software* ObiTools3 è un file in formato tab; il file è una matrice, dalla quale si possono osservare i nomi identificativi dei campioni con le singole sequenze identificate (e non) e relativa associazione tassonomica. Per ogni assegnazione della sequenza ad un determinato campione, è associata una quantità in termini di *count*, che corrisponde al numero di volte che la specifica sequenza è stata identificata in quello specifico campione.

È così possibile svolgere una analisi qualitativa di quali siano le specie trovate in un singolo campione, ma anche quantitativa, ovvero è possibile, sulla base del valore *count*, stabilire quale sia il numero di sequenze amplificate riconducibili a quello specifico campione.

Durante l'analisi informatica sono stati analizzati i risultati di tre diversi sequenziamenti. Il primo era un sequenziamento effettuato in precedenza (Cananzi 2019), Illumina HiSeq, per il quale si è presentata la necessità di utilizzare il server Capri ad alte prestazioni, fornito dal Dei dell'Università di Padova. Grazie a questo server è stato possibile processare dati molto grandi in tempi ridotti. Per gli altri due sequenziamenti MiSeq non si è reso necessario l'utilizzo di tale server, in quanto i risultati forniti dal sequenziamento hanno un peso minore.

Per i singoli comandi specifici effettuati in ObiTools3, si rimanda all'appendice II.

### 3.6. Analisi statistiche ed ecologiche

In questo lavoro di tesi sono stati analizzati con sistemi bioinformatici i risultati di tre diversi sequenziamenti: uno svolto con tecnologia Illumina HiSeq, gli altri due con la tecnologia Illumina MiSeq.

Una volta ottenuto il file in formato tab dal software ObiTools3, questo è stato analizzato mediante Microsoft Office Excel e R per ottenere come risultato la lista delle OTUs (*Operative Taxonomic Units*) assegnate per



ogni campione. A partire da questa lista è stato fatto un controllo e una filtrazione qualitativa dell'assegnazione alle sequenze.

In particolare, la lista di specie assegnata alle sequenze è stata confrontata con la letteratura e raffinata grazie al parere esperto della Professoressa Carlotta Mazzoldi (Dipartimento di Biologia - Unipd), che si ringrazia. In diversi casi dubbi, nei quali una sequenza era stata assegnata ad una specie non nota nella laguna di Venezia, si è proceduto con un controllo più approfondito attraverso BLAST (*Basic Local Alignment Search Tool*), che ha permesso di sostituire una specie con un'altra la cui sequenza aveva una percentuale di identità maggiore, o ricondurre a livello di specie una sequenza che era stata assegnata a livello di genere. Si è ricavata in questo modo la lista definitiva di specie.

Successivamente, si è provveduto a correggere i dati ottenuti per ogni specie ed ogni campione dal segnale legato a possibili contaminazioni o artefatti. A tal fine, dal valore *count* delle sequenze, che identifica quante sequenze assegnate ad una specie vengono ricondotte ad uno specifico campione, è stato sottratto il segnale ottenuto per quella stessa specie nei diversi controlli (bianchi di estrazione, filtrazione, bianchi di sequenziamento e controlli negativi di PCR); nello specifico, si è deciso di rimuovere da ogni campione un numero di sequenze pari al terzo quartile del *count* di quella stessa specie presente nei bianchi ottenuti con la stessa tipologia di sequenziamento (HiSeq o MiSeq). In questo modo si è ottenuta una filtrazione rispetto ai contaminanti presenti, che ha riportato come dato una matrice di abbondanza assoluta, nella quale per ogni specie viene riportato il valore di *count* privo delle sequenze di quella specie trovate nei bianchi.

Successivamente, i dati sono stati approssimativamente normalizzati al fine di rendere possibile il confronto tra il sequenziamento HiSeq e i due sequenziamenti MiSeq analizzati. In particolare, si è tenuto conto del diverso *throughput* dei due tipi di sequenziamento, e i dati sono stati corretti tenendo conto del rapporto tra il numero delle sequenze ottenute mediante sequenziamento HiSeq e la media dei due sequenziamenti MiSeq.

A partire dalla matrice, filtrata e normalizzata in modo preliminare per rendere confrontabili i due sequenziamenti, sono stati creati due file necessari alle successive analisi: una matrice che presenta le specie sull'asse y e i campioni sull'asse x e al suo interno il *count* della specie assegnato allo specifico campione; un secondo file che contiene i metadati, ovvero è un file nel quale sono state inserite le informazioni riguardanti i campioni: sito e data di campionamento, tipologia filtro ed ogni altra informazione utile.

I due file descritti sono stati importati nell'ambiente R e analizzati mediante pacchetti specifici per le analisi ecologiche (alcuni esempi sono *vegan*,

*phyloseq*, *microbiome*) al fine di analizzare statisticamente i dati e creare dei grafici esplicativi.

Nell'ambiente R sono state effettuate delle analisi multivariate per verificare che i dati avessero delle significative differenze tra di loro e che si potessero considerare, come variabili esplicative, anche la variazione dell'abbondanza della specie in base alla stagione e al sito di campionamento, al di là del dato di presenza/assenza.

A partire da una matrice di abbondanza relativa, creata in R, delle sequenze in ogni campione, è stata svolta un'analisi statistica mediante *Adonis test* (Anderson, 2001), ovvero una analisi permutazionale multivariata della varianza che utilizza una matrice di distanze euclidee tra abbondanze relative dei campioni.

L'Adonis test restituisce un *P-value*, considerato significativo sotto lo 0.05, che descrive la probabilità di osservare la distribuzione della varianza nel dataset per solo effetto del caso. Il test restituisce inoltre un valore di  $R^2$ , che indica, per ogni variabile, la percentuale di variabilità spiegata dalla stessa nell'intero dataset. In base al risultato dell'Adonis test sono state effettuate due diverse analisi delle componenti principali (PCA) al fine di visualizzare i dati in grafico, in particolare per valutare la variazione dei campioni.

Lo stesso test è stato svolto anche, a parte, per il sequenziamento in cui i campioni sono stati sequenziati in replica, ma con l'aggiunta del *blocking primer*, al fine di stimare statisticamente la variazione apportata dall'aggiunta del primer. Mediante il pacchetto R *pheatmap* è stata anche creata una *heatmap*, che rappresenta un modo per visualizzare facilmente la differenza di distribuzione dell'abbondanza delle sequenze, per ogni specie, in ogni campione. Al variare dell'intensità di colore infatti è possibile apprezzare come una sequenza attribuita ad una specie sia più o meno abbondante in un determinato campione. Questa è stata creata a partire da una matrice di abbondanza relativa convertita in scala logaritmica (matrice argomento del  $\log_{10}$ ) ed è stata utilizzata per vedere graficamente la differenza tra i campioni senza l'utilizzo del *blocking primer* e con l'utilizzo del *blocking primer*.

## 4. Risultati

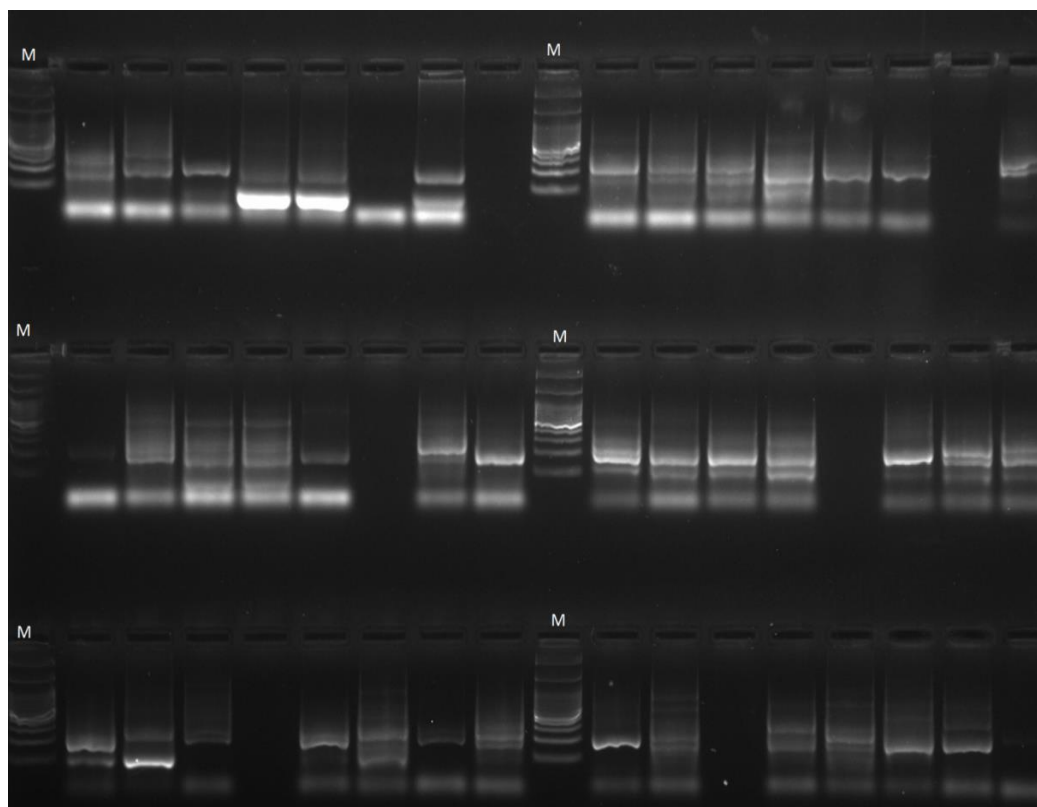
### 4.1. Estrazione del DNA da filtro e prove di amplificazione

Tutti i campioni appartenenti al primo ciclo di campionamento hanno restituito quantificazioni soddisfacenti, da un minimo di 11.9 ng/μl ad un

massimo di 74.9 ng/μl. Le estrazioni da bianchi di filtrazione e di estrazione hanno fornito invece quantificazioni massime di 4 ng/μl.

Le quantificazioni delle estrazioni da filtro del secondo ciclo di campionamento hanno restituito valori mediamente minori rispetto alle precedenti, con notevoli variazioni in base alla tipologia di filtro. In particolare, i filtri nei quali la quantità di DNA risulta maggiore sono i filtri con maglia più stringente (0.22 μm), mentre la quantificazione diminuisce all'allargarsi della maglia.

Uno dei gel dell'amplificazione mediante PCR è riportato in Fig.9; come si può osservare, il frammento 12S rDNA è stato amplificato con successo mediante i primer Tele02; infatti, tutti i pozzetti contenenti i campioni di filtro hanno restituito diverse bande, corrispondenti alle diverse sequenze amplificate. I pozzetti senza banda corrispondono invece a bianchi di sequenziamento o a controlli negativi di PCR, nei quali l'amplificazione di DNA non è avvenuta, o è avvenuta in modo troppo debole per restituire delle bande visibili.



**Fig. 9** - Esempio di gel di amplificato per la verifica dell'avvenuta amplificazione. I pozzetti privi di banda corrispondono a bianchi di sequenziamento o a controlli negativi di PCR. I pozzetti esterni e centrali (contrassegnati con la lettera "M"), che presentano molte bande di diverse grandezze, presentano al loro interno il marcatore di peso molecolare ExcelBand™ 100 bp+3K (Smobio®).

## 4.2. Sequenziamento

I due sequenziamenti MiSeq hanno restituito rispettivamente circa 12 e 13 milioni di sequenze grezze. Dopo l'allineamento e l'identificazione bioinformatica delle sequenze che presentano il primer con il relativo tag, sono state ottenute circa 3 milioni di sequenze dal primo sequenziamento e 4 dal secondo. Di queste, il 59-69% sono state assegnate tassonomicamente con una percentuale di identità minima del 97%.

Il sequenziamento HiSeq, effettuato in precedenza rispetto a questo lavoro di tesi, ha restituito 240 milioni di sequenze. Di queste, una volta applicati i diversi filtri bioinformatici, sono rimaste circa 15 milioni sequenze identificate. Delle sequenze identificate, il 95% sono state assegnate tassonomicamente con una percentuale di identità minima del 97%. Per i dettagli numerici, riporto alla tabella 2 sottostante nella quale si esplicitano i diversi sequenziamenti, le sequenze restituite nel totale, quelle che sono state identificate e quelle assegnate tassonomicamente.

**Tabella 3** - Conteggio delle sequenze processate in questo lavoro di tesi.

Sequenziamento	Totale	Identificate	Assegnate
HiSeq	240.000.000	14.818.575	14.007.216
MiSeq1	12.000.000	2.922.705	2.017.441
MiSeq2	13.000.000	3.968.416	2.345.358

## 4.3. Identificazione di OTU

Nonostante il primer fosse classificato come specifico per l'infraclassa dei teleostei, l'assegnazione tassonomica ha restituito 105 diverse OTU (*Operational Taxonomic Units*), comprendenti anche specie non appartenenti all'infraclassa dei teleostei (es. uccelli e mammiferi) che sono state escluse dalle successive analisi (ad eccezione del delfino, *Tursiops truncatus*).

Oltre a queste, state escluse dalle analisi anche le sequenze assegnate a OTU che vengono considerate, con tutta probabilità, frutto di una contaminazione di laboratorio. Nello specifico, sono state escluse le reads di *Chionodraco* (icefish antartico) e *Huso huso*, (storione beluga) entrambi teleostei trattati nei laboratori di ricerca condivisi e con alta probabilità non riconducibili alla laguna di Venezia. In due casi si è presentato il problema legato al fatto che le specie (*Acipenser naccarii* e *Carcharhinus plumbeus*), seppur molto rare, potrebbero essere state rinvenute in laguna; essendo

però specie che vengono anche trattate nei laboratori condivisi, al fine di aumentare l'affidabilità della lista di specie sono state escluse, in modo conservativo, dall'analisi successiva, classificandole come contaminante di laboratorio.

Dalla lista rimanente, è stato effettuato un confronto con le informazioni presenti in letteratura e una consulenza con la Professoressa Carlotta Mazzoldi (Dipartimento di Biologia – Unipd), esperta di fauna ittica in laguna di Venezia, e in casi dubbi, il controllo incrociato mediante BLAST. Questo confronto ha permesso di attribuire la specie a diversi casi in cui ObiTools3 aveva assegnato la sequenza solo a livello di genere. In altri casi, svolgendo una ricerca su BLAST delle sequenze ricondotte a specie non presenti in laguna di Venezia, ma filogeneticamente simili ad alcune presenti invece in laguna, è stata assegnata un'altra specie alla sequenza dubbia.

Dopo la correzione di queste imprecisioni da parte delle analisi informatiche, sono state confermate 64 pesci potenzialmente presenti in laguna, dei quali 58 sono stati identificati a livello di specie.

In tabella 4 è presente la legenda con la sigla della *guild* ecologica (Cavraro *et al.* 2017), e le sigle delle possibili aree di provenienza. In tabella 5 viene riportata la lista di specie con la *guild* ecologica, quando conosciuta, e l'area nella quale ci si aspetta il rilevamento della specie.

**Tabella 4** - Legenda con sigla e spiegazione estesa.

Sigla	Guild	Sigla2	Area
RL	Residenti in laguna	I	Interno della laguna
MM	Migratori marini	L	Lagunare
MO	Migratori occasionali	E	Esterno della laguna

**Tabella 5** - In tabella, la lista di specie individuate mediante analisi di eDNA in questo lavoro di tesi e il count assoluto delle sequenze ottenute dai sequenziamenti dopo la preliminare normalizzazione per HiSeq e MiSeq. A fianco, l'assegnazione alla *guild* ecologica (Cavraro *et al.* 2017) e l'area di provenienza della specie.

OTU	COUNT	Guild	Area
<i>Abramis brama</i>	1291		I
<i>Alosa fallax</i>	5547		L
<i>Anguilla anguilla</i>	27585		L
<i>Aphia minuta</i>	1323		E
<i>Arnoglossus laterna</i>	5044	MM	E
<i>Atherina boyeri</i>	1351239	RL	L

<i>Barbus barbus</i>	141		I
<i>Carassius auratus</i>	24135	MM	I
<i>Chelon auratus</i>	218278	MM	L
<i>Chelon labrosus</i>	2152548	MM	L
<i>Chelon saliens</i>	185688	MM	L
<i>Conger conger</i>	1630	MO	E
<i>Cottus gobio</i>	1118		I
<i>Cyprinus carpio</i>	32881		I
<i>Ctenopharyngodon idella</i>	27645		I
<i>Dicentrarchus labrax</i>	327188	MM	L
<i>Engraulis encrasicolus</i>	41719	MM	E
<i>Esox flaviae</i>	7459	D	I
<i>Esox lucius</i>	2681	D	I
<i>Gadus sp.</i>	35918		E
<i>Gambusia affinis</i>	39	D	I
<i>Gambusia holbrooki</i>	13413	D	I
<i>Gobius cobitis</i>	29884	MO	I
<i>Gobius niger</i>	1701152	RL	L
<i>Gobius paganellus</i>	248979	RL	E
<i>Ictalurus sp.</i>	9608		I
<i>Lagocephalus lagocephalus</i>	991	invasiva	E
<i>Leuciscus leuciscus</i>	62889		I
<i>Merluccius merluccius</i>	70084		E
<i>Micromesistius poutassou</i>	2138		E
<i>Mugil cephalus</i>	13748	MM	L
<i>Mustelus mustelus</i>	2675		E
<i>Myliobatis aquila</i>	1139		E
<i>Oncorhynchus mykiss</i>	34238	introdotta	I
<i>Padogobius martensii</i>	2761		I
<i>Pagellus erythrinus</i>	195	MO	L
<i>Phoxinus phoxinus tumensis</i>	2688		I
<i>Poecilia reticulata</i>	6089	introdotta	I
<i>Pomatomus saltatrix</i>	1959		E
<i>Pomatoschistus microps</i>	129		L
<i>Pomatoschistus minutus</i>	10399	ER	L
<i>Pseudorasbora parva</i>	24498	D	I
<i>Raja sp.</i>	34490		E
<i>Rhodeus sericeus</i>	58143	D	I
<i>Rutilus rutilus</i>	7427	introdotta	I
<i>Salmo salar</i>	4472		L
<i>Sander luciperca</i>	1601	introdotta	I
<i>Sardina pilchardus</i>	111991		E

<i>Sardinella aurita</i>	1264		L
<i>Scomber scombrus</i>	3738	MO	E
<i>Scophthalmus maximus</i>	1717	MO	L
<i>Silurus sp.</i>	6873		I
<i>Solea solea</i>	86215	MM	L
<i>Sparidae sp.</i>	172739	MO	L
<i>Sprattus sprattus</i>	2609	MM	L
<i>Squalius cephalus</i>	22224		I
<i>Symphodus cinereus</i>	2842	MO	E
<i>Syngnathus typhle</i>	80566	RL	L
<i>Trachinus draco</i>	48031		E
<i>Trachurus mediterraneus</i>	32665		E
<i>Trachurus trachurus</i>	9706		E
<i>Triglidae sp.</i>	6232		L
<i>Tursiops truncatus</i>	515		E
<i>Zosterisessor ophiocephalus</i>	2278357	RL	L

Nell'esame della lista di specie totale va considerato come la laguna di Venezia sia un ambiente di transizione; come atteso infatti, la lista comprende sia pesci d'acqua salata che d'acqua dolce.

Il 70% di queste specie appartengono a pesci tipicamente lagunari o marini, presenti occasionalmente in laguna. Il restante 30% sono pesci di acqua dolce; al di là del dato presenza di queste specie, la stima dell'abbondanza di queste specie in laguna è da osservare con occhio critico. A causa infatti di uno dei limiti intrinseci dell'eDNA, non possiamo stabilire con certezza se il segnale di abbondanza sia da ricondursi alla numerosità degli individui presenti in laguna, o a eDNA della specie dulciacquicola che viene trasportato in laguna dagli affluenti.

Le specie maggiormente rappresentate, sia in termini di *count*, sia in termini di distribuzione all'interno dei campioni, sono, come atteso, specie residenti lagunari. Tra queste si nominano *Zosterisessor ophiocephalus* (Fig. 10), *Atherina boyeri* e *Gobius niger* (Fig. 11).





**Fig. 10** - *Zosterisessor ophiocephalus*. Fonte: [www.istitutoveneto.org](http://www.istitutoveneto.org).



**Fig. 11** - *Gobius niger*. Fonte: <https://spain.inaturalist.org/taxa/118704-Gobius-niger>.

La *guild* ecologica dei migratori marini, ovvero teleostei che migrano, in base al loro ciclo di riproduzione, dal mare alla laguna, è ampiamente rappresentata nei risultati. In particolare, molto abbondanti risultano le sequenze del genere *Chelon*, rappresentato dalle specie *Chelon auratus*, *Chelon labrosus* (Fig. 13) e *Chelon saliens*.



Molto abbondante in tutto il dataset anche *Dicentrarchus labrax* (Branzino, Fig. 14) e *Solea solea*, anch'essi migratori marini.

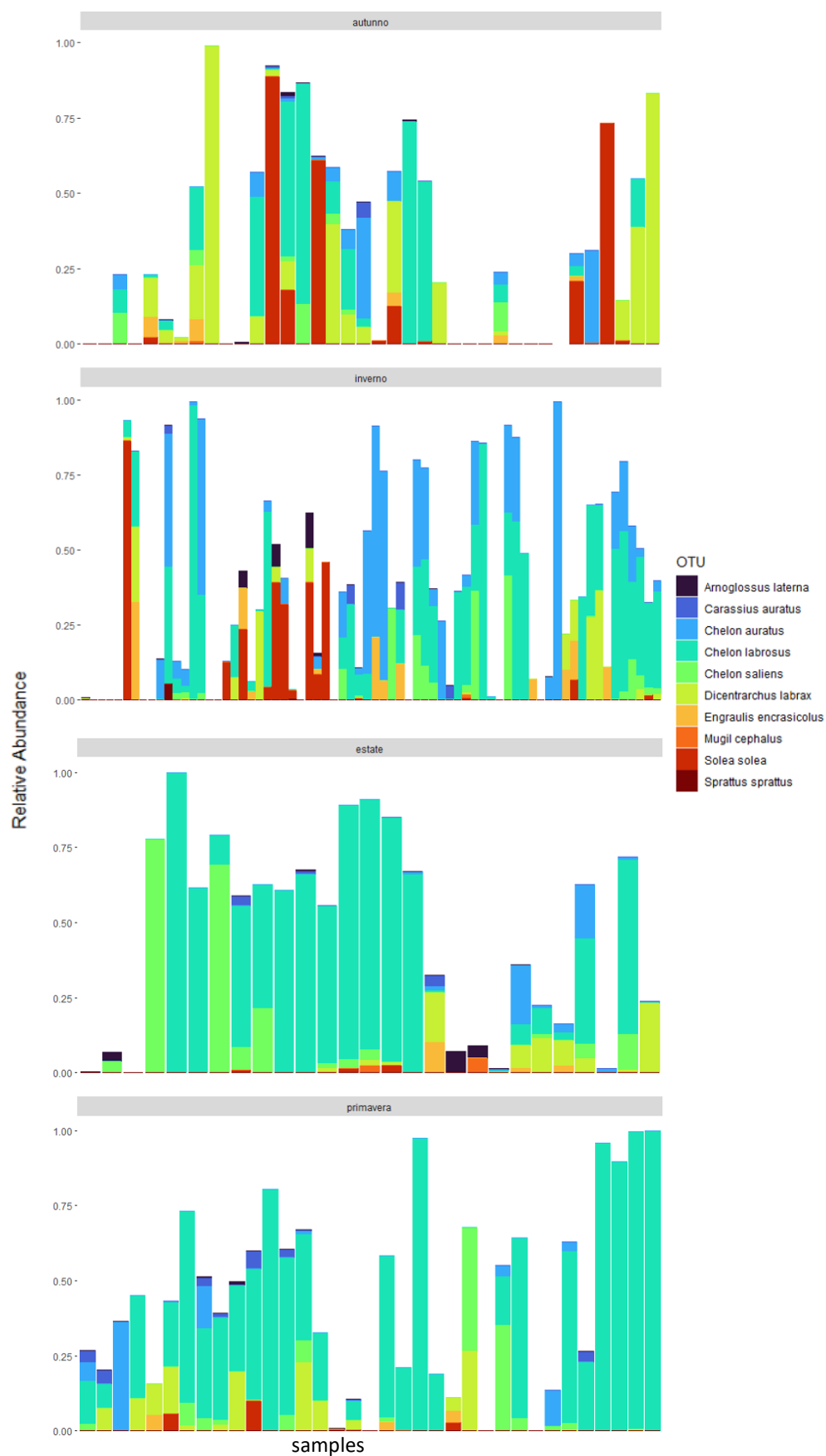


**Fig. 13** - *Chelon labrosus*. Fonte: <https://www.snorkeling-report.com/species/chelon-labrosus>.



**Fig. 13** - *Dicentrarchus labrax*. Fonte: [https://en.wikipedia.org/wiki/European\\_bass](https://en.wikipedia.org/wiki/European_bass).

Al fine di osservare la variazione di abbondanza relativa di sequenze assegnate alle specie classificate come migratori marini, è stata svolta una analisi sulla base di queste sole specie, al fine di individuare una variazione significativa della loro presenza sulla base della stagione di campionamento. In Fig. 14 è possibile apprezzare la diversa abbondanza relativa delle specie di migratori marini al cambiamento nelle diverse stagioni.



**Fig. 14** - Rappresentazione dell'abbondanza relativa delle specie di migratori marini nelle quattro stagioni.

Dal grafico si può osservare che l'abbondanza di alcune specie rispetto ad altre cambia al variare della stagione. In alcuni casi specifici, la distribuzione della specie risulta coerente con l'atteso. Si può osservare, ad esempio, che in estate e in primavera la specie maggiormente rappresentata è *Chelon labrosus*. Questo dato è concorde con la distribuzione attesa della specie, che è solita migrare dal mare alle acque interne in primavera (aprile-maggio), quando sono maggiormente presenti le prede (Quartarella 2013), e a nord in estate, con l'innalzarsi delle temperature dell'acqua ([www.fishbase.se/summary/Chelon-labrosus](http://www.fishbase.se/summary/Chelon-labrosus)).

*Solea solea* è maggiormente rappresentata in autunno e in inverno, quando inizia la sua stagione riproduttiva ([www.fishbase.se/summary/Solea-solea](http://www.fishbase.se/summary/Solea-solea)). Allo stesso modo, *Dicentrarchus labrax* risulta abbondante in autunno, momento in cui la specie si prepara ad affrontare la stagione riproduttiva ([www.fishbase.se/summary/dicentrarchus-labrax](http://www.fishbase.se/summary/dicentrarchus-labrax)).

Dalle analisi risulta la presenza del pesce palla (*Lagocephalus lagocephalus*) (Fig. 15), specie altamente invasiva, in due siti e in diverse date di campionamento: in particolare, la specie è stata identificata in campioni di Marghera (febbraio) e Palude della Rosa (settembre e ottobre). L'areale di questa specie comprende tutte le coste tropicali dell'Africa, ed è stata documentata, anche se molto raramente, nel Mar Mediterraneo ([it.wikipedia.org/wiki/Lagocephalus\\_lagocephalus](http://it.wikipedia.org/wiki/Lagocephalus_lagocephalus)); tuttavia, la sua presenza non è mai stata registrata nel medio e alto Adriatico ([www.iucn.it/scheda.php?id=-2081245894](http://www.iucn.it/scheda.php?id=-2081245894)).

Nonostante sia inaspettata, l'individuazione di questa specie nei risultati non può essere giustificata da transito nel mercato ittico o da contaminante di laboratorio, quindi si è potuta considerare la presenza nel dataset del suo DNA come affidabile.



**Fig. 15** - Due esemplari di *Lagocephalus lagocephalus*. Fonte: <https://www.fishbase.se/photos/PicturesSummary.php?StartRow=2&ID=4292&what=species&TotRec=7>

Anche se il dato non è da considerare una prova della presenza *in loco* della specie, l'individuazione della specie invasiva a livello di eDNA può essere considerata come un segnale precoce di allerta, che può indicare la necessità di una sorveglianza attiva.

Indagando in modo critico tra le specie identificate, si è notata la presenza del pesce spada *Xiphias gladius*, del tonno (identificato a livello di genere, *Thunnus sp.*) e dello squalo mako (*Isurus oxyrinchus*), non attesi in laguna. Infatti, il pesce spada è atteso nell'oceano Indiano, Atlantico e Pacifico ([www.fishbase.se/summary/226](http://www.fishbase.se/summary/226)); il tonno è stato identificato solo a livello di genere, che comprende diverse specie tipicamente con una distribuzione molto ampia, ma principalmente in aree oceaniche e solo raramente nel Mar Adriatico; lo squalo mako, infine, tipico delle acque tropicali ([www.fishbase.de/summary/isurus-oxyrinchus.html](http://www.fishbase.de/summary/isurus-oxyrinchus.html)).

La caratteristica che accomuna queste tre specie nel dataset è la presenza di molte sequenze, ciascuna con un *count* molto basso; questo, insieme al fatto che sono specie non autoctone della laguna di Venezia, ha portato ad indagare sulla loro individuazione. Così, si è scoperto che nei mesi di campionamento sono stati registrati in transito nel mercato ittico di Chioggia; quindi, consideriamo queste specie come contaminante alimentare dovuto a scarti di lavorazione del mercato ittico.

#### 4.4. Analisi statistiche

Il risultato dell'Adonis test (riportato in Fig. 18) ha evidenziato che sia la variabile sito, sia la variabile stagione sono risultati significativi nello spiegare la variabilità dei risultati. C'è tuttavia una netta differenza tra i due, che si apprezza dalla differenza nel valore  $R^2$ . Se la variabile stagione, nonostante restituisca un *P-value* significativo (0.005), spiega solo il 6% della variabilità dei dati, il valore sito spiega invece il 18% della variabilità del dataset e risulta quindi estremamente significativo (*P-value* 0.001).

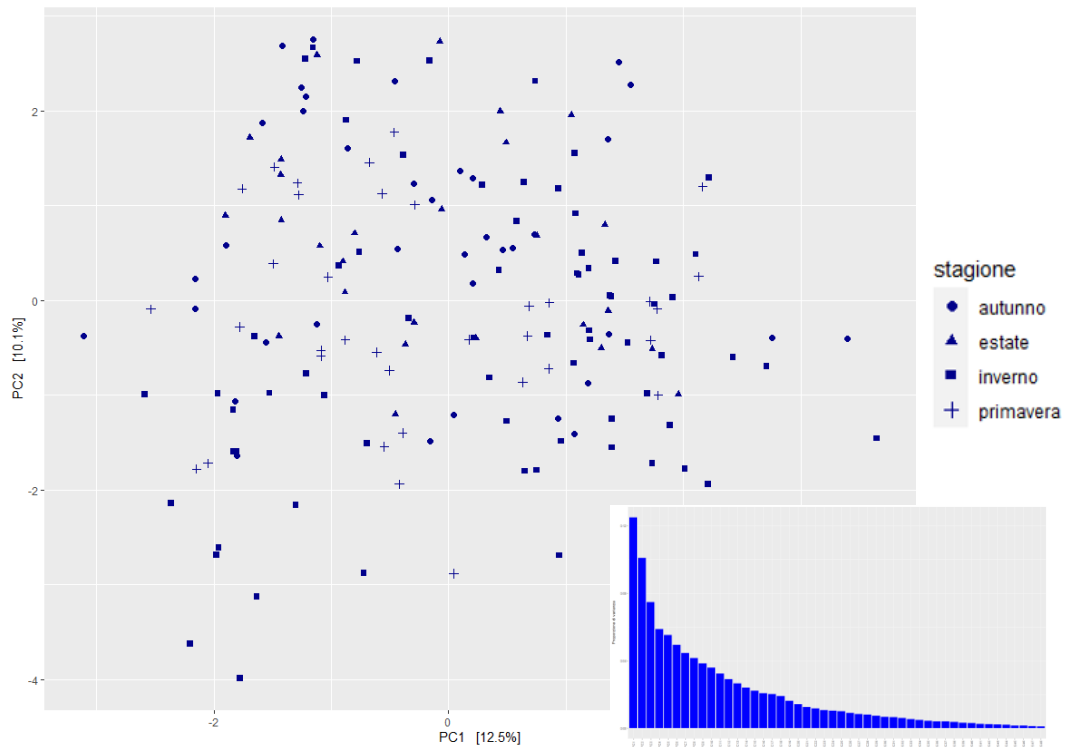
Per questo motivo, nelle analisi successive sono state considerate queste due variabili, in analisi descrittive, per trarre alcune conclusioni ecologiche preliminari.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	
\$sito	4	6814	1703.58	3.8603	0.18261	0.001	***
\$stagione	3	2259	753.10	1.7065	0.06054	0.005	**
Residuals	64	28244	441.31		0.75685		
Total	71	37317			1.00000		
---							
Signif. codes:	0	****	0.001	***	0.01	**	0.05
					.	0.1	' ' 1

**Figura 17** - Risultato dell'Adonis test (in R) effettuato sull'intero dataset.

#### 4.4.1 Analisi delle componenti principali

Alla luce dei precedenti risultati, e al fine di stimare la variazione dei campioni sulla base della variazione di stagione o sito, riproducendola graficamente, è stata svolta una analisi delle componenti principali (PCA). In particolare, sono state prese in considerazione la prima (asse x) e la seconda (asse y) componente principale (Fig. 18), che insieme spiegano il 22,6% della variabilità all'interno del dataset.



**Fig. 18** - Grafico a barre delle componenti principali del dataset.

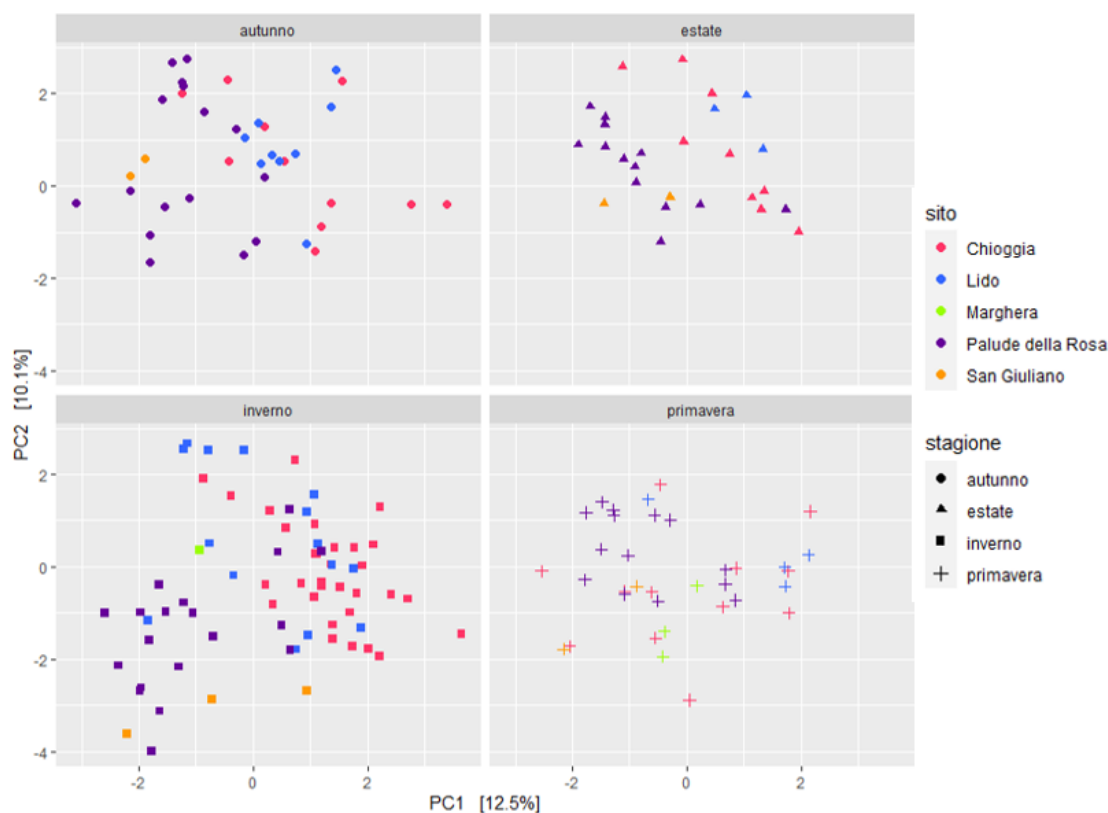
**Fig. 19** - Grafico delle componenti principali dell'intero dataset. Le stagioni sono identificabili mediante le diverse forme.

Da una prima analisi che considera assieme le diverse stagioni (Fig. 19), sembra che i campioni non abbiano una particolare distribuzione sulla base delle stagioni, e sembrano tutti simili tra loro.

Diversamente, se le stagioni vengono separate e i campioni vengono colorati in base al loro sito di campionamento (Fig. 20) si possono apprezzare delle differenze. Infatti, i campioni prelevati a Palude della Rosa (nord) e Chioggia (sud) sono sempre nettamente separati in autunno, inverno ed estate. Come atteso dalla prima osservazione, anche i campioni di Lido e San



Giuliano sono ben divisi tra loro, con San Giuliano vicino a Palude della Rosa (più a nord) e Lido più vicino a Chioggia (più a sud).



**Fig. 20** - Grafico di analisi delle componenti principali dell'intero dataset. All'interno di ogni stagione, i siti sono identificabili mediante i diversi colori.

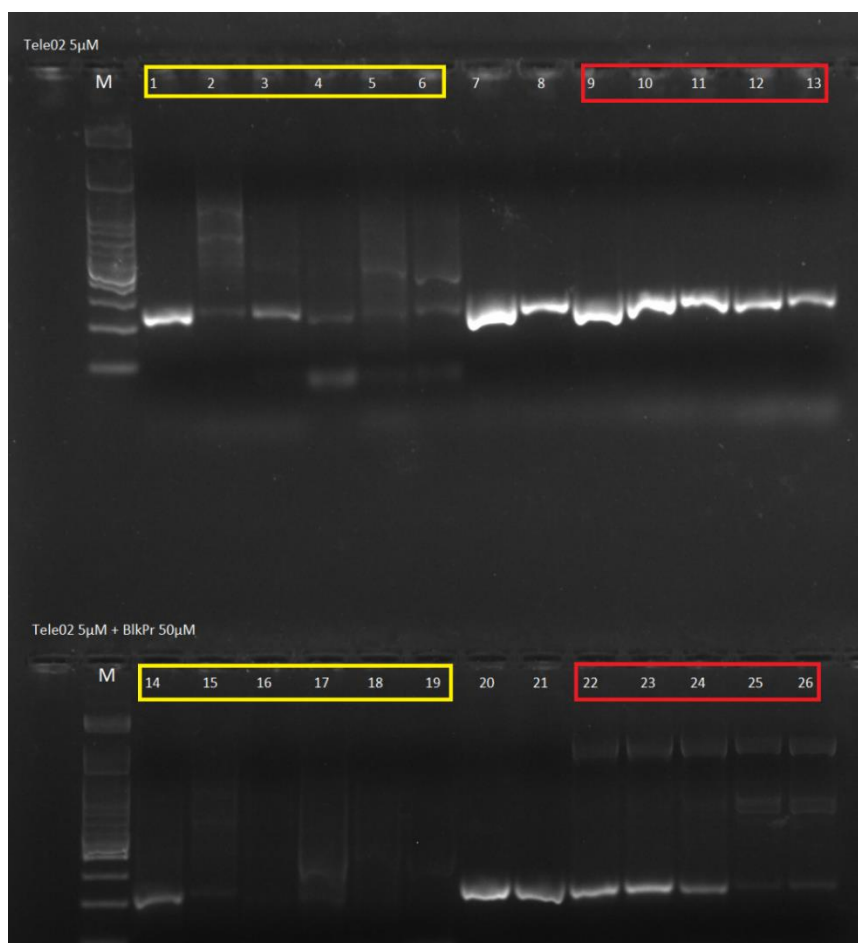
Il grafico evidenzia dunque che, sulla base delle componenti principali, in tre stagioni i campioni appartenenti allo stessa zona lagunare o a zone contigue sono meno variabili tra loro rispetto al totale del dataset. Più che il sito in sé, sembra quindi essere importante l'effetto Laguna Nord/Sud. Questo effetto non sembra esserci in primavera, periodo in cui i campioni sembrano omogenei in tutti i siti.

Per l'aiuto con le analisi statistiche e l'utilizzo di R, si ringrazia la Dott.ssa Alessia Buratin del gruppo di ricerca della Professoressa Stefania Bortoluzzi (Dipartimento di Biologia – Unipd).

#### 4.5. Blocking primer

Parte dei risultati delle prove di amplificazione con *blocking primer* sono visualizzati in Fig. 21, nella quale in rosso sono stati riquadrati i campioni di DNA umano, in concentrazioni scalari da 1:25 a 1:1000, mentre nei riquadri gialli sono stati inseriti dei campioni di filtro come controllo

positivo dell'avvenuta amplificazione. I campioni 7-8 e 10-21 nell'immagine corrispondono a controlli positivi rappresentati da DNA di teleosteo.



**Fig. 21** - Gel d'agarosio per verifica del funzionamento del blocking primer *Tele02\_Flong\_blkhomo*. Come si nota dall'immagine, la banda corrispondente ai pozzetti riquadrati di rosso (DNA umano) diminuisce visibilmente con l'aggiunta del blocking primer. I pozzetti di sinistra (contrassegnati con la lettera "M"), che presentano molte bande di diverse grandezze, presentano al loro interno il marcatore di peso molecolare *ExcelBand™ 100 bp+3K (Smobio<sup>R</sup>)*.

Il protocollo che ha restituito i risultati migliori è quello che prevede una concentrazione del blocking primer 10X rispetto a quella del primer universale *Tele02* e la temperatura di *annealing* del profilo di amplificazione di 50°C.

Dei due *blocking primer* testati, "*Tele02\_Flong\_blkhomo*" ha restituito i migliori risultati. È stato quindi inserito nel protocollo di amplificazione per il successivo sequenziamento.

Dalle analisi informatiche si è potuto osservare che l'utilizzo del *blocking primer* ha diminuito con successo il numero di sequenze identificate e attribuite ad *Homo sapiens* e *Homininae* presenti in ciascun campione, con poche eccezioni; nel totale dei campioni, si ha infatti una diminuzione del

12% delle sequenze di uomo rispetto ai campioni amplificati esclusivamente mediante Tele02. Si è osservata inoltre una diminuzione di altri contaminanti, appartenenti soprattutto agli ordini di mammiferi e uccelli. Questi risultati hanno permesso, in termini di numero di specie di teleostei identificate, di incrementare la lista di 3 specie; sono stati infatti identificati, soltanto nei campioni amplificati con il *blocking primer*, anche *Esox flaviae*, *Esox lucius* e *Rutilus rutilus*. Oltre a questo risultato, le sequenze attribuite ai pesci presentano un *count* mediamente doppio rispetto ai campioni senza *blocking primer*.

A partire da questa analisi preliminare è stato svolto un test statistico che ha evidenziato come la variabile utilizzo/non utilizzo del *blocking primer* abbia un effetto significativo sulla distribuzione dell'abbondanza relativa delle specie identificate. In Fig. 22 sono riportati i risultati del test statistico "Adonis test" effettuato mediante R.

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)	
\$Blocking	1	1806	1805.81	4.3032	0.04839	0.001	***
\$sito	4	6814	1703.58	4.0596	0.18261	0.001	***
\$stagione	3	2259	753.10	1.7946	0.06054	0.002	**
Residuals	63	26438	419.65		0.70846		
Total	71	37317			1.00000		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

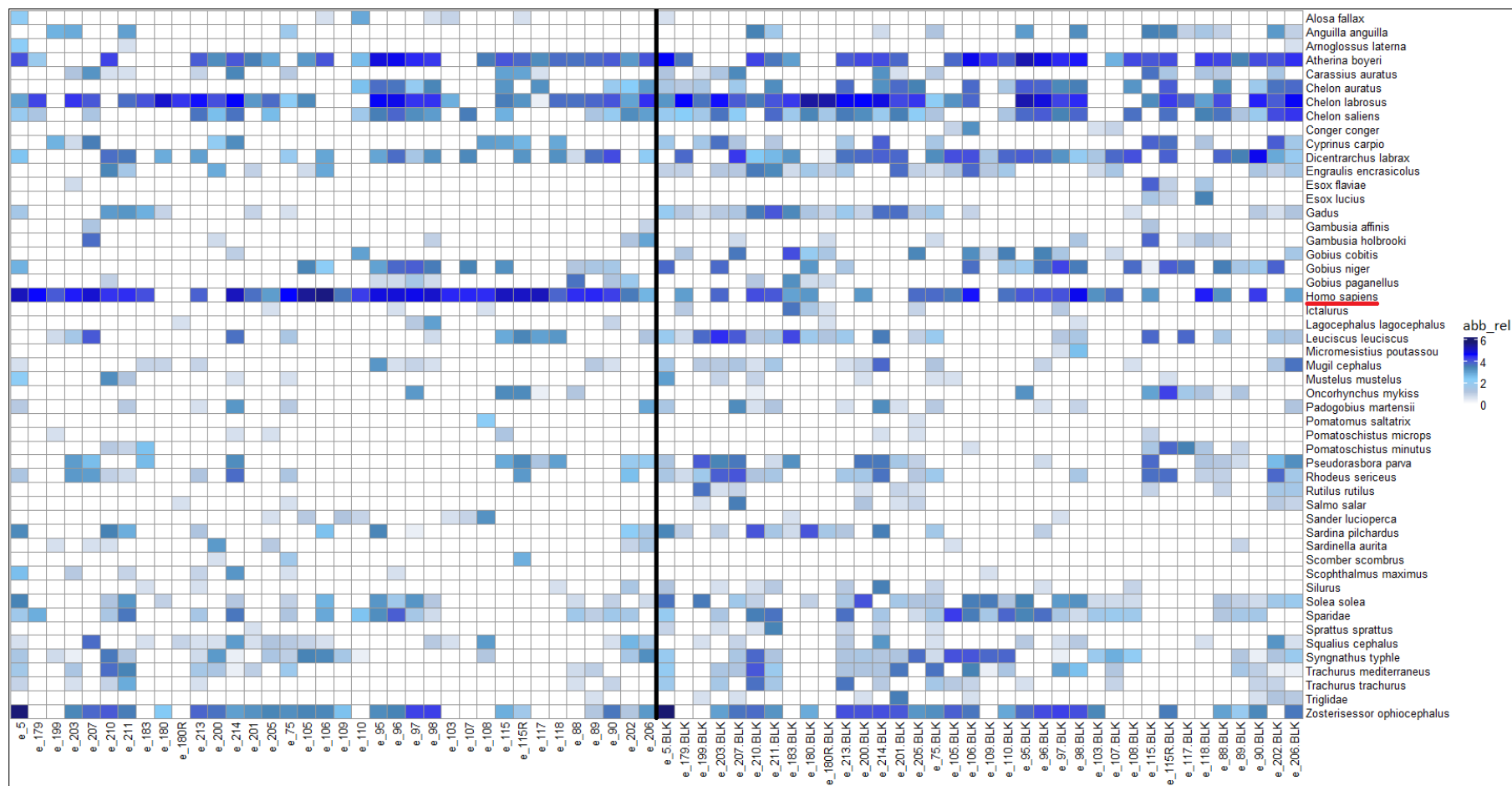
**Fig. 22** - Risultato dell'analisi dei dati mediante Adonis Test.

Dal risultato si può osservare come, insieme a stagione e sito di campionamento, anche l'utilizzo del *blocking primer* ha un effetto significativo sulla variabilità dell'abbondanza relativa all'interno del dataset.

Per rendere visualizzabile l'effetto del *blocking primer*, riporto in Fig. 23 una *heatmap*, nella quale viene rappresentata l'abbondanza relativa delle sequenze attribuite ad una specie per ogni campione. All'aumentare dell'abbondanza relativa delle sequenze, la corrispondente casella della *heatmap* sarà colorata più intensamente.

Per chiarezza, il grafico è stato diviso in due parti da una linea nera. Nella prima metà, con l'etichetta es. "e\_5" viene indicato un campione; nella seconda metà della *heatmap*, con l'etichetta es. "e\_5 BLK" viene indicato lo stesso campione, il cui DNA è stato però amplificato con l'utilizzo del *blocking primer*. Dalla *heatmap* è apprezzabile come ci sia una netta diminuzione delle sequenze attribuite a *Homo sapiens* (sottolineato in rosso nel grafico) e come aumentino invece le abbondanze relative delle altre specie individuate.





**Fig. 23** – Heatmap con campioni amplificati senza e con l'utilizzo del blocking primer Tele02\_Flong\_blk homo. La linea nera divide i campioni nel quale il protocollo è stato svolto senza l'utilizzo del blocking primer (sinistra) dagli stessi campioni, a destra, amplificati però secondo protocollo che prevede l'utilizzo del primer disegnato. Sottolineato in rosso, la specie *Homo sapiens*.

## 5. Discussione

Lo scopo di questo lavoro di tesi è stato quello di ottimizzare una metodica di analisi di DNA ambientale per il rilevamento della fauna ittica e di ottenere una prima stima della biodiversità all'interno della laguna di Venezia con questa metodica innovativa.

Al termine del lavoro di tesi, si possono considerare standardizzati definitivamente i protocolli sperimentali, sia per quanto riguarda l'estrazione del DNA genomico da filtro e la sua amplificazione, mediante PCR, con l'utilizzo del primer universale Tele02, sia per la successiva analisi bioinformatica dei dati. Inoltre, l'analisi di DNA ambientale, svolta per questa tesi, ha permesso l'identificazione di un buon numero di specie di pesci appartenenti alla laguna di Venezia e il rilevamento della specie invasiva *Lagocephalus lagocephalus*, specie mai registrata prima in laguna, la cui presenza è da confermare mediante indagini tradizionali. L'identificazione di una specie invasiva all'interno del dataset fornisce una conferma della potenzialità del *metabarcoding* con approccio passivo per il DNA ambientale nell'identificare anche specie rare, come visto in studi precedenti (Mathon *et al.* 2021). Anche se il dato, a causa delle problematiche intrinseche del DNA ambientale (spostamento del DNA ambientale causato dalle correnti, degradazione dello stesso, influenza di fattori ambientali sulla sua individuazione), deve essere valutato con occhio critico e non è mai da ritenersi una prova assoluta della presenza della specie, il risultato può essere considerato come un segnale precoce di allerta, che può indicare la necessità di una sorveglianza attiva.

Sono state inoltre identificate tre OTU (*Xiphias gladius*, *Thunnus sp.* e *Isurus oxyrinchus*), la cui presenza nella laguna di Venezia è da escludere o, per il tonno, da considerarsi assai poco probabile. Queste sequenze sono state ricondotte a contaminazioni da scarti di pescheria del mercato ittico. Si manifesta quindi un ulteriore limite dell'analisi di DNA ambientale, che potrebbe identificare come presenti specie il cui DNA deriva invece da scarti alimentari riversati in acqua.

Oltre al dato di presenza/assenza delle specie, le distribuzioni delle abbondanze relative osservate dai risultati sono in larga parte conformi alle attese, ovvero le specie maggiormente presenti in laguna, e in particolare le specie residenti lagunari, sono quelle presenti in maggiore abbondanza nel dataset. Inoltre, da una prima analisi statistica ed ecologica è stato possibile notare delle differenze nella distribuzione delle specie nei campioni, sulla base di variabili spazio-temporali, per quanto riguarda le loro abbondanze relative.

Dalle analisi risulta quindi fondamentale, come già evidenziato in letteratura (Taberlet *et al.* 2018) la scelta del barcode più appropriato; è da ricercarsi infatti il primer universale che permetta la miglior amplificazione della classe di interesse, che amplifichi quindi una regione conservata, ma eviti l'amplificazione eccessiva dei contaminanti, mai del tutto eliminabile quando si svolge un approccio passivo di DNA *metabarcoding*. In questo ambito, il primer universale Tele02 ha permesso l'individuazione di specie

appartenenti all'infraclasse dei teleostei, ma, essendo il frammento da esso amplificato (12S rDNA) molto conservato, ha portato all'amplificazione anche di molte sequenze attribuibili a specie di mammiferi e uccelli, che non sono oggetto di studio e che competono in PCR con le sequenze di interesse. Infatti, tra le sequenze analizzate sono state identificate numerose specie di mammiferi, quali *Homo sapiens*, *Canis lupus familiaris*, *Myocastor coypus* e di uccelli, come *Larus sp.*, *Meleagris gallopavo* e altri. A causa dell'incorrere di tali contaminanti, e al fine di limitarne la presenza nei risultati, si conferma la necessità di protocolli che prevedano la completa separazione tra gli spazi pre e post PCR (Taberlet *et al.* 2018) e la costante pulizia degli strumenti utilizzati (De Barba *et al.* 2014). Nonostante tutte le accortezze prese, è inevitabile l'incorrere di qualche tipo di contaminante (Ficetola *et al.* 2016). In questo lavoro di tesi, l'inclusione di numerosi controlli negativi nel protocollo sperimentale si è rilevata fondamentale da questo punto di vista, perché ha permesso di stimare l'entità della presenza di tali contaminanti e di filtrare, rispetto a questi, i risultati dei campioni, così da ottenere una stima più affidabile dell'abbondanza delle sequenze di interesse e di evitare i falsi positivi, errore più comune effettuato con il DNA *metabarcoding* (Ficetola *et al.* 2016).

Un modo per diminuire ulteriormente le contaminazioni da parte di DNA "esterno", che hanno anche l'effetto negativo di competere in PCR con le sequenze di interesse, riducendo il potere del DNA *metabarcoding* di rilevare le specie più rare (Harper *et al.* 2019, Bylemans *et al.* 2019), prevede l'utilizzo di *blocking primer* (Vestheim & Jarman 2008). Questo approccio è stato utilizzato nel presente lavoro di tesi, utilizzando un *blocking primer* in grado di bloccare l'amplificazione del DNA umano, che dalle analisi precedenti era stato rilevato come il maggior contaminante. Il *blocking primer*, disegnato sulla sequenza della corrispondente regione del genoma mitocondriale di *Homo sapiens*, ha ottenuto il risultato atteso. Infatti, ha permesso di bloccare, anche se non totalmente, l'amplificazione di DNA umano e, in parte, anche quella di altri contaminanti, come mammiferi e uccelli. Grazie al suo utilizzo, inoltre, sono state identificate tre specie non presenti nel dataset prodotto dal protocollo precedente. Infine, il suo utilizzo si è rivelato utile soprattutto nella determinazione più precisa dell'abbondanza delle specie nel campione, perché la mancanza di una grande parte delle sequenze attribuite al DNA di uomo ha permesso di aumentare in numero le sequenze attribuite invece ai teleostei. Risulta quindi chiaro che il DNA umano, insieme a quello di altri mammiferi, rappresenta un contaminante ambientale che compete, nella reazione di PCR, con le sequenze delle specie di interesse. Mediante l'utilizzo del *blocking primer* però è possibile, se non eliminare del tutto, diminuire significativamente le sequenze di contaminanti amplificate e permettere dunque di diminuire l'effetto competitivo sulle altre sequenze.

Indagando in modo critico il risultato dell'assegnazione tassonomica ottenuta dagli strumenti bioinformatici, si è ricercata la causa della mancata identificazione di alcune specie di teleostei, che sono stati invece rilevati precedentemente mediante indagini tradizionali e, in un progetto di tesi svolto in parallelo, con un primer non specifico per l'infraclasse, che amplifica la regione 16S rDNA mitocondriale. In particolare, sono state

individuate le specie *Aphanius fasciatus*, *Salaria pavo*, *Belone belone*, *Knipowitschia panizzae* e *Syngnathus abaster* attraverso l'utilizzo del primer Echi02, disegnato per l'amplificazione preferenziale di Echinodermi. La causa del mancato rilevamento con Tele02 è stata ricondotta all'assenza, nel database utilizzato per l'analisi, del frammento 12S mitocondriale delle suddette specie. Allo stesso modo, la mancanza della sequenza di riferimento nel database utilizzato ha causato, in alcuni casi, l'errata assegnazione ad una specie, che è stata corretta mediante la verifica della sequenza tramite BLAST, che, apparentemente, effettua la ricerca su un database più completo ed aggiornato di quello di riferimento utilizzato per le analisi. Ad esempio, *Sardinella aurita* è stata ricondotta alla specie *Sardinella longiceps*, e *Mustelus manazo* è stata ricondotta alla specie *Mustelus mustelus*. In altri casi, mediante il controllo manuale è stata resa più specifica l'assegnazione che da una prima a livello di genere è stata ricondotta a livello di specie. A titolo di esempio, il genere *Pseudorasbora* sp. è stato ricondotto alla specie *Pseudorasbora parva*, e il genere *Chelon* sp. alla specie *Chelon labrosus*. Alcuni errori nelle assegnazioni tassonomiche e la presenza di molte sequenze non assegnate ad una OTU (il 25% di sequenze in media), sono da ricondursi al fatto che il database utilizzato risulta poco specifico per le specie della laguna di Venezia, in quanto è un database non curato e le sequenze di riferimento al suo interno provengono da informazioni raccolte a livello mondiale. Questo rende quindi necessaria la costruzione di un database locale, maggiormente specifico, che comprenda tutte le specie di interesse nell'ambiente che si va a considerare. Nella costruzione di tale database è importante assicurarsi di includere un livello di specificità adeguato; un database troppo specifico porterebbe infatti alla possibile perdita di segnali, quali l'individuazione di specie invasive, non attese in laguna, a causa dell'esclusione a priori delle stesse dalle analisi. Un approccio integrativo potrebbe prevedere di utilizzare il database pubblico scaricato da EMBL, integrato con il database locale costruito *ad hoc* per la laguna di Venezia.

In questo lavoro di tesi è stata anche standardizzata l'analisi bioinformatica mediante il pacchetto Obitools3 (Boyer *et al.* 2016), che è stato ampiamente utilizzato e studiato in ogni suo aspetto. L'utilizzo di tale pacchetto, mediante i vari passaggi di filtrazione, ha permesso di ottenere sequenze di alta qualità e assegnazioni tassonomiche con identità molto alta.

Successivamente, si è cercato il modo migliore per effettuare analisi statistiche, che devono tenere conto di una grande mole di dati di partenza, e delle diverse variabili che possono influenzare la distribuzione delle specie nei campioni. In questa tesi è stato effettuato un primo tentativo di analisi ecologica, che potrà proseguire in modo più approfondito in futuro. In particolare, sulla base dell'esperienza effettuata, un approccio che unisca analisi descrittive, come analisi delle componenti principali, con un'analisi multivariata, sembra appropriato per visualizzare i risultati dal DNA ambientale e ottenere indicazioni sulla variazione spazio-temporale della biodiversità rilevata e sull'effetto significativo delle variabili in gioco.

## 6. Conclusioni

L'analisi di DNA ambientale si è dimostrata un utile strumento per la stima della biodiversità in laguna di Venezia. A partire da questo dato, la standardizzazione di un piano di campionamento che copra i siti di interesse in un arco temporale più lungo può portare ad una stima del cambiamento della biodiversità nel tempo e alla stima degli impatti umani sulla distribuzione delle specie. Una ricerca futura dovrà concentrarsi anche sulla costruzione di un database locale di riferimento che permetta una analisi più veloce e accurata delle sequenze identificate. Si dovranno inoltre standardizzare le analisi statistiche, che necessitano di un test adeguato ad una grande quantità di dati e molte variabili, e analisi ecologiche, al fine di ottenere dei dati confrontabili delle distribuzioni della fauna ittica. Sarà inoltre necessario indagare in modo più accurato sull'effetto degli scarti di pescheria sulla persistenza del DNA ambientale in laguna e la sua individuazione tramite l'analisi.

Per limitare l'effetto dei contaminanti sui risultati della ricerca, il protocollo utilizzato potrà essere integrato con l'utilizzo sistematico del *blocking primer* disegnato per *Homo sapiens*, ed eventualmente il disegno di ulteriori *blocking primer* per altre specie presenti in quantità rilevante.

Nell'ambito del progetto Venezia2021, un piano di indagine tramite analisi di DNA ambientale potrà fornire utili informazioni sull'impatto del MOSE sulla comunità ittica della laguna. Le criticità nell'analisi del DNA ambientale riguardano, in particolar modo, la stima quantitativa delle specie in un ambiente, più che il dato di presenza/assenza. A partire da questo presupposto, si necessita di guardare ai risultati in ottica critica, e considerare sempre la qualità dei dati prima di trarne conclusioni affrettate. Risulta infatti chiaro, da questo lavoro di tesi, che l'analisi del DNA ambientale accurata non può prescindere da un confronto con informazioni a priori o dal consulto di esperti di ecologia tradizionale, al fine di stabilire quanto siano accurate le analisi e fornire ipotesi riguardo contaminazioni, specie autoctone, specie invasive.

## 7. Bibliografia

Aglieri, G., Baillie, C., Mariani, S., Cattano, C., Calò, A., Turco, G., ... & Milazzo, M. (2021). Environmental DNA effectively captures functional diversity of coastal fish communities. *Molecular Ecology*, 30(13), 3127-3139.

Anderson, M.J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26: 32--46.

Barnes, M. A., & Turner, C. R. (2016). The Ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17(1), 1-17.

Begon, M., & Townsend, C. R. (2020). *Ecology: from individuals to ecosystems*. John Wiley & Sons.

Bellemain, E., Davey, M. L., Kauserud, H., Epp, L. S., Boessenkool, S., Coissac, E., ... & Brochmann, C. (2013). Fungal palaeodiversity revealed using high-throughput metabarcoding of ancient DNA from arctic permafrost. *Environmental microbiology*, 15(4), 1176-1189.

Binladen, J., Thomas, M., Gilbert, P., Campos, P. F., & Willerslev, E. (2007). 5'-tailed sequencing primers improve sequencing quality of PCR products. *Biotechniques*, 42(2), 174-176.

Blaber, S. J., Cyrus, D. P., Albaret, J. J., Ching, C. V., Day, J. W., Elliott, M., ... & Silvert, W. (2000). Effects of fishing on the structure and functioning of estuarine and nearshore ecosystems. *iCES Journal of Marine Science*, 57(3), 590-602.

Boessenkool, S., Epp, L. S., Haile, J., Bellemain, E. V. A., Edwards, M., Coissac, E., ... & Brochmann, C. (2012). Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, 21(8), 1806-1815.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., ... & De Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29(6), 358-367.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology resources*, 16(1), 176-182.

Bylemans, J., Gleeson, D. M., Duncan, R. P., Hardy, C. M., & Furlan, E. M. (2019). A performance evaluation of targeted eDNA and eDNA metabarcoding analyses for freshwater fishes. *Environmental DNA*, 1(4), 402-414.

Cilleros, K., Valentini, A., Allard, L., Dejean, T., Etienne, R., Grenouillet, G., ... & Brosse, S. (2019). Unlocking biodiversity and conservation studies in high-diversity environments using environmental DNA (eDNA): A test with Guianese freshwater fishes. *Molecular Ecology resources*, 19(1), 27-46.

Coissac, E., Riaz, T., & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, 21(8), 1834-1847.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., ... & Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7(9), 1008-1018.

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology resources*, 14(2), 306-323.

Deagle, B. E., Kirkwood, R., & Jarman, S. N. (2009). Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, 18(9), 2022-2038.

Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. *PloS one*, 9(2), e88786.

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872-5895.

Deiner, K., Walser, J. C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological conservation*, 183, 53-63.

Díaz-Ferguson, E., Herod, J., Galvez, J., & Moyer, G. (2014). Development of molecular markers for eDNA detection of the invasive African jewelfish (*Hemichromis letourneuxi*): a new tool for monitoring aquatic invasive species in National Wildlife Refuges. *Management of Biological Invasions*, 5(2), 121.

Doi, H., Katano, I., Sakata, Y., Souma, R., Kosuge, T., Nagano, M., ... & Tojo, K. (2017). Detection of an endangered aquatic heteropteran using environmental DNA in a wetland ecosystem. *Royal Society Open Science*, 4(7), 170568.

Egan, S. P., Barnes, M. A., Hwang, C. T., Mahon, A. R., Feder, J. L., Ruggiero, S. T., ... & Lodge, D. M. (2013). Rapid invasive species detection by combining environmental DNA with light transmission spectroscopy. *Conservation Letters*, 6(6), 402-409.

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic acids research*, 43(5), 2513-2524.

Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., ... & Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74(9), 1362-1374.

F. Cavraro, S. Redolfi Bristol, V. Georgalas, P. Torricelli, M. Zucchetta e P. Franzoi (2017) "Ingresso e distribuzione di uova, larve e giovanili di teleostei marini in laguna di Venezia: connettività mare-laguna e funzione di nursery" in *Il controllo ambientale della costruzione del MOSE* edito da P. Campostrini, C. Dabalà, P. Del Negro, L. Tosi, 375-409.

Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., ... & Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. *BMC genomics*, 11(1), 1-10.

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology letters*, 4(4), 423-425.

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding?.

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P. J., Griggo, C., Gielly, L., ... & Taberlet, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications*, 5(1), 1-7.

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PloS one*, 6(7), e22746.

Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J., & Waits, L. P. (2013). Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarum*). *Freshwater Science*, 32(3), 792-800.

Goldberg, C. S., Strickler, K. M., & Pilliod, D. S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, 183, 1-3.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., ... & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7(11), 1299-1307.

Hansen, B. K., Bekkevold, D., Clausen, L. W., & Nielsen, E. E. (2018). The sceptical optimist: challenges and perspectives for the application of



environmental DNA in marine fisheries. *Fish and Fisheries*, 19(5), 751-768.

Harper, L. R., Handley, L. L., Carpenter, A. I., Ghazali, M., Di Muri, C., Macgregor, C. J., ... & Hänfling, B. (2019). Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals. *Biological Conservation*, 238, 108225.

Hebert, P. D., Ratnasingham, S., & De Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(suppl\_1), S96-S99.

Herbert, N. A., & Steffensen, J. F. (2005). The response of Atlantic cod, *Gadus morhua*, to progressive hypoxia: fish swimming speed and physiological stress. *Marine Biology*, 147(6), 1403-1412.

Huggins, L. G., Koehler, A. V., Schunack, B., Inpankaew, T., & Traub, R. J. (2020). A host-specific blocking primer combined with optimal DNA extraction improves the detection capability of a metabarcoding protocol for canine vector-borne bacteria. *Pathogens*, 9(4), 258.

Jane, S. F., Wilcox, T. M., McKelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., ... & Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology resources*, 15(1), 216-227.

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation letters*, 4(2), 150-157.

Jeunen, G. J., Knapp, M., Spencer, H. G., Taylor, H. R., Lamare, M. D., Stat, M., ... & Gemmell, N. J. (2019). Species-level biodiversity assessment using marine environmental DNA metabarcoding requires protocol optimization and standardization. *Ecology and Evolution*, 9(3), 1323-1335.

Karahan, A., Douek, J., Paz, G., Stern, N., Kideys, A. E., Shaish, L., ... & Rinkevich, B. (2017). Employing DNA barcoding as taxonomy and conservation tools for fish species censuses at the southeastern Mediterranean, a hot-spot area for biological invasion. *Journal for Nature Conservation*, 36, 1-9.

Kircher, M., Heyn, P., & Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC genomics*, 12(1), 1-14.

Kowalczyk, R., Taberlet, P., Coissac, E., Valentini, A., Miquel, C., Kamiński, T., & Wójcik, J. M. (2011). Influence of management practices on large herbivore diet—Case of European bison in Białowieża Primeval Forest (Poland). *Forest Ecology and Management*, 261(4), 821-828.

Kwok, S. A., & Higuchi, R. (1989). Avoiding false positives with PCR. *Nature*, 339(6221), 237-238.

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H., Hart, M. M., Powell, J. R., Klironomos, J. N., ... & Dunfield, K. E. (2007). Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, 39(12), 2977-2991.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *nature*, 362(6422), 709-715.

Lodge, D. M., Williams, S., MacIsaac, H. J., Hayes, K. R., Leung, B., Reichard, S., ... & McMichael, A. (2006). Biological invasions: recommendations for US policy and management. *Ecological applications*, 16(6), 2035-2054.

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific reports*, 8(1), 1-11.

Maruyama, A., Nakamura, K., Yamanaka, H., Kondoh, M., & Minamoto, T. (2014). The release rate of environmental DNA from juvenile and adult fish. *PLoS One*, 9(12), e114639.

Mathon, L., Valentini, A., Guérin, P. E., Normandeau, E., Noel, C., Lionnet, C., ... & Manel, S. (2021). Benchmarking bioinformatic tools for fast and accurate eDNA metabarcoding species identification. *Molecular Ecology Resources*, 21(7), 2565-2579.

McDonald, L. L., & Thompson, W. L. (2004). Sampling rare populations. *Sampling rare or elusive species: concepts, designs, and techniques for estimating population parameters*, 11-42.

McLusky & Elliott, 2007. Transitional waters: A new approach, semantics or just muddying the waters? *Estuarine, Coastal and Shelf Science* 71: 359-363.

Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS One*, 9(11), e113346.

Nielsen, K. M., Johnsen, P. J., Bensasson, D., & Daffonchio, D. (2007). Release and persistence of extracellular DNA in the environment. *Environmental biosafety research*, 6(1-2), 37-53.

Pansu, J., Giguët-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., ... & Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, 24(7), 1485-1498.

Pavan, C. (1989). *Sile. Alla scoperta del fiume: Immagini, storia, itinerari*. pp. 368. La Bottega di Manuzio, Treviso.

Pérez-Ruzafa, A., & Marcos, C. (2012). Fisheries in coastal lagoons: An assumed but poorly researched aspect of the Ecology and functioning of coastal lagoons. *Estuarine, Coastal and Shelf Science*, 110, 15-31.

Piaggio, A. J., Engeman, R. M., Hopken, M. W., Humphrey, J. S., Keacher, K. L., Bruce, W. E., & Avery, M. L. (2014). Detecting an elusive invasive species: A diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology resources*, 14(2), 374-380.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, 45(3), 219-235.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70(8), 1123-1130.

Post, J. R., & Lee, J. A. (1996). Metabolic ontogeny of teleost fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, 53(4), 910-923.

Prosser, J. I. (2010). Replicate or lie. *Environmental microbiology*, 12(7), 1806-1810.

Quartarella, E. (2013). Analisi del popolamento ittico della Laguna di Venezia: relazioni con le variabili e le pressioni ambientali.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in Ecology. *Journal of applied Ecology*, 51(5), 1450-1459.

Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., & Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Molecular Ecology resources*, 15(1), 168-176.

Rinaldo, A. (1997). *Equilibrio fisico e idrogeologico della laguna*. Fondazione ENI Enrico Mattei.

Rojahn, J., Gleeson, D. M., Furlan, E., Haeusler, T., & Bylemans, J. (2021). Improving the detection of rare native fish species in environmental DNA metabarcoding surveys. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 31(4), 990-997.

- Sassoubre, L. M., Yamahara, K. M., Gardner, L. D., Block, B. A., & Boehm, A. B. (2016). Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environmental Science & Technology*, 50(19), 10456-10464.
- Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS one*, 6(12), e27310.
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15(6), 1289-1303.
- Schurmann, H., & Steffensen, J. F. (1997). Effects of temperature, hypoxia and activity on the metabolism of juvenile Atlantic cod. *Journal of fish Biology*, 50(6), 1166-1180.
- Shaw, J. L., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, 131-138.
- Shehzad, W., Riaz, T., Nawaz, M. A., Miquel, C., Poillot, C., Shah, S. A., ... & Taberlet, P. (2012). Carnivore diet analysis based on next-generation sequencing: Application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, 21(8), 1951-1965.
- Signa, G. (2011). Interconnessioni trofiche mediate da fattori biotici in aree di transizione.
- Simon, C., & Daniel, R. (2011). Metagenomic analyses: past and future trends. *Applied and environmental microbiology*, 77(4), 1153-1161.
- Spear, S. F., Groves, J. D., Williams, L. A., & Waits, L. P. (2015). Using environmental DNA methods to improve detectability in a hellbender (*Cryptobranchus alleganiensis*) monitoring program. *Biological Conservation*, 183, 38-45.
- Su, M., Liu, H., Liang, X., Gui, L., & Zhang, J. (2018). Dietary analysis of marine fish species: enhancing the detection of prey-specific DNA sequences via high-throughput sequencing using blocking primers. *Estuaries and coasts*, 41(2), 560-571.
- Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For biodiversity research and monitoring*. Oxford University Press.
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental dna. *Molecular Ecology*, 21(8), 1789-1793.
- Tagliapietra, D., Magni, P., Basset, A., & Viaroli, P. (2014). Ecosistemi costieri di transizione: trasformazioni recenti, pressioni antropiche dirette

e possibili impatti del cambiamento climatico. *Biologia Ambientale*, 28(2), 101-111.

Takahara, T., Minamoto, T., Yamanaka, H., Doi, H., & Kawabata, Z. I. (2012). Estimation of fish biomass using environmental DNA. *PloS one*, 7(4), e35868.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4-18.

Thomsen, P. F., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... & Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11), 2565-2573.

Tréguier, A., Paillisson, J. M., Dejean, T., Valentini, A., Schlaepfer, M. A., & Roussel, J. M. (2014). Environmental DNA surveillance for invertebrate species: Advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, 51(4), 871-879.

Tsuji, S., Ushio, M., Sakurai, S., Minamoto, T., & Yamanaka, H. (2017). Water temperature-dependent degradation of environmental DNA and its relation to bacterial abundance. *PLoS One*, 12, e0176608.

Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, 24(2), 110-117.

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... & Dejean, T. (2016). Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25(4), 929-942.

Vestheim, H., & Jarman, S. N. (2008). Blocking primers to enhance PCR amplification of rare sequences in mixed samples—a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, 5(1), 1-11.

Wilcox, T. M., Schwartz, M. K., McKelvey, K. S., Young, M. K., & Lowe, W. H. (2014). A blocking primer increases specificity in environmental DNA detection of bull trout (*Salvelinus confluentus*). *Conservation Genetics Resources*, 6(2), 283-284.

Yoccoz, N. G. (2012). The future of environmental DNA in Ecology. *Molecular Ecology*, 21(8), 2031-2038.

Zinger, L., Gobet, A., & Pommier, T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, 21(8), 1878-1896.

## 8. Sitografia

Data di ultima consultazione di tutti i siti: 14 febbraio 2022.

<https://earth.google.com/web/@45.30568737,12.37494061,1.4832982a,52270.99429557d,35y,-29.65474209h,20.43267667t,359.99999999r>

[https://en.wikipedia.org/wiki/European\\_bass](https://en.wikipedia.org/wiki/European_bass).

<https://spain.inaturalist.org/taxa/118704-Gobius-niger>.

[www.corila.it/](http://www.corila.it/)

[www.corila.it/it/Venezia2021](http://www.corila.it/it/Venezia2021)

[www.ebi.ac.uk/](http://www.ebi.ac.uk/)

[www.fishbase.de/summary/isurus-oxyrinchus.html](http://www.fishbase.de/summary/isurus-oxyrinchus.html)

[www.fishbase.se/photos/PicturesSummary.php?StartRow=2&ID=4292&what=species&TotRec=7](http://www.fishbase.se/photos/PicturesSummary.php?StartRow=2&ID=4292&what=species&TotRec=7)

[www.fishbase.se/summary/226](http://www.fishbase.se/summary/226)

[www.fishbase.se/summary/Chelon-labrosus](http://www.fishbase.se/summary/Chelon-labrosus)

[www.fishbase.se/summary/dicentrarchus-labrax](http://www.fishbase.se/summary/dicentrarchus-labrax)

[www.fishbase.se/summary/Solea-solea](http://www.fishbase.se/summary/Solea-solea)

[www.fishbase.se/summary/thunnus-albacares.html](http://www.fishbase.se/summary/thunnus-albacares.html)

[www.istitutoveneto.org](http://www.istitutoveneto.org)

[www.italianostravenezia.org/2020/10/18/autorita-per-la-laguna-di-venezia-le-nostre-considerazioni/](http://www.italianostravenezia.org/2020/10/18/autorita-per-la-laguna-di-venezia-le-nostre-considerazioni/)

[www.iucn.it/scheda.php?id=-2081245894](http://www.iucn.it/scheda.php?id=-2081245894)

[www.mosevenezia.eu/mose/](http://www.mosevenezia.eu/mose/)

[www.snorkeling-report.com/species/chelon-labrosus](http://www.snorkeling-report.com/species/chelon-labrosus)

## 9. Appendici

**Appendice I:** filtri campionati per le analisi di DNA ambientale. Nelle informazioni vengono specificati sito e data di campionamento e tipologia di filtro. I campioni sono ordinati per data in ordine cronologico.

ID_sample	Sito	Data	Tipologia filtro
CH_79	Chioggia	13/11/2018	Cellulosa
CH_80	Chioggia	13/11/2018	Cellulosa
CH_81	Chioggia	13/11/2018	Cellulosa
CH_82	Chioggia	13/11/2018	Cellulosa
CH_83	Chioggia	13/11/2018	Cellulosa
CH_84	Chioggia	13/11/2018	Cellulosa
CH_85	Chioggia	13/11/2018	Cellulosa
CH_86	Chioggia	13/11/2018	Cellulosa
CH_71	Chioggia	14/11/2018	GF/C
CH_72	Chioggia	14/11/2018	GF/C
CH_73	Chioggia	14/11/2018	GF/C
CH_74	Chioggia	14/11/2018	GF/C
CH_75	Chioggia	14/11/2018	GF/C
CH_76	Chioggia	14/11/2018	GF/C
CH_77	Chioggia	14/11/2018	GF/C
CH_78	Chioggia	14/11/2018	GF/C
LD_162A	Lido	15/11/2018	0.22 µm
LD_162B	Lido	15/11/2018	3 µm
LD_162C	Lido	15/11/2018	20 µm
MG_174A	Marghera	15/11/2018	0.22 µm
MG_174B	Marghera	15/11/2018	3 µm
MG_174C	Marghera	15/11/2018	20 µm
PR_186A	Palude della Rosa	15/11/2018	0.22 µm
PR_186B	Palude della Rosa	15/11/2018	3 µm
PR_186C	Palude della Rosa	15/11/2018	20 µm
SG_198A	San Giuliano	15/11/2018	0.22 µm
SG_198B	San Giuliano	15/11/2018	3 µm
SG_198C	San Giuliano	15/11/2018	20 µm
LD_163A	Lido	13/02/2019	0.22 µm
LD_163B	Lido	13/02/2019	3 µm
LD_163C	Lido	13/02/2019	20 µm
MG_175A	Marghera	13/02/2019	0.22 µm
MG_175B	Marghera	13/02/2019	3 µm
MG_175C	Marghera	13/02/2019	20 µm
PR_187A	Palude della Rosa	13/02/2019	0.22 µm
PR_187B	Palude della Rosa	13/02/2019	3 µm
PR_187C	Palude della Rosa	13/02/2019	20 µm
SG_199A	San Giuliano	13/02/2019	0.22 µm
SG_199B	San Giuliano	13/02/2019	3 µm

SG_199C	San Giuliano	13/02/2019	20 µm
CH_1	Chioggia	08/03/2019	GF/C
CH_2	Chioggia	08/03/2019	GF/C
CH_3	Chioggia	08/03/2019	GF/C
CH_4	Chioggia	08/03/2019	GF/C
CH_5	Chioggia	08/03/2019	GF/C
CH_6	Chioggia	08/03/2019	Cellulosa
CH_7	Chioggia	08/03/2019	GF/C
CH_8	Chioggia	08/03/2019	Cellulosa
CH_9	Chioggia	08/03/2019	GF/C
CH_10	Chioggia	08/03/2019	Cellulosa
CH_11	Chioggia	08/03/2019	GF/C
CH_12	Chioggia	08/03/2019	Cellulosa
CH_13	Chioggia	08/03/2019	GF/C
CH_14	Chioggia	08/03/2019	Cellulosa
CH_15	Chioggia	08/03/2019	GF/C
CH_16	Chioggia	08/03/2019	Cellulosa
CH_17	Chioggia	08/03/2019	GF/C
CH_18	Chioggia	08/03/2019	Cellulosa
CH_19	Chioggia	14/03/2019	GF/C
CH_20	Chioggia	14/03/2019	GF/C
CH_21	Chioggia	14/03/2019	Cellulosa
CH_22	Chioggia	14/03/2019	Cellulosa
PR_23	Palude della Rosa	14/03/2019	GF/C
PR_24	Palude della Rosa	14/03/2019	GF/C
PR_25	Palude della Rosa	14/03/2019	Cellulosa
PR_26	Palude della Rosa	14/03/2019	Cellulosa
PR_27	Palude della Rosa	12/4/2019	GF/C
PR_28	Palude della Rosa	12/4/2019	Cellulosa
PR_29	Palude della Rosa	12/4/2019	GF/C
PR_30	Palude della Rosa	12/4/2019	Cellulosa
CH_31	Chioggia	17/4/2019	GF/C
CH_32	Chioggia	17/4/2019	GF/F
CH_33	Chioggia	17/4/2019	Cellulosa
PR_34	Palude della Rosa	18/04/2019	GF/C
PR_35	Palude della Rosa	18/04/2019	GF/F
PR_36	Palude della Rosa	18/04/2019	Cellulosa
PR_37	Palude della Rosa	18/04/2019	GF/C
PR_38	Palude della Rosa	18/04/2019	GF/F
PR_39	Palude della Rosa	18/04/2019	Cellulosa
PR_40	Palude della Rosa	18/04/2019	GF/C
PR_41	Palude della Rosa	18/04/2019	GF/F
PR_42	Palude della Rosa	18/04/2019	Cellulosa
CH_152A	Chioggia	15/05/2019	0.22 µm
CH_152B	Chioggia	15/05/2019	3 µm
CH_152C	Chioggia	15/05/2019	20 µm



LD_164A	Lido	15/05/2019	0.22 µm
LD_164B	Lido	15/05/2019	3 µm
LD_164C	Lido	15/05/2019	20 µm
MG_176A	Marghera	15/05/2019	0.22 µm
MG_176B	Marghera	15/05/2019	3 µm
MG_176C	Marghera	15/05/2019	20 µm
PR_188A	Palude della Rosa	15/05/2019	0.22 µm
PR_188B	Palude della Rosa	15/05/2019	3 µm
PR_188C	Palude della Rosa	15/05/2019	20 µm
SG_200A	San Giuliano	15/05/2019	0.22 µm
SG_200B	San Giuliano	15/05/2019	3 µm
SG_200C	San Giuliano	15/05/2019	20 µm
CH_43	Chioggia	16/5/2019	GF/C
CH_44	Chioggia	16/5/2019	GF/C
CH_47	Chioggia	16/5/2019	Cellulosa
CH_48	Chioggia	16/5/2019	Cellulosa
PR_45	Palude della Rosa	16/5/2019	GF/C
PR_46	Palude della Rosa	16/5/2019	GF/C
PR_49	Palude della Rosa	16/5/2019	Cellulosa
PR_50	Palude della Rosa	16/5/2019	Cellulosa
PR_51	Palude della Rosa	12/06/2019	GF/C
PR_52	Palude della Rosa	12/06/2019	Cellulosa
PR_53	Palude della Rosa	12/06/2019	GF/C
PR_54	Palude della Rosa	12/06/2019	Cellulosa
PR_55	Palude della Rosa	08/07/2019	GF/C
PR_56	Palude della Rosa	08/07/2019	GF/C
PR_57	Palude della Rosa	08/07/2019	Cellulosa
PR_58	Palude della Rosa	08/07/2019	Cellulosa
PR_59	Palude della Rosa	08/07/2019	GF/C
PR_60	Palude della Rosa	08/07/2019	GF/C
PR_61	Palude della Rosa	08/07/2019	Cellulosa
PR_62	Palude della Rosa	08/07/2019	Cellulosa
CH_67	Chioggia	22/08/2019	GF/C
CH_68	Chioggia	22/08/2019	GF/C
CH_69	Chioggia	22/08/2019	Cellulosa
CH_70	Chioggia	22/08/2019	Cellulosa
CH_153A	Chioggia	22/08/2019	0.22 µm
CH_153B	Chioggia	22/08/2019	3 µm
CH_153C	Chioggia	22/08/2019	20 µm
LD_165A	Lido	22/08/2019	0.22 µm
LD_165B	Lido	22/08/2019	3 µm
LD_165C	Lido	22/08/2019	20 µm
MG_177A	Marghera	22/08/2019	0.22 µm
MG_177B	Marghera	22/08/2019	3 µm
MG_177C	Marghera	22/08/2019	20 µm
PR_63	Palude della Rosa	22/08/2019	GF/C

PR_64	Palude della Rosa	22/08/2019	GF/C
PR_65	Palude della Rosa	22/08/2019	Cellulosa
PR_66	Palude della Rosa	22/08/2019	Cellulosa
PR_189A	Palude della Rosa	22/08/2019	0.22 µm
PR_189B	Palude della Rosa	22/08/2019	3 µm
PR_189C	Palude della Rosa	22/08/2019	20 µm
SG_201A	San Giuliano	22/08/2019	0.22 µm
SG_201B	San Giuliano	22/08/2019	3 µm
SG_201C	San Giuliano	22/08/2019	20 µm
PR_87	Palude della Rosa	23/09/2019	GF/C
PR_88	Palude della Rosa	23/09/2019	GF/C
PR_89	Palude della Rosa	23/09/2019	Cellulosa
PR_90	Palude della Rosa	23/09/2019	Cellulosa
PR_95	Palude della Rosa	08/10/2019	GF/C
PR_96	Palude della Rosa	08/10/2019	GF/C
PR_97	Palude della Rosa	08/10/2019	Cellulosa
PR_98	Palude della Rosa	08/10/2019	Cellulosa
CH_105	Chioggia	04/11/2019	GF/C
CH_106	Chioggia	04/11/2019	GF/C
CH_109	Chioggia	04/11/2019	Cellulosa
CH_110	Chioggia	04/11/2019	Cellulosa
CH_154A	Chioggia	04/11/2019	0.22 µm
LD_166A	Lido	04/11/2019	0.22 µm
LD_166B	Lido	04/11/2019	3 µm
LD_166C	Lido	04/11/2019	20 µm
MG_178A	Marghera	04/11/2019	0.22 µm
MG_178B	Marghera	04/11/2019	3 µm
MG_178C	Marghera	04/11/2019	20 µm
PR_103	Palude della Rosa	04/11/2019	GF/C
PR_104	Palude della Rosa	04/11/2019	GF/C
PR_107	Palude della Rosa	04/11/2019	Cellulosa
PR_108	Palude della Rosa	04/11/2019	Cellulosa
PR_190A	Palude della Rosa	04/11/2019	0.22 µm
PR_190B	Palude della Rosa	04/11/2019	3 µm
PR_190C	Palude della Rosa	04/11/2019	20 µm
SG_202A	San Giuliano	04/11/2019	0.22 µm
SG_202B	San Giuliano	04/11/2019	3 µm
SG_202C	San Giuliano	04/11/2019	20 µm
PR_115	Palude della Rosa	04/12/2019	GF/C
PR_116	Palude della Rosa	04/12/2019	GF/C
PR_117	Palude della Rosa	04/12/2019	Cellulosa
PR_118	Palude della Rosa	04/12/2019	Cellulosa
PR_123	Palude della Rosa	16/01/2020	GF/C
PR_124	Palude della Rosa	16/01/2020	GF/C
PR_125	Palude della Rosa	16/01/2020	GF/C
PR_126	Palude della Rosa	16/01/2020	GF/C

PR_127	Palude della Rosa	16/01/2020	Cellulosa
PR_128	Palude della Rosa	16/01/2020	Cellulosa
PR_129	Palude della Rosa	16/01/2020	Cellulosa
PR_130	Palude della Rosa	16/01/2020	Cellulosa
CH_135	Chioggia	17/02/2020	GF/C
CH_136	Chioggia	17/02/2020	GF/C
CH_155A	Chioggia	17/02/2020	0.22 µm
CH_155B	Chioggia	17/02/2020	3 µm
CH_155C	Chioggia	17/02/2020	20 µm
LD_167A	Lido	17/02/2020	0.22 µm
LD_167B	Lido	17/02/2020	3 µm
LD_167C	Lido	17/02/2020	20 µm
MG_179A	Marghera	17/02/2020	0.22 µm
MG_179B	Marghera	17/02/2020	3 µm
MG_179C	Marghera	17/02/2020	20 µm
PR_137	Palude della Rosa	17/02/2020	GF/C
PR_138	Palude della Rosa	17/02/2020	GF/C
PR_139	Palude della Rosa	17/02/2020	GF/C
PR_140	Palude della Rosa	17/02/2020	GF/C
PR_141	Palude della Rosa	17/02/2020	GF/C
PR_142	Palude della Rosa	17/02/2020	GF/C
PR_143	Palude della Rosa	17/02/2020	GF/C
PR_144	Palude della Rosa	17/02/2020	GF/C
PR_191A	Palude della Rosa	17/02/2020	0.22 µm
PR_191B	Palude della Rosa	17/02/2020	3 µm
PR_191C	Palude della Rosa	17/02/2020	20 µm
SG_203A	San Giuliano	17/02/2020	0.22 µm
SG_203B	San Giuliano	17/02/2020	3 µm
SG_203C	San Giuliano	17/02/2020	20 µm
CH_156A	Chioggia	29/05/2020	0.22 µm
CH_156B	Chioggia	29/05/2020	3 µm
CH_156C	Chioggia	29/05/2020	20 µm
LD_168A	Lido	29/05/2020	0.22 µm
LD_168B	Lido	29/05/2020	3 µm
LD_168C	Lido	29/05/2020	20 µm
MG_180A	Marghera	29/05/2020	0.22 µm
MG_180B	Marghera	29/05/2020	3 µm
MG_180C	Marghera	29/05/2020	20 µm
PR_192A	Palude della Rosa	29/05/2020	0.22 µm
PR_192B	Palude della Rosa	29/05/2020	3 µm
PR_192C	Palude della Rosa	29/05/2020	20 µm
SG_204A	San Giuliano	29/05/2020	0.22 µm
SG_204B	San Giuliano	29/05/2020	3 µm
SG_204C	San Giuliano	29/05/2020	20 µm
CH_157A	Chioggia	27/08/2020	0.22 µm
CH_157B	Chioggia	27/08/2020	3 µm

CH_157C	Chioggia	27/08/2020	20 µm
LD_169A	Lido	27/08/2020	0.22 µm
LD_169B	Lido	27/08/2020	3 µm
LD_169C	Lido	27/08/2020	20 µm
MG_181A	Marghera	27/08/2020	0.22 µm
MG_181B	Marghera	27/08/2020	3 µm
MG_181C	Marghera	27/08/2020	20 µm
PR_193A	Palude della Rosa	27/08/2020	0.22 µm
PR_193B	Palude della Rosa	27/08/2020	3 µm
PR_193C	Palude della Rosa	27/08/2020	20 µm
SG_205A	San Giuliano	27/08/2020	0.22 µm
SG_205B	San Giuliano	27/08/2020	3 µm
SG_205C	San Giuliano	27/08/2020	20 µm
CH_158A	Chioggia	09/11/2020	0.22 µm
CH_158B	Chioggia	09/11/2020	3 µm
CH_158C	Chioggia	09/11/2020	20 µm
LD_170A	Lido	09/11/2020	0.22 µm
LD_170B	Lido	09/11/2020	3 µm
LD_170C	Lido	09/11/2020	20 µm
MG_182A	Marghera	09/11/2020	0.22 µm
MG_182B	Marghera	09/11/2020	3 µm
MG_182C	Marghera	09/11/2020	20 µm
PR_194A	Palude della Rosa	09/11/2020	0.22 µm
PR_194B	Palude della Rosa	09/11/2020	3 µm
PR_194C	Palude della Rosa	09/11/2020	20 µm
SG_206A	San Giuliano	09/11/2020	0.22 µm
SG_206B	San Giuliano	09/11/2020	3 µm
SG_206C	San Giuliano	09/11/2020	20 µm
CH_159A	Chioggia	19/02/2021	0.22 µm
CH_159B	Chioggia	19/02/2021	3 µm
CH_159C	Chioggia	19/02/2021	20 µm
LD_171A	Lido	19/02/2021	0.22 µm
LD_171B	Lido	19/02/2021	3 µm
LD_171C	Lido	19/02/2021	20 µm
MG_183A	Marghera	19/02/2021	0.22 µm
MG_183B	Marghera	19/02/2021	3 µm
MG_183C	Marghera	19/02/2021	20 µm
PR_195A	Palude della Rosa	19/02/2021	0.22 µm
PR_195B	Palude della Rosa	19/02/2021	3 µm
PR_195C	Palude della Rosa	19/02/2021	20 µm
SG_207A	San Giuliano	19/02/2021	0.22 µm
SG_207B	San Giuliano	19/02/2021	3 µm
SG_207C	San Giuliano	19/02/2021	20 µm
CH_210A	Chioggia	19/05/2021	0.22 µm
CH_210B	Chioggia	19/05/2021	3 µm
CH_210C	Chioggia	19/05/2021	20 µm

LD_211A	Lido	19/05/2021	0.22 µm
LD_211B	Lido	19/05/2021	3 µm
LD_211C	Lido	19/05/2021	20 µm
MG_212A	Marghera	19/05/2021	0.22 µm
MG_212B	Marghera	19/05/2021	3 µm
MG_212C	Marghera	19/05/2021	20 µm
PR_213A	Palude della Rosa	19/05/2021	0.22 µm
PR_213B	Palude della Rosa	19/05/2021	3 µm
PR_213C	Palude della Rosa	19/05/2021	20 µm
SG_214A	San Giuliano	19/05/2021	0.22 µm
SG_214B	San Giuliano	19/05/2021	3 µm
SG_214C	San Giuliano	19/05/2021	20 µm

## Appendice II: comandi estesi e completi effettuati su Obitools3 per le analisi bioinformatiche.

### 1. Import dataset e allineamento delle reads

```
> obi import directory_raw_sequences/reads_Forward.fastq dms/reads1
> obi import directory_raw_sequences/reads_Forward.fastq dms/reads2
> obi alignpairedend -R dms/reads2 dms/reads1 dms/aligned_reads
```

### 2. Selezione qualitative delle sequenze allineate

```
> obi stats -a score_norm dms/aligned_reads
> obi grep -p "sequence['score_norm'] > 0.3" dms/aligned_reads dms/good_sequences
> obi import -ngsfilter directory_ngsfilter/ngsfilter.txt dms/ngsfile
> obi ngsfilter -t dms/ngsfile -u dms/unidentified_sequences dms/good_sequences dms/identified_sequences
> obi uniq -m sample dms/identified_sequences dms/dereplicated_sequences
> obi annotate -k COUNT -k MERGED_sample dms/dereplicated_sequences dms/cleaned_metadata_sequences
> obi grep -p "len(sequence)>=80 and sequence ['COUNT']>=10" dms/cleaned_metadata_sequences dms/denoised_sequences
> obi clean -s MERGED_sample -r 0.05 -H dms/denoised_sequences dms/cleaned_sequences
```

### 3. EMBL

```
> obi import -embl EMBL dms/embl_refs
> wget https://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdump.tar.gz
> obi import - - taxdump taxdump.tar.gz dms/taxonomy/my_tax
> obi ecopcr -e 2 -l 100 -L 300 -F AAACTCGTGCCAGCCACC -R GGGTATCTAATCCCAGTTTG -- taxonomy dms/taxonomy/my_tax dms/embl_refs dms/v_refs
> obi grep -- require-rank=species -- require-rank=genus -- require-rank=family -- taxonomy dms/taxonomy/my_tax dms/v_refs_clean
> obi uniq - - taxonomy dms/taxonomy/my_tax dms/v_refs_uniq dms/v_refs_uniq
> obi grep - - require-rank=family - - taxonomy dms/taxonomy/my_tax dms/v_refs_uniq_clean
> obi build_ref_db -t 0.97 - - taxonomy dms/taxonomy/my_tax dms/c_refs_uniq_clean dms/v_db_97
```

### 4. Assegnazione tassonomica

```
> obi ecotag -m 0.97 - - taxonomy dms/taxonomy/my_tax -R dms/v_db_97 dms/cleaned_sequences dms/assigned_sequences
> obi stat SCIENTIFIC_NAME dms/assigned_sequences
```

### 5. Esportazione dei dati

```
> obi export - - fasta-output dms/assigned_sequences > risultati.tab
```